N d'ordre: **224/2024-C/MT**

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université des Sciences et de la Technologie Houari Boumediène

Faculté de Mathématiques

THÈSE DE DOCTORAT

Présentée pour l'obtention du grade de Docteur

En: **MATHEMATIQUES**

Spécialité: **Probabilités, Statistiques et Applications**

Par: **Bayarassou Nassima**

Sujet

Estimation non paramétrique de la fonction de régression par erreur relative pour des données tronquées à gauche et censurées à droite soumises à une structure de dépendance

Soutenue publiquement le 03/11/2024, devant le jury composé de:

M. Abdelkader TATACHAK	Professeur	à l' USTHB, Algérie	Président
M. Salim BOUZEBDA	Professeur	à l'UTC, France	Examinateur
Mme. Zohra GUESSOUM	Professeur	à l'USTHB, Algérie	Examinatrice
Mme. Farida HAMRANI	Maître de conférence/B	à l'UMMTO, Algérie	Co-Directrice de Thèse
M. Elias OULD SAÏD	Professeur	à l'IUT, Calais, France	Directeur de Thèse
Mme. Ourida SADKI	Professeur	à l'USTHB, Algérie	Examinatrice

Order N: **224/2024-C/MT**

Democratic and Popular Republic of Algeria

Ministry of Higher Education and Scientific Research

University of Sciences and Technology Houari Boumediene

Faculty of Mathematics

DOCTORAL THESIS

Presented in view of obtaining the degree of Doctor of Philosophy

In: **MATHEMATICS**

Speciality: **Probability, Statistics and Applications**

By: **Nassima Bayarassou**

Topic

Nonparametric estimation of the relative error regression function for left truncated and right censored data with a dependence structure

Publicly defended on 03/11/2024, to the jury composed of:

Mr. Abdelkader TATACHAK	Professor	at USTHB, Algeria	Chairman
Mr. Salim BOUZEBDA	Professor	at UTC, France	Examiner
Mrs. Zohra GUESSOUM	Professor	at USTHB, Algeria	Examiner
Mrs. Farida HAMRANI	Senior Lecturer/B	at UMMTO, Algeria	Co-Supervisor
Mr. Elias OULD SAÏD	Professor	at IUT, Calais, France	Supervisor
Mrs. Ourida SADKI	Professor	at USTHB, Algeria	Examiner

This thesis is dedicated to
my beloved parents.

Acknowledgement

First and foremost, I would like to express my deep gratitude to my thesis supervisors, Mr. Elias Ould Saïd and Mrs. Farida Hamrani, for their kind guidance and constant support. Their valuable advice and availability, even outside conventional hours, have greatly contributed to the progress of my research. It has been a real pleasure working with them, and I hope this collaboration will continue in the future.

I would like to thank Mr. Abdelkader Tatachak for the honor of chairing the jury of this thesis. His presence, despite his numerous responsibilities, means a great deal to me. I also wish to express my profound gratitude to Mrs. Ourida Sadki and Mrs. Zohra Guessoum, as well as Mr. Salim Bouzebda, for accepting to review and evaluate this work.

My thanks also go to all the members of the MSTD laboratory, especially Latifa and Adel, for their invaluable assistance throughout this thesis. A big thank you to my fellow doctoral students, Nour, Dounia, and Yamina, for their unwavering friendship and moral support, which have made these years of study unforgettable.

Finally, my most sincere thanks go to my family, and especially to my father, for his constant and unwavering support throughout this academic journey.

Résumé

Dans cette thèse, nous nous intéressons à l'estimation non paramétrique de la fonction de régression par erreur relative. Dans un premier temps, nous étudions les propriétés asymptotiques d'un estimateur introduit par Jones et al. (2008), dans le cadre de données indépendantes et complètement observées. Par la suite, nous proposons la construction d'un estimateur à noyau adapté aux données tronquées à gauche et censurées à droite. Nous établissons, sur un compact, la convergence uniforme presque sûre avec vitesse ainsi que la normalité asymptotique de cet estimateur en présence d'une dependance de type α-mélange. Les résultats sont illustrés à travers des simulations et des applications pratiques.

Mots clés: α-mélange, convergence uniforme presque sûre, données tronquées-censurées, estimateur à noyau, normalité asymptotique, régression non paramétrique par erreur relative.

Abstract

In this thesis, we focus on the nonparametric estimation of the relative error regression function. Initialy, we study the asymptotic properties of an estimator introduced by Jones et al. (2008) in the context of independent and fully observed data. Subsequently, we propose the construction of a kernel estimator adapted to left truncated and right censored data. We establish, on a compact set, the almost sure uniform convergence with a rate, as well as the asymptotic normality of this estimator under the assumption of α-mixing dependency. The results are illustrated through simulations and practical applications.

Keywords: α-mixing, almost sure uniform convergence, asymtotic normality, kernel estimator, nonparametric relative error regression, truncated-censored data.

Scientific productions

Articles

- Bayarassou, N., Hamrani, F., and Ould Saïd, E. (2024), "Asymptotic normality of a relative regression function estimator for left truncated and right censored data under α-mixing condition ", *Communications in Statistics - Theory and Methods*, Submitted.

- Bayarassou, N., Hamrani, F., and Ould Saïd, E. (2023), "Nonparametric relative error estimation of the regression function for left truncated and right censored time series data ", *Journal of Nonparametric Statistics*, https://doi.org/10.1080/10485252.2023.2241572.

Communications

- Presented at the National Conference: Mathematical Modeling for Dynamic Systems, 26-27 Juin 2024, Constantine, Algeria.

- Presented at the Second National Conference on Pure and Applied Mathematics, 04 July 2023, Tébessa, Algeria.

- Presented at the First International Workshop of Statistics and its Applications, 01-02 March 2023, Saida, Algeria.

- Presented at the Scientific Day of the MSTD Laboratory, 26 February 2022, Alger, Algeria.

Contents

Chapter 1

General introduction

1.1 Regression

In the contemporary world, the ability to extract meaningful insights from data is crucial for making informed decisions and solving complex problems across diverse fields, including medicine, economics, engineering and social sciences. Among the numerous techniques available for data analysis, regression analysis remains a fundamental tool, enabling researchers and practitioners to explore relationships between variables and make predictions.

Regression analysis has a long and rich history. Its roots can be traced back to the work of various pioneers such as Legendre (1805) and Gauss (1809). However, it was Galton (1890) who coined the term "regression" while studying the connection between the heights of parents and their offspring. Since then, regression has evolved and expanded in scope to encompass various techniques and approaches, progressing from simple linear regression to more complex and nonparametric models.

At its core, regression aims to model the relationship between a response variable (denoted by Y) and one or more explanatory variables (denoted by X). Consider a dataset with N data points, each representing a pair of values $(X_i, Y_i)_{i=1,...,N}$ for the variables X and Y. The equation

$$Y_i = m(X_i) + \epsilon_i, \quad i = 1, ..., N,$$

captures this relationship. Here, $m(\cdot)$ represents the unknown regression function and the term ϵ_i signifies the error term associated with the i-th observation.

In social sciences, for instance, regression serves to understand how factors like parental education level X might influence a child's academic performance Y. By analyzing data from different students, researchers can construct regression models to quantify the influence of parental education on academic performance.

The process of building a regression model, specifically estimating the regression function $m(\cdot)$, typically involves two principal stages:

- Choosing the approach: there are two primary approaches to explore the relationship between X and Y: parametric and nonparametric. The parametric approach assumes that the regression function $m(\cdot)$ follows a predefined form, often expressed as a mathematical equation containing a set of unknown parameters. A widely used example is linear regression, where the relationship is modeled as ($Y = \beta_0 + \beta_1 X + \epsilon$), with β_0 and β_1 representing parameters to be estimated. The advantage of the parametric approach lies in its interpretability, as the estimated parameters provide direct insight into the influence of each explanatory variable on the response variable. Additionally, parametric models generally require less data to achieve satisfactory predictive performance compared to their nonparametric counterparts. However, a significant limitation of this approach is its reliance on the assumption that the chosen functional form adequately represents the underlying data. If the data deviate substantially from the assumed model, the predictions may lack reliability.

Conversely, the nonparametric approach does not impose any predefined structure on the regression function $m(\cdot)$, allowing the data to determine the functional form. Common nonparametric techniques include methods such as k-nearest neighbors, which predicts the response variable for a new data point based on the values of its closest neighbors in the training set, and kernel smoothing, which uses a weighted average of nearby data points to estimate the response variable. The strength of nonparametric regression lies in its flexibility and adaptability, particularly when modeling complex or unknown relationships in the data. However, this flexibility often comes at the expense of interpretability, as nonparametric models do not typically offer clear insights into the individual effects of explanatory variables. Moreover, nonparametric methods can be computationally intensive, especially when applied to large datasets.

In recent years, nonparametric regression has garnered substantial attention from researchers due to the growing need to model complex relationships present in real world data. As a result, many works have been published and have contributed to its development in both the theoretical and practical aspects (e.g., Härdle (1990), and Tsybakov (2009)).

- Selecting the estimation method: once the approach has been determined, the next step involves choosing an appropriate method for estimating the regression function $m(\cdot)$. In nonparametric regression, various estimation techniques are available, with the least squares method being the most commonly employed. This method seeks to minimize a loss function, which in this context is the mean squared error (MSE), expressed as:

$$\mathbb{E}\left[(Y - m(X))^2 \,|X\right].$$

Minimizing this MSE provides a solution that corresponds to the conditional expectation of Y given X, denoted as $\mathbb{E}[Y|X]$, representing the average value of Y for a given value of X. Numerous techniques exist for estimating this conditional expectation. Notably, Nadaraya (1964) and Watson (1964) independently introduced a kernel estimator for this purpose. In addition, Révész (1973) proposed a recursive kernel estimator, and Stone (1977) defined a k-nearest neighbors estimator.

While the least squares method is widely used due to its simplicity and efficiency, it has limitations, especially in the presence of outliers. Outliers, which are data points that deviate significantly from the rest of the dataset, can arise from various sources, such as measurement errors (e.g., malfunctioning equipment), data entry mistakes (e.g., incorrect decimal placement), or genuine extreme values within the population being studied. The least squares method is particularly sensitive to outliers, as it squares the residuals (the differences between predicted and observed values), causing outliers with large residuals to have a disproportionate influence on the estimation, thereby leading to inaccurate models that may fail to capture the true relationship between X and Y.

To address this issue, several alternative estimation methods have been developed to mitigate the impact of outliers on the estimation process. One such method is the M-estimation, introduced by Huber (1964), which is considered a robust alternative to the least squares. The M-estimation modifies the MSE loss function as follows:

$$\mathbb{E}\left[\rho\left(Y - m(X)\right)|X\right],$$

where ρ is a function designed to reduce the influence of outliers on the overall estimate. For further exploration of M-estimation, we refer to the works of Härdle (1984), Fan et al. (1994), Boente and Rodriguez (2006), and Azzedine et al. (2008). Another method gaining recognition is the relative error regression (RER). This method focuses on minimizing the mean squared relative error (MSRE) instead of the MSE, defined as:

$$\mathbb{E}\left[\left(\frac{Y - m(X)}{Y}\right)^2 |X\right], \text{ for } Y > 0. \tag{1.1}$$

By minimizing this MSRE, the regression process becomes less susceptible to the influence of outliers. This is because MSRE evaluates the proportional difference between predicted and true values, thus mitigating the impact of extreme deviations on the error measure. Park and Stefanski (1998) showed that the solution of the minimisation problem of (1.1) is explicitly expressed by:

$$m(X) = \frac{\mathbb{E}[Y^{-1}|X]}{\mathbb{E}[Y^{-2}|X]}, \tag{1.2}$$

where the first two conditional inverse moments of Y given X are finite almost surly. The first kernel and local linear estimator of equation (1.2) was introduced by Jones et al. (2008) for the case when the explanatory variable is a real number, while Demongeot et al. (2016) proposed a kernel estimator of the same equation for the functional case.

The selection of an appropriate statistical approach and estimation method critically depends on the characteristics of the data and the specific research objectives.

1.2 Incomplete survival data

Survival data, also known as time-to-event data, comprises observations that record the duration until a specific event of interest occurs. This event could be the death of an individual, the failure of a machine or the end of a contract. A characteristic feature of survival data is its inherent incompleteness, primarily due to censoring and/or truncation. These phenomena are distinct and can manifest in various forms, leading to diverse types of incomplete data.

Censoring arises when the exact timing of the event of interest is unknown, resulting in only partial information being available. The three principal types of censoring are right censoring, left censoring and interval censoring.

1.2.1 Right censoring

Right censoring occurs when the exact value of a lifetime variable Y is unknown, but it is known to be greater than or equal to a specific time W. Right censored data typically falls into the following categories:

1.2.1.1 Type I censoring

In Type I censoring, a study has a predefined observation period, and at the conclusion of this fixed time, some individuals may not have experienced the event of interest. Let $Y_1, \ldots, Y_N$ represent the random variables of interest and W the fixed censoring time. Instead of observing $Y_1, \ldots, Y_N$, the observed data consists of pairs $(Z_1, \delta_1), \ldots, (Z_N, \delta_N)$, where:

$$Z_i = \min(Y_i, W) \text{ and } \delta_i = \mathbb{1}_{\{Y_i \leq W\}}, \text{ for } i = 1, \ldots, N.$$

Example 1.1. *A pharmaceutical company tests the efficacy of a new drug over a fixed period (e.g., six months). At the end of the six-month period, some patients may not have experienced the desired outcome. The event times for these patients are considered Type I censored because the observation period ended before the event of interest occurred.*

1.2.1.2 Type II censoring

In Type II censoring, the study continues until a predetermined number r of failures occurr. Thus, only the r smallest lifetimes out of the N samples are observed, where $(1 \leq r \leq N)$. Let $Y_1, \ldots, Y_N$ represent a random sample of lifetimes. A Type II censored sample is defined by observing $(Z_1, \delta_1), \ldots, (Z_N, \delta_N)$, where:

$$Z_i = \min(Y_i, Y_{\{r\}}) \text{ and } \delta = \mathbb{1}_{\{Y_i \leq Y_{\{r\}}\}}, \text{ for } i = 1, \ldots, N.$$

Here, $Y_{\{r\}}$ denotes the r-th order statistic of the sample $Y_1, \ldots, Y_N$.

Example 1.2. *In a quality control study, the durability of a material is tested by applying stress until a predetermined number of samples (e.g., $r = 10$) fail. Only the lifetimes of the first 10 failures are recorded for analysis, constituting Type II censored data, where only the r smallest lifetimes are observed.*

1.2.1.3 Type III censoring

Type III censoring generalizes Type I censoring by allowing each individual to have its own censoring time. Let $Y_1, \ldots, Y_N$ represent the random variables of interest and $W_1, \ldots, W_N$ the individual censoring times. The observed data consists of pairs $(Z_1, \delta_1), \ldots, (Z_N, \delta_N)$, where:

$$Z_i = \min(Y_i, W_i) \text{ and } \delta = \mathbb{1}_{\{Y_i \leq W_i\}}, \text{ for } i = 1, \ldots, N.$$

Example 1.3. *A medical study tracks the time until a specific symptom appears in patients. Each patient is monitored for a different duration depending on their health condition. For instance, Patient A may be followed for six months, while Patient B is observed for one year. If a patient's follow-up ends before the event of interest occurs, the observation is classified as Type III censored.*

1.2.2 Left censoring

Left censoring occurs when the event of interest has already occurred before an individual enters the study, meaning their true lifetime is known to be less than a specific censoring time W. Let $Y_1, \ldots, Y_N$ represent the random variables of interest and W the left censoring time. Instead of observing $Y_1, \ldots, Y_N$, the observed data consists of pairs $(Z_1, \delta_1), \ldots, (Z_N, \delta_N)$, where:

$$Z_i = \max(Y_i, W) \text{ and } \delta = \mathbb{1}_{\{Y_i \geq W\}}, \text{ for } i = 1, \ldots, N.$$

Example 1.4. *In a study on early childhood learning, researchers may investigate the age at which children learn to accomplish specific tasks, such as tying their shoes. If some children already know how to tie their shoes at the start of the study, their event times are classified as left censored.*

1.2.3 Interval censoring

Interval censoring occurs when the exact time of the event is unknown, but it is known to have occurred within a specific interval (W_a, W_b), where W_a and W_b are the endpoints of the censoring interval. This type of censoring is common in clinical trials or longitudinal studies with periodic follow-up.

Example 1.5. *A study monitoring the hatching time of turtle eggs provides an example of interval censoring. Researchers visit nesting sites periodically, as they cannot observe the eggs continuously. If an egg hatches between two visits, the exact time of hatching is unknown, but it is known to have occurred between the two visit times.*

Truncation, on the other hand, arises when certain observations are systematically excluded from a study based on predefined criteria. The two main types of truncation are left truncation and right truncation.

1.2.4 Left truncation

In the case of left truncation, only individuals whose event of interest occurs after a specified time T are included in the analysis. Events that occur prior to this time are not observed, and thus, no information is available for these cases.

Example 1.6. *A study aiming to estimate the size distribution of microscopic particles may face left truncation. Due to the limited resolution of the microscope, only particles exceeding a certain size are detectable. Consequently, smaller particles, which may also be relevant to the analysis, are excluded from the dataset, leading to left truncation.*

1.2.5 Right truncation

Right truncation, in contrast to left truncation, occurs when only individuals whose event times fall below a specific time T are included in the study. Events occurring after this time are excluded from the analysis.

Example 1.7. *In astronomical research, studies investigating the lifespans of stars often encounter right truncation. Stars located beyond a certain distance are not detectable due to the limitations of observational instruments. As a result, the dataset is right truncated, as only stars within the observable range are included in the analysis.*

In various real world contexts, data may exhibit both left truncation and right censoring, giving rise to a more complex form of incomplete data known as left truncated and right censored (LTRC) data.

1.2.6 Left truncation and right censoring

LTRC data arises when both left truncation and right censoring are present in a study. Let $Y_1, \ldots, Y_N$ represent the lifetime variables, $W_1, \ldots, W_N$ the censoring times, and $T_1, \ldots, T_N$ the truncation times. Instead of observing the full lifetimes $Y_1, \ldots, Y_N$, the observed data consists of the triplet (Z_i, T_i, δ_i), where:

$$Z_i = \min(Y_i, W_i) \text{ and } \delta_i = \mathbb{1}_{\{Y_i \leq W_i\}}, \text{ for } i = 1, \ldots, n.$$

Here, δ is an indicator function, with $\delta = 1$ if the lifetime is observed and $\delta = 0$ if it is censored. Observations are only recorded for cases where $Z_i \geq T_i$, and if $Z_i < T_i$, no observation is made.

Example 1.8. *A study examining the longevity of vintage cars registered in a specific city provides an example of LTRC data. Left truncation occurs because only cars that remained operational beyond a certain age (e.g., 25 years) are included in the study. Cars that were no longer functional before reaching this age are excluded. Right censoring arises when the study concludes before some cars have ceased operation, leaving their exact lifespans unknown.*

Example 1.9. *A study of mortality in a retirement community also exemplifies LTRC data. Left truncation is present by including only individuals who joined the community after reaching a certain age (e.g., 55 years), excluding those who died prior to reaching that age. Right censoring arises for residents who are still alive at the conclusion of the study, as their survival times remain unknown.*

Several researchers have explored the estimation of the regression functionin the presence of incomplete data. For right censored data, Carbonez et al. (1995) introduced a partition-based estimator and demonstrated its strong consistency. Kohler et al. (2002) proved the strong consistency of various estimators, including kernel, k-nearest neighbors, least squares and smoothing splines. Guessoum and Ould Saïd (2009) investigated a kernel estimator, establishing its pointwise and uniform strong consistency with a rate over a compact set, as well as its asymptotic normality. Lemdani and Ould Saïd (2017) defined an M-estimator and studied its asymptotic properties for real-valued explanatory variables, later extended to the functional case by Ait Hennani et al. (2019). Bouhadjera et al. (2022) constructed a local linear regression estimator that minimizes the sum of squared relative errors and obtained its uniform almost sure consistency with a rate over a compact set.

For left truncated data. Ould Saïd and Lemdani (2006) proposed a new kernel estimator and investigated its asymtotic properties, including pointwise consistency, uniform strong convergence and asymptotic normality. In the infinite-dimensional case, Derrar et al. (2015) introduced an M-estimator and established its almost complete convergence with a rate and asymptotic normality. Altendji et al. (2018) demonstrated the almost sure consistency and the asymptotic normality of a kernel relative regression estimator. Bouabsa (2021) obtained a uniform consistency with a rate of a k-nearest neighbors relative regression estimator. Most recently, Bouhadjera et al. (2023) proposed a local linear relative error regression estimator for a real-valued explanatory variable and proved its strong uniform consistency.

For LTRC data, Cao et al. (2005) proposed a relative hazard rate estimator, establishing its asymptotic distribution and asymptotic mean squared error. Molanes-López and Cao (2008) defined a kernel estimator for the relative density, obtaining its bias, variance and limit distribution.

1.3 Dependance

All the aforementioned works on incomplete data have assumed independence between observations. However, this assumption may not always hold in practice, especially in contexts such as the retirement community study. Participants in such a setting often share similar characteristics due to the nature of the community, leading to potential dependence in the data. Consequently, recent research studies have turned their attention toward modeling dependent data, introducing concepts like association and mixing to capture this dependence.

The concept of association, proposed by Esary et al. (1967) for reliability applications, represents a particular case of the weak dependence formalized by Doukhan and Louhichi (1999). Mathematically, a finite sequence of random variables $\{U_i; 1, \dots, n\}$ is considered associated if:

$$\mathrm{Cov}\left(f(U_1, \dots, U_n), g(U_1, \dots, U_n)\right) \geq 0,$$

for all non-decreasing function f and g from $\mathbb{R}^n \to \mathbb{R}$, where the covariance exists. An infinite sequence of random variables $\{U_i;\ i \geq 1\}$ is said to be associated if every finite subsequence is associated. For a more detailed discussion on association and its applications, we refer the reader to Cai and Roussas (1998), Masry (2002), Bulinski and Shashkin (2007), Guessoum et al. (2012), Menni and Tatachak (2016), Hamrani and Guessoum (2017), Adjoudj and Tatachak (2018), and Bey et al. (2022).

Mixing is another approach to describing weak dependence between random variables. It characterizes the gradual weakening of dependence as the time lag between variables increases. There are various types of mixing conditions, with α-mixing (strong mixing) being the weakest and therefore the least restrictive. The strong mixing condition is defined as follows:

Definition 1.1. *A sequence $\{U_i;\ i \geq 1\}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be α-mixing if the mixing* *coefficient:*

$$\alpha(n) = \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)| : A \in \mathcal{F}_1^k \text{ and } B \in \mathcal{F}_{k+n}^{\infty},\ k \in \mathbb{N}^*\}$$

converges to zero as $n \to \infty$, *where* $\mathcal{F}_l^m$ *denotes the* σ*-algebra generated by* $\{U_j;\ l \le j \le m\}$. *(See, Doukhan (1994)).*

There are two primary subclasses of strong mixing sequences:

Definition 1.2. *The sequence* $\{U_i;\ i \ge 1\}$ *is said to be arithmetically* $\alpha-$*mixing with rate* $a > 0$ *if*

$$\exists C > 0, \alpha(n) \le C n^{-a}.$$

It is called geometrically α*-mixing if*

$$\exists C > 0, \exists t \in (0,1), \alpha(n) \le C t^n.$$

Strong mixing can also be defined for multiple sequences of random variables:

Proposition 1.1. *Let* $\{U_i, i \ge 1\}$ *and* $\{V_i, i \ge 1\}$ *be two sequences of* α*-mixing random variables with mixing coefficient* $\alpha_1(\cdot)$ *and* $\alpha_2(\cdot)$*, respectively, then* $\{(U_i, V_i), i \ge 1\}$ *is a sequence of* α*-mixing random variables with mixing coefficient* $4\alpha(n)$*, where* $\alpha(n) = max\{\alpha_1(n), \alpha_2(n)\}$. *(See, Cai (2001))*

Several important inequalities are frequently used when dealing with α-mixing sequences:

Proposition 1.2 (Fuk-Nagaev's inequality)**.** *Let* $\{U_i, i \ge 1\}$ *be a sequence of random variables, with strong mixing coefficient* $\alpha(n) = O(n^{-v})$*, where* $v > 1$. *If there exists* $M < \infty$ *such that* $|U_1| \le M$*, then we have for any* $\varepsilon > 0$*,* $q > 1$ *and for some* $C < \infty$*:*

$$\boldsymbol{P}\left(\left|\sum_{i=1}^{n} U_i\right| > \varepsilon\right) \le C\left\{\left(1 + \frac{\varepsilon^2}{q S_n^2}\right)^{-\frac{q}{2}} + n q^{-1}\left(\frac{q}{\varepsilon}\right)^{v+1}\right\},$$

where $S_n^2 = \displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} |Cov(U_i, U_j)|$. *(See, Rio (2000)).*

Proposition 1.3 (Davydov covariance inequality)**.** *Let* $(U_n)_{n\in\mathbb{Z}}$ *be a sequence of stationary* α*-mixing random variables. For some* $k \in \mathbb{Z}$*, consider a real variable* U *(resp.* U'*) which is* $\mathcal{F}_{-\infty}^{k}$*-measurable (resp.* $\mathcal{F}_{n+k}^{+\infty}$*-measurable). If for some positive numbers* p, q, r *such that* $p^{-1} + q^{-1} + r^{-1} = 1$*, we have* $\mathbb{E}(X)^p < \infty$ *and* $\mathbb{E}(X')^p < \infty$*, then*

$$\exists C, 0 < C < +\infty,\ Cov(U, U') \le C(\mathbb{E}(U)^p)^{\frac{1}{p}}(\mathbb{E}(U')^q)^{\frac{1}{q}}\alpha(n)^{\frac{1}{r}},$$

where the σ*-algebra* $\mathcal{F}_a^b = \sigma\{U_i, a \le i \le b\}$. *(See, Rio (2000)).*

Proposition 1.4 (Volkonskii and Rozanov covariance inequality)**.** *Consider a set of* σ*-algebras of events* $\mathcal{F}_i^j$*, where* $-\infty < i < j < \infty$*, and let* $U_1, \ldots, U_n$ *be* α*-mixing random variables satisfying the condition* $|X_k| \le 1$. *Assume that each* U_k *is measurable with respect to the* σ*-algebra* $\mathcal{F}_{i_k}^{j_k}$*, where* $i_1 < j_1 < \ldots < i_n < j_n$*,* $i_{k+1} - j_k \ge w \ge 1$. *Then,*

$$\left| \left(\mathbb{E} \prod_{j=1}^{n} U_j \right) - \prod_{j=1}^{n} \mathbb{E}[U_j] \right| \leq 16(n-1)\alpha(w).$$

(See, Volkonskii and Rozanov (1959)).

The study of incomplete data under the strong mixing condition is a relatively recent field, with limited results on nonparametric regression estimation. For left truncated data, Liang et al. (2009) established the uniform strong convergence with a rate of a kernel estimator. Wang and Liang (2012) constructed an M-estimator and demonstrated its weak and strong consistency, along with asymptotic normality. For right censored data, Guessoum and Ould Saïd (2012) proved the convergence in probability of a kernel estimator and its asymtotic normality. Bouhadjera and Ould Saïd (2019) obtained the uniform almost sure convergence with a rate of the relative error regression estimator. For LTRC data, to the best of our knowledge, the only contribution is by Benseradj and Guessoum (2022). They proposed a robust M-estimator and established its strong uniform consistency with a rate.

1.4 Organization of the thesis

The primary motivation for this thesis is to address the nonparametric estimation of the relative error regression function for left truncated and right censored data under the assumption of α-mixing dependency. The manuscript is structured into three main chapters, in addition to an introduction and a conclusion that includes future research directions.

Chapter 2 focuses on the asymptotic properties of a relative regression function estimator proposed by Jones et al. (2008) in the context of complete data. Under the assumption of independent and identically distributed observations, we establish the almost sure uniform convergence with a rate and asymptotic normality of the estimator. These results are illustrated through a simulation study.

Chapter 3 introduces a nonparametric estimator for the regression function associated with LTRC data. The estimator is constructed by minimizing the mean squared relative error. We demonstrate its strong uniform convergence with a rate under α-mixing condition and evaluate its performance via an extensive simulation study. A real world case study further underscores its practical applicability.

Chapter 4 extends the work of Chapter 3 by deriving the asymptotic normality of the proposed estimator. This enables the construction of confidence intervals. Simulation studies and a real world data application validate the theoretical findings.

It is worth noting that the work presented in Chapter 3 has been published as a journal article, while the work in Chapter 4 is currently under review.

Chapter 2

Asymptotic properties of the relative regression function estimator: complete and iid data

In this chapter, we investigate the asymptotic properties of a relative regression function estimator introduced by Jones et al. (2008) in the context of complete data. Specifically, we establish the almost sure uniform convergence and the asymptotic normality of the estimator under the assumption that the observations are independent and identically distributed (iid). Furthermore, we conduct a simulation study to illustrate and validate these theoretical results.

2.1 Introduction

In a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, consider a sequence of iid random vectors $\{(\mathbf{X}_i, Y_i);\ i = 1, \ldots, N\}$ sampled from $(\mathbf{X}, Y)$, taking values in $\mathbb{R}^d \times \mathbb{R}_+^*$. The primary objective in studying the relationship between these two variables is often to predict one variable given the other. One widely used method for addressing this prediction problem is regression based on the mean squared error (MSE) criterion. However, this method can be sensitive to outliers, which may significantly influence the results. To overcome this issue, Park and Stefanski (1998) proposed the relative error regression approach, which minimizes the mean squared relative error (MSRE), defined as:

$$\mathbb{E}\left[\left(\frac{Y - m(\mathbf{x})}{Y}\right)^2 \mid \mathbf{X} = \mathbf{x}\right], \text{ for } Y > 0.$$

Minimizing the MSRE leads to the following expression for the regression function

$$m(\mathbf{x}) = \frac{\mathbb{E}[Y^{-1} \mid \mathbf{X} = \mathbf{x}]}{\mathbb{E}[Y^{-2} \mid \mathbf{X} = \mathbf{x}]}.$$

To substantiate this result, we can expand the MSRE as follows

$$MSRE = \mathbb{E}\left[\frac{(Y - m(\mathbf{x}))^2}{Y^2} \mid \mathbf{X} = \mathbf{x}\right] = \mathbb{E}\left[\frac{Y^2 - 2Ym(\mathbf{x}) + m(\mathbf{x})^2}{Y^2} \mid \mathbf{X} = \mathbf{x}\right].$$

Upon simplifying this expression, we obtain

$$MSRE = 1 - 2m(\mathbf{x})\mathbb{E}[Y^{-1} \mid \mathbf{X} = \mathbf{x}] + m(\mathbf{x})^2\mathbb{E}[Y^{-2} \mid \mathbf{X} = \mathbf{x}].$$

By differentiating the above expression with respect to $m(\mathbf{x})$ and set the derivative equal to zero, we find

$$\frac{\partial}{\partial m(\mathbf{x})} MSRE = -2\mathbb{E}[Y^{-1} \mid \mathbf{X} = \mathbf{x}] + 2m(\mathbf{x})\mathbb{E}[Y^{-2} \mid \mathbf{X} = \mathbf{x}] = 0.$$

Solving for $m(\mathbf{x})$, we arrive at the desired expression for the regression function. In this context, Jones et al. (2008) introduced a nonparametric estimator for $m(\mathbf{x})$, given by

$$\widehat{m}_N(\mathbf{x}) =: \frac{\widehat{\psi}_{N,1}(\mathbf{x})}{\widehat{\psi}_{N,2}(\mathbf{x})},$$

with

$$\widehat{\psi}_{N,\ell} = \frac{1}{Nh_N^d} \sum_{i=1}^{N} Y_i^{-\ell} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_N}\right), \text{ for } \ell = 1, 2.$$

Here, $K_d : \mathbb{R}^d \to \mathbb{R}$ is a kernel function and h_N is a sequence of positive real numbers that approaches zero as $N \to \infty$. The formula for $\widehat{m}_N(\cdot)$ resembles the Nadaraya-Watson estimator developed by Nadaraya (1964) and Watson (1964) in the MSE setting.

The remainder of this chapter is organized as follows: Section 2.2 outlines the assumptions and presents the theoretical results. Section 2.3 discusses the simulation study. Section 2.4 provides auxiliary results and proofs.

2.2 Assumptions and main results

To formulate the asymptotic properties of the estimator $\widehat{m}_N(\cdot)$, we first introduce the following notation

$$\psi_\lambda(\mathbf{u}) = \int_{\mathbb{R}_+^*} t^{-\lambda} f_{\mathbf{X},Y}(\mathbf{u}, t)dt, \text{ for } \lambda = 1, 2, 3, 4.$$

Let $\mathcal{C}$ denote a compact set in $\mathbb{R}^d$, and assume that $\inf_{\mathbf{x}\in\mathcal{C}} \widehat{\psi}_{N,2}(\mathbf{x}) > 0$. Throughout this chapter, when no confusion is possible, C will be used to denote any generic constant.

2.2.1 Assumptions

K. The kernel $K_d(\cdot)$ is a bounded probability density with compact support and satisfies:

(a) $\int_{\mathbb{R}^d} K_d(\mathbf{r})\mathbf{dr} = 1$, $\int_{\mathbb{R}^d} r_i K_d(\mathbf{r})\mathbf{dr} = 0$ and $\int_{\mathbb{R}^d} |r_i||r_j| K_d(\mathbf{r})\mathbf{dr} < \infty$, for $i, j = 1, \ldots, d$.

(b) $\int_{\mathbb{R}^d} K_d^2(\mathbf{r})\mathbf{dr} < \infty$ and $\int_{\mathbb{R}^d} r_i K_d^2(\mathbf{r})\mathbf{dr} < \infty$, for $i = 1, \ldots, d$.

(c) $\forall(\mathbf{t}, \mathbf{s}) \in \mathcal{C}^2$ $|K_d(\mathbf{t}) - K_d(\mathbf{s})| \leq \| \mathbf{t} - \mathbf{s} \|^\gamma$, for $\gamma > 0$.

H. The bandwidth h_N satisfies:

(a) $\lim_{N\to\infty} h_N = 0$, $\lim_{N\to\infty} N h_N^d = \infty$ and $\lim_{N\to\infty} \frac{\log N}{N h_N^d} = 0$.

(b) $\lim_{N\to\infty} N h_N^{d+4} = 0$.

D. (a) The function $\psi_\lambda(\cdot)$ is twice continuously differentiable and $\sup_{\mathbf{x}\in\mathcal{C}} \left| \frac{\partial^2 \psi_\lambda(\mathbf{x})}{\partial x_i \partial x_j} \right| < \infty$, for $\lambda = 1, 2, 3, 4$ and $i, j = 1, \ldots, d$.

(b) $\forall Y > 0$, $\exists C > 0$, such that $Y^{-\ell} \leq C$, for $\ell = 1, 2$.

2.2.1.1 Comments on the Assumptions

Assumptions **K**, **H**(a), and **D**(a) are typical in nonparametric regression estimation. Assumption **D**(b) is particularly useful for establishing consistency. Assumption **H**(b) is a technical condition necessary for the study of asymptotic normality.

2.2.2 Main results

The first theorem establishes a strong uniform consistency result with a rate over a compact set for the estimator $\widehat{m}_N(\cdot)$.

Theorem 2.1 (Strong uniform consistency)**.** *Under Assumptions **K** and **D**, we have*

$$\sup_{\boldsymbol{x}\in\mathcal{C}} |\widehat{m}_N(\boldsymbol{x}) - m(\boldsymbol{x})| = O(h_N^2) + O\left(\sqrt{\frac{\log N}{N h_N^d}}\right) \quad a.s. \text{ as } N \to \infty.$$

The seond theorem focuses on the asymtotic normality of the estimator $\widehat{m}_N(\cdot)$.

Theorem 2.2 (Asymptotic normality)**.** *Under Assumptions **K**(a), **K**(b), **H**, **D**(a), we have :*

$$\sqrt{N h_N^d}\,(\widehat{m}_N(\boldsymbol{x}) - m(\boldsymbol{x})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_N^2(\boldsymbol{x})) \quad as\ N \to \infty,$$

where $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution and*

$$\sigma_N^2(\boldsymbol{x}) = \frac{\psi_2^3(\boldsymbol{x}) - 2\psi_1(\boldsymbol{x})\psi_2(\boldsymbol{x})\psi_3(\boldsymbol{x}) + \psi_1^2(\boldsymbol{x})\psi_4(\boldsymbol{x})}{\psi_2^4(\boldsymbol{x})}\kappa,$$

with $\kappa = \int_{\mathbb{R}^d} K_d^2(\boldsymbol{s})d\boldsymbol{s}$.

Remark 2.1. *A computable estimator* $\widehat{\sigma}_N^2(\boldsymbol{x})$ *for the asymptotic variance* $\sigma_N^2(\boldsymbol{x})$ *can be easily obtained using a plug-in method by replacing* $\psi_\lambda(\cdot)$ *with its estimator* $\widehat{\psi}_{N,\lambda}(\cdot)$ *for* $\lambda = 1, 2, 3, 4$*:*

$$\widehat{\psi}_{N,\lambda}(\boldsymbol{x}) = \frac{1}{Nh_N^d}\sum_{i=1}^{N} Y_i^{-\lambda} K_d\left(\frac{\boldsymbol{x} - X_i}{h_N}\right).$$

As a direct consequence of Theorem 2.2, we derive the following confidence interval.

Corollary 2.1. *Under the assumptions of Theorem 2.2, and for each fixed* $\zeta \in (0,1)$*, a confidence interval of level* $(1-\zeta)$ *for* $m(\boldsymbol{x})$ *is given by*

$$\left[\widehat{m}_N(\boldsymbol{x}) - t_{1-\frac{\zeta}{2}}\frac{\widehat{\sigma}_N(\boldsymbol{x})}{\sqrt{Nh_N^d}}, \widehat{m}_N(\boldsymbol{x}) + t_{1-\frac{\zeta}{2}}\frac{\widehat{\sigma}_N(\boldsymbol{x})}{\sqrt{Nh_N^d}}\right],$$

where $t_{1-\frac{\zeta}{2}}$ *denotes the* $\left(1-\frac{\zeta}{2}\right)$*-quantile of the standard normal distribution.*

2.3 Simulation study

The goal of this section is to assess the performance of the estimator $\widehat{m}_N(\cdot)$ in the context of a one-dimensional covariante $\mathbf{X}$ (i.e., d=1). To achieve this, we simulate N observations of the pairs $\{(X_i, Y_i); i = 1, \ldots, N\}$, where $X_i \rightsquigarrow \mathcal{N}(3,1)$ and $Y_i = 2X_i + \epsilon_i$, with $\epsilon_i \rightsquigarrow \mathcal{N}(0, 0.1)$. Then, we calculate the estimator $\widehat{m}_N(x)$ for $x \in \mathcal{C} = [1,4]$, using a Gaussian kernel and the bandwidth that minimizes the global mean squared error (GMSE):

$$GMSE = \frac{1}{A.B}\sum_{j=1}^{B}\sum_{i=1}^{A}(\widehat{m}_{N,j}(x_i) - m(x_i))^2,$$

for B = 50 replications, where $\widehat{m}_{N,j}(\cdot)$ represents the estimator at iteration j and A is the number of equidistant points x_i within the compact set $\mathcal{C}$.

2.3.1 Consistency

To evaluate the consistency of the estimator, we plot $\widehat{m}_N(x)$ alongside the true function $m(x) = 2x$.

- Sample size effect: As demonstrated in Figure 2.1, the accuracy of the estimator improves with increasing sample size.
- Outliers effect: Figure 2.2 illustrates the robustness of the estimator in the presence of outliers, introduced by multiplying 4 % of each sample by a multiplier factor (MF).

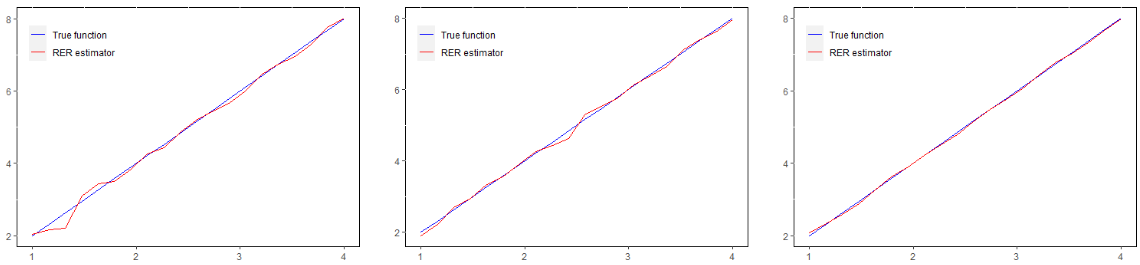

Figure 2.1: $m(\cdot)$ and $\widehat{m}_N(\cdot)$ with N=50,100, and 300 respectively.

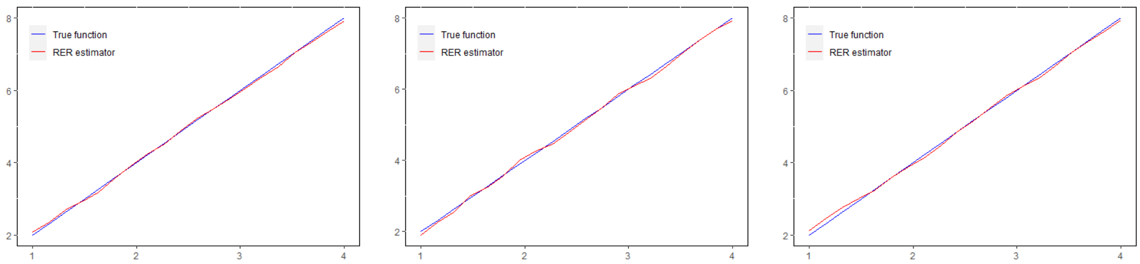

Figure 2.2: $m(\cdot)$ and $\widehat{m}_N(\cdot)$ with N=300, MF=50, 100, and 150 respectively.

2.3.2 Asymptotic normality

To examine the asymptotic normality of the estimator, we compare the shape of the estimated density with that of a standard normal density. We begin by estimating the regression function $m(x) = 2x$ using $\widehat{m}_N(x)$, and subsequently compute the normalized deviation between this estimator and the theoretical regression function at $x = 2$ as follows:

$$m_N^\star = m_N^\star(2) := \frac{\sqrt{Nh_N}}{\widehat{\sigma}_N(2)}(\widehat{m}_N(2) - m(2)).$$

Following this approach, we generate a sequence of size $B = 100$. We then estimate its density function using the kernel method with a Gaussian kernel and the bandwidth $h_B = C \cdot B^{-\frac{1}{5}}$ (as proposed by Silverman (1986)), where C is an appropriately chosen constant. The estimated density $\widehat{m}_N^\star$ is plotted alongside the standard normal density for various values of N. As shown in Figure 2.3, the two distributions become closer as the sample size increases.

Next, we construct confidence intervals for $m(x)$ at a level of 95 %, following Corollary 2.1:

$$\widehat{m}_N(x) \pm 1.96 \times \frac{\widehat{\sigma}_N(x)}{\sqrt{Nh_N^d}},$$

applying the same data and methodology used in computing the estimator $\widehat{m}_N(x)$. Figure 2.4 highlights the reduction in confidence interval width as the sample size increases.

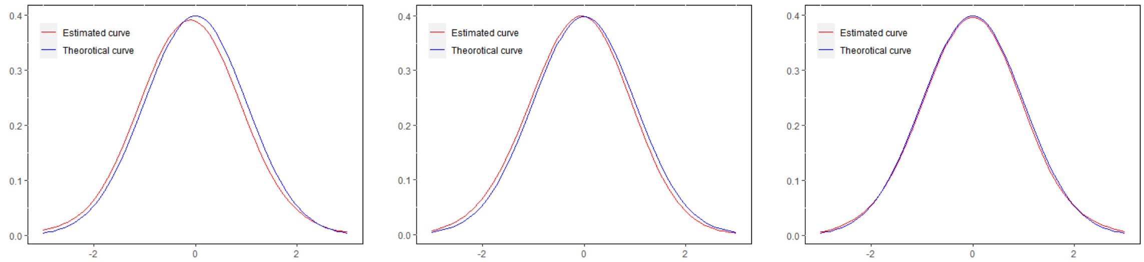

Figure 2.3: N=50,100, and 300 respectively.

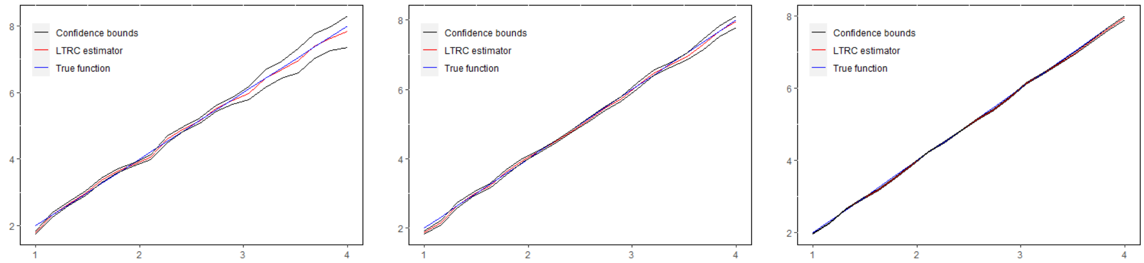

Figure 2.4: N=50,100, and 300 respectively.

2.4 Auxiliary results and proofs

2.4.1 Proofs for consistency result

A classical decomposition allows us to express the difference as follows

$$\begin{aligned}
\widehat{m}_N(\mathbf{x}) - m(\mathbf{x}) &= \frac{\widehat{\psi}_{N,1}(\mathbf{x})}{\widehat{\psi}_{N,2}(\mathbf{x})} - \frac{\psi_1(\mathbf{x})}{\psi_2(\mathbf{x})} \\
&= \frac{1}{\widehat{\psi}_{N,2}(\mathbf{x})}\left\{\left[\widehat{\psi}_{N,1}(\mathbf{x}) - \psi_1(\mathbf{x})\right] - m(\mathbf{x})\left[\widehat{\psi}_{N,2}(\mathbf{x}) - \psi_2(\mathbf{x})\right]\right\} \\
&= \frac{1}{\widehat{\psi}_{N,2}(\mathbf{x})}\left\{\left[(\widehat{\psi}_{N,1}(\mathbf{x}) - \mathbb{E}[\widehat{\psi}_{N,1}(\mathbf{x})]) + (\mathbb{E}[\widehat{\psi}_{N,1}(\mathbf{x})] - \psi_1(\mathbf{x}))\right]\right. \\
&\quad \left. -m(\mathbf{x})\left[(\widehat{\psi}_{N,2}(\mathbf{x}) - \mathbb{E}[\widehat{\psi}_{N,2}(\mathbf{x})]) + (\mathbf{E}[\widehat{\psi}_{N,2}(\mathbf{x})] - \psi_2(\mathbf{x}))\right]\right\} \\
&=: \frac{1}{\widehat{\psi}_{N,2}(\mathbf{x})}\left\{[\Xi_{1,1}(\mathbf{x}) + \Xi_{2,1}(\mathbf{x})] - m(\mathbf{x})\left[\Xi_{1,2}(\mathbf{x}) + \Xi_{2,2}(\mathbf{x})\right]\right\}. \qquad (2.1)
\end{aligned}$$

Applying the triangle inequality to (2.1), we obtain

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{C}} |\widehat{m}_n(\mathbf{x}) - m(\mathbf{x})| \leq\ & \frac{1}{\inf_{\mathbf{x}\in\mathcal{C}} |\widehat{\psi}_{N,2}(\mathbf{x})|}\left\{\sup_{\mathbf{x}\in\mathcal{C}}\left[|\Xi_{1,1}(\mathbf{x})| + |\Xi_{2,1}(\mathbf{x})|\right]\right. \\
& \left. + \sup_{\mathbf{x}\in\mathcal{C}} |m(\mathbf{x})|\left[|\Xi_{1,2}(\mathbf{x})| + |\Xi_{2,2}(\mathbf{x})|\right]\right\}.
\end{aligned}$$

The proof of Theorem 2.1 follows by using Lemma 2.1 and Lemma 2.2 below, which handle the terms $\Xi_{2,\ell}$ and $\Xi_{1,\ell}$, respectively, for $\ell = 1, 2$.

Lemma 2.1. *Under Assumptions* $\boldsymbol{K}(a)$ *and* $\boldsymbol{D}(a)$*, we have for* $\ell = 1, 2$

$$\sup_{\boldsymbol{x}\in\mathcal{C}} |\Xi_{2,\ell}(\boldsymbol{x})| = O_{a.s.}(h_N^2) \text{ as } N \to \infty.$$

Proof. Using a change of variable and a Taylor expansion, we derive

$$\begin{aligned}
\Xi_{2,\ell}(\mathbf{x}) &= \mathbb{E}\left[\frac{1}{Nh_N^d}\sum_{i=1}^{N} Y_i^{-\ell} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right)\right] - \psi_\ell(\mathbf{x}) \\
&= \mathbb{E}\left[\frac{1}{h_N^d} Y_1^{-\ell} K_d\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_N}\right)\right] - \psi_\ell(\mathbf{x}) \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}_+^*} \frac{t^{-\ell}}{h_N^d} K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_N}\right) f_{\mathbf{X},Y}(\mathbf{u},t)\mathbf{du}dt - \psi_\ell(\mathbf{x}) \\
&= \int_{\mathbb{R}^d} \frac{1}{h_N^d} K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_N}\right) \psi_\ell(\mathbf{u})\mathbf{du} - \psi_\ell(\mathbf{x}) \\
&= \int_{\mathbb{R}^d} K_d(\mathbf{r})\psi_\ell(\mathbf{x}-\mathbf{r}h_N)\mathbf{dr} - \psi_\ell(\mathbf{x}) \\
&= \int_{\mathbb{R}^d} K_d(\mathbf{r})\left[\psi_\ell(\mathbf{x}-\mathbf{r}h_N) - \psi_\ell(\mathbf{x})\right]\mathbf{dr} \\
&= \int_{\mathbb{R}^d} K_d(\mathbf{r})\left[-h_N\sum_{i=1}^{d} r_i\frac{\partial\psi_\ell(\mathbf{x})}{\partial x_i} + \frac{h_N^2}{2}\left\{\sum_{i=1}^{d} r_i^2\frac{\partial^2\psi_\ell(\mathbf{x}_0)}{\partial x_i^2}\right.\right. \\
&\quad \left.\left. +2\sum_{i=1}^{d}\sum_{j=1}^{d} r_ir_j\frac{\partial^2\psi_\ell(\mathbf{x}_0)}{\partial x_i\partial x_j}\right\}\right]\mathbf{dr} \\
&\le \frac{h_N^2}{2}\left\{\sum_{i=1}^{d}\sup_{\mathbf{x}\in\mathcal{C}}\left|\frac{\partial^2\psi_\ell(\mathbf{x}_0)}{\partial x_i^2}\right|\int_{\mathbb{R}^d} r_i^2 K_d(\mathbf{r})\mathbf{dr}\right. \\
&\quad \left. +2\sum_{i=1}^{d}\sum_{j=1}^{d}\sup_{\mathbf{x}\in\mathcal{C}}\left|\frac{\partial^2\psi_\ell(\mathbf{x}_0)}{\partial x_i\partial x_j}\right|\int_{\mathbb{R}^d}|r_i||r_j|K_d(\mathbf{r})\mathbf{dr}\right\}.
\end{aligned}$$

Thus, under Assumptions $\mathbf{K}$(a) and $\mathbf{D}$(a), we conclude that

$$\sup_{\mathbf{x}\in\mathcal{C}} |\Xi_{2,\ell}(\mathbf{x})| = O(h_N^2).$$

This completes the proof. □

Lemma 2.2. *Under Assumptions* $\boldsymbol{K}(a)$*-*$\boldsymbol{K}(c)$*,* $\boldsymbol{D}(a)$ *and* $\boldsymbol{D}(b)$*, we have for* $\ell = 1, 2$

$$\sup_{\boldsymbol{x}\in\mathcal{C}} |\Xi_{1,\ell}(\boldsymbol{x})| = O_{a.s.}\left(\sqrt{\frac{\log N}{Nh_N^d}}\right) \text{ as } N \to \infty.$$

Proof. Given that $\mathcal{C}$ is a compact set, it can be covered by a finite number ω_N of balls $\mathcal{B}_k(\mathbf{x}_k, a_N^d)$ centred at $\mathbf{x}_k = (x_{1,k}, \dots, x_{d,k})$, $1 \le k \le \omega_N$, where ω_N and a_N^d satisfy the relations:

$$\omega_N \le M a_N^{-d} \text{ and } a_N^d = h_N^{d(1+\frac{1}{2\gamma})} N^{-\frac{1}{2\gamma}},$$

with M is a positive constant and γ represents the Lipschitz condition stated in Assumption $\mathbf{K}$(c). Then, for any $\mathbf{x} \in \mathcal{C}$, there exists a ball B_k containing $\mathbf{x}$ such that

$$\| \mathbf{x} - \mathbf{x}_k \| \leq a_N^d. \tag{2.2}$$

For $\ell = 1, 2$, we have

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{C}} |\Xi_{1,\ell}(\mathbf{x})| &= \sup_{\mathbf{x}\in\mathcal{C}} \Big| (\widehat{\psi}_{N,\ell}(\mathbf{x}) - \widehat{\psi}_{N,\ell}(\mathbf{x}_k)) + (\mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x}_k)] - \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x})]) \\
&\quad + (\widehat{\psi}_{N,\ell}(\mathbf{x}_k) - \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x}_k)]) \Big| \\
&\leq \max_{1\leq k\leq\omega_N} \sup_{\mathbf{x}\in\mathcal{C}} \left| \widehat{\psi}_{N,\ell}(\mathbf{x}) - \widehat{\psi}_{N,\ell}(\mathbf{x}_k) \right| + \max_{1\leq k\leq\omega_N} \sup_{\mathbf{x}\in\mathcal{C}} \left| \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x}_k)] - \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x})] \right| \\
&\quad + \max_{1\leq k\leq\omega_N} \left| \widehat{\psi}_{N,\ell}(\mathbf{x}_k) - \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x}_k)] \right| \\
&=: I_1 + I_2 + I_3.
\end{aligned}$$

We begin by addressing the term I_1. Under Assumptions $\mathbf{D}$(b) and $\mathbf{K}$(c), we obtain

$$\begin{aligned}
\left| \widehat{\psi}_{N,\ell}(\mathbf{x}) - \widehat{\psi}_{N,\ell}(\mathbf{x}_k) \right| &= \left| \frac{1}{Nh_N^d} \sum_{i=1}^{N} Y_i^{-\ell} \left[K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right) - K_d\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_N}\right) \right] \right| \\
&\leq \frac{C}{h_N^d} \left\| \frac{\mathbf{x}-\mathbf{X}_1}{h_N} - \frac{\mathbf{x}_k-\mathbf{X}_1}{h_N} \right\|^{\gamma} \\
&\leq \frac{C \| \mathbf{x}-\mathbf{x}_k \|^{\gamma}}{h_N^{d+\gamma}}.
\end{aligned}$$

Then, by using (2.2), we get

$$\sup_{\mathbf{x}\in\mathcal{C}} \left| \widehat{\psi}_{N,\ell}(\mathbf{x}) - \widehat{\psi}_{N,\ell}(\mathbf{x}_k) \right| \leq \frac{C a_N^{d\gamma}}{h_N^{d+\gamma}} = \frac{C}{\sqrt{Nh_N^d}} h_N^{\gamma(d-1)}.$$

Therefore

$$I_1 = \max_{1\leq k\leq\omega_N} \sup_{\mathbf{x}\in\mathcal{C}} \left| \widehat{\psi}_{N,\ell}(\mathbf{x}) - \widehat{\psi}_{N,\ell}(\mathbf{x}_k) \right| = O\left(\frac{1}{\sqrt{nh_N^d}} \right).$$

Similarly, for the second term I_2, we derive

$$I_2 = \max_{1\leq k\leq\omega_N} \sup_{\mathbf{x}\in\mathcal{C}} \left| \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x})] - \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x}_k)] \right| = O\left(\frac{1}{\sqrt{Nh_N^d}} \right).$$

For I_3, we appy Bernstein's inequality. This inequality asserts that for a sequence of independent random variables $\{\phi_i, i \geq 1\}$, if there exists a positive constant $M < \infty$ such that $|\phi_1| < \infty$, then for any $\varepsilon > 0$, the following holds

$$\mathbb{P}\left\{ \left| \sum_{i=1}^{N} \phi_i \right| > \varepsilon N \right\} \leq 2\exp\left(-\frac{\varepsilon^2 N}{2\sigma^2 \left(1 + \varepsilon \frac{M}{\sigma^2}\right)} \right), \tag{2.3}$$

where $\sigma^2 = \mathbb{E}(\phi_1^2)$.
To apply this, we define

$$\phi_{i,\ell}(\mathbf{x}_k) = Y_i^{-\ell} K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_N}\right) - \mathbb{E}\left[Y_1^{-\ell} K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_1}{h_N}\right)\right], \text{ for } \ell = 1, 2.$$

It follows that

$$\Xi_{1,\ell}(\mathbf{x}_k) = \frac{1}{N h_N^d} \sum_{i=1}^{N} \phi_{i,\ell}(\mathbf{x}_k).$$

Next, we compute

$$\begin{aligned}
\mathbb{E}(\phi_{1,\ell}^2(\mathbf{x}_k)) &= \mathbb{E}\left[Y_1^{-2\ell} K_d^2\left(\frac{\mathbf{x}_k - \mathbf{X}_1}{h_N}\right)\right] - \mathbb{E}^2\left[Y_1^{-\ell} K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_1}{h_N}\right)\right] \\
&=: \mathcal{R}_1 - \mathcal{R}_2.
\end{aligned}$$

By employing a change of variable, a Taylor expansion and under Assumptions **K**(b) and **D**(a), we obtain for $\lambda = 2\ell$ with $\ell = 1, 2$

$$\begin{aligned}
\mathcal{R}_1 &= \mathbb{E}\left[Y_1^{-2\ell} K_d^2\left(\frac{\mathbf{x}_k - \mathbf{X}_1}{h_N}\right)\right] \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}_+^*} t^{-2\ell} K_d^2\left(\frac{\mathbf{x}_k - \mathbf{u}}{h_N}\right) f_{\mathbf{X},Y}(\mathbf{u}, t)\mathbf{du}dt \\
&= \int_{\mathbb{R}^d} K_d^2\left(\frac{\mathbf{x}_k - \mathbf{u}}{h_N}\right) \psi_\lambda(\mathbf{u})\mathbf{du} \\
&= h_N^d \int_{\mathbb{R}^d} K_d^2(\mathbf{s})\psi_\lambda(\mathbf{x}_k - \mathbf{s}h_N)\mathbf{ds} \\
&= O(h_N^d). \qquad (2.4)
\end{aligned}$$

Conversely, for $\mathcal{R}_2$, we have

$$\begin{aligned}
\sqrt{\mathcal{R}_2} &= \mathbb{E}\left[Y_1^{-\ell} K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_1}{h_N}\right)\right] \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}_+^*} t^{-\ell} K_d\left(\frac{\mathbf{x}_k - \mathbf{u}}{h_N}\right) f_{\mathbf{X},Y}(\mathbf{u}, t)\mathbf{du}dt \\
&= \int_{\mathbb{R}^d} K_d\left(\frac{\mathbf{x}_k - \mathbf{u}}{h_N}\right) \psi_\ell(\mathbf{u})\mathbf{du} \\
&= h_N^d \int_{\mathbb{R}^d} K_d(\mathbf{s})\psi_\ell(\mathbf{x}_k - \mathbf{s}h_N)\mathbf{ds},
\end{aligned}$$

using a Taylor expansion around $\mathbf{x}_k$ and under Assumptions **K**(a), **D**(a), we get

$$\mathcal{R}_2 = O(h_N^{2d}). \qquad (2.5)$$

Consequently, from (2.4) and (2.5), we obtain

$$\mathbb{E}(\phi_{1,\ell}(\mathbf{x}_k)) = O(h_N^d).$$

We are currently in a position to apply Bernstein's inequality given in equation (2.3). For $\varepsilon \geq 0$, we have

$$\begin{aligned}
\mathbb{P}\{|\widehat{\psi}_{N,\ell}(\mathbf{x}_k) - \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x}_k)]| > \varepsilon\} &= \mathbb{P}\left\{\left|\sum_{i=1}^{N} \phi_{i,\ell}(\mathbf{x}_k)\right| > N h_N^d \varepsilon\right\} \\
&\leq 2\exp\left(-\frac{N h_N^{2d}\varepsilon^2}{2\sigma^2\left(C + h_N^d \varepsilon \frac{M}{\sigma^2}\right)}\right) \\
&\leq 2\exp\left(-\frac{N h_N^d \varepsilon^2}{2\left(C + M\varepsilon\right)}\right) \qquad (2.6)
\end{aligned}$$

By setting $\varepsilon = \varepsilon_0\left(\sqrt{\frac{\log N}{N h_N^d}}\right) =: \varepsilon_N$ for all $\varepsilon_0 > 0$, the inequality (2.6) simplifies to

$$\begin{aligned}
\mathbb{P}\{|\widetilde{\psi}_{N,\ell}(\mathbf{x}_k) - \mathbf{E}[\widetilde{\psi}_{N,\ell}(\mathbf{x}_k)]| > \varepsilon_N\} &\leq 2\exp\left(-\frac{\varepsilon_0^2 \log N}{2\left(C + M\varepsilon_0\sqrt{\frac{\log N}{N h_N^d}}\right)}\right) \\
&\leq N^{-\frac{\varepsilon_0^2}{2\left(C+M\varepsilon_0\sqrt{\frac{\log N}{N h_N^d}}\right)}}.
\end{aligned}$$

We can write now

$$\begin{aligned}
\mathbb{P}\{\max_{1\leq k\leq \omega_N} |\widehat{\psi}_{N,\ell}(\mathbf{x}_k) - \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x}_k)]| > \varepsilon_N\} &\leq \sum_{i=1}^{\omega_N} \mathbb{P}\{|\widehat{\psi}_{N,\ell}(\mathbf{x}_k) - \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x}_k)]| > \varepsilon_N\} \\
&\leq M a_N^{-d} N^{-\frac{\varepsilon_0^2}{2\left(C+M\varepsilon_0\sqrt{\frac{\log N}{N h_N^d}}\right)}} \\
&\leq M h_N^{-d\left(1+\frac{1}{2\gamma}\right)} N^{\frac{1}{2\gamma}} N^{-\frac{\varepsilon_0^2}{2\left(C+M\varepsilon_0\sqrt{\frac{\log N}{N h_N^d}}\right)}} \\
&\leq M h_N^{-d\left(1+\frac{1}{2\gamma}\right)} N^{\frac{1}{2}\left(\frac{1}{\gamma} - \frac{\varepsilon_0^2}{C}\right)}.
\end{aligned} \qquad (2.7)$$

With a suitable choice of ε_0, the above series is summable. Then, by applying Borel-Cantelli's lemma to (2.7), we conclude the proof. □

2.4.2 Proofs for asymtotic normality

To prove the asymtotic normality result presented in Theorem 2.2, we utilize the following decomposition

$$\begin{aligned}
\sqrt{N h_N^d}\left(\widehat{\psi}_{N,\ell}(\mathbf{x}) - \psi_{N,\ell}(\mathbf{x})\right) &= \sqrt{N h_N^d}\left\{\left(\widehat{\psi}_{N,\ell}(\mathbf{x}) - \mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x})]\right)\right. \\
&= \left.+\left(\mathbb{E}[\widehat{\psi}_{N,\ell}(\mathbf{x})] - \psi_\ell(\mathbf{x})\right)\right\} \\
&=: \sqrt{N h_N^d}\left\{\Xi_{1,\ell}(\mathbf{x}) + \Xi_{2,\ell}(\mathbf{x})\right\}, \text{ for } \ell = 1, 2. \qquad (2.8)
\end{aligned}$$

We begin by demonstrating in the subsequent lemma that the term $\sqrt{Nh_N^d}\Xi_{2,\ell}(\mathbf{x})$ is negligible. Following this, we show that the leading term $\sqrt{Nh_N^d}\Xi_{1,\ell}(\mathbf{x})$ is asymptotically normal using Lyapunov's central limit theorem.

Lemma 2.3. *Under Assumptions* ***K****(a),* ***H****(b),* ***D****(a), we have for* $\ell = 1, 2$

$$\sqrt{Nh_N^d}|\Xi_{2,\ell}(\boldsymbol{x})| = o_{a.s.}(1) \text{ as } N \to \infty.$$

Proof. Building upon the result of Lemma 2.1 and using Assumption **H**(b), we find

$$\sqrt{Nh_N^d}|\Xi_{2,\ell}(\mathbf{x})| = O_{a.s.}\left(\sqrt{Nh_N^{d+4}}\right) = o_{a.s.}(1),$$

which concludes the proof. □

Before advancing to the asymptotic normality of the dominant term, we first compute the necessary variance and covariance in Lemma 2.4 and Lemma 2.5.

Lemma 2.4. *Under Assumptions* ***K****(a),* ***K****(b),* ***D****(a), we have for* $\lambda = 2\ell$ *with* $\ell = 1, 2$

$$Var\left(\sqrt{Nh_N^d}\Xi_{1,\ell}(\boldsymbol{x})\right) \to \psi_\lambda(\boldsymbol{x})\kappa.$$

Proof. For $\ell = 1, 2$, we have

$$\begin{aligned} Var\left(\sqrt{Nh_N^d}\Xi_{1,\ell}(\mathbf{x})\right) &= Nh_N^d\left\{\mathbb{E}\left(\widehat{\psi}^2_{N,\ell}(\mathbf{x})\right) - \mathbb{E}^2\left(\widehat{\psi}_{N,\ell}(\mathbf{x})\right)\right\} \\ &= \frac{1}{h_N^d}\left\{\mathbb{E}\left[Y_1^{-2\ell}K_d^2\left(\frac{\mathbf{x}-X_1}{h_N}\right)\right] - \mathbb{E}^2\left[Y_1^{-\ell}K_d\left(\frac{\mathbf{x}-X_1}{h_N}\right)\right]\right\} \\ &=: \frac{1}{h_N^d}\left\{\mathcal{I}_1 - \mathcal{I}_2\right\}. \end{aligned}$$

Using Assumptions **K**(b) and **D**(a), for $\lambda = 2\ell$ with $\ell = 1, 2$, we obtain

$$\begin{aligned} \mathcal{I}_1 &= \mathbb{E}\left[Y_1^{-2\ell}K_d^2\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_N}\right)\right] \\ &= \int_{\mathbb{R}^d}\int_{\mathbb{R}_+^*} t^{-2\ell}K_d^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_N}\right) f_{\mathbf{X},Y}(\mathbf{u},t)\mathbf{du}dt \\ &= \int_{\mathbb{R}^d} K_d^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_N}\right)\psi_\lambda(\mathbf{u})\mathbf{du} \\ &= h_N^d\int_{\mathbb{R}^d} K_d^2(\mathbf{s})\psi_\lambda(\mathbf{x}-\mathbf{s}h_n)\mathbf{ds} \\ &= h_N^d\psi_\lambda(\mathbf{x})\int_{\mathbb{R}^d} K_d^2(\mathbf{s})\mathbf{ds} \\ &= h_n^d\psi_\lambda(\mathbf{x})\kappa. \end{aligned} \tag{2.9}$$

Similarly, for $\mathcal{I}_2$, applying Assumptions **K**(a) and **D**(a), we get

$$\begin{aligned}
\sqrt{\mathcal{I}_2} &= \mathbb{E}\left[Y_1^{-\ell} K_d\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right] \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}_+^*} t^{-\ell} K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_N}\right) f_{\mathbf{X},Y}(\mathbf{u},t)\mathbf{du}dt \\
&= \int_{\mathbb{R}^d} K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_N}\right) \psi_\ell(\mathbf{u})\mathbf{du} \\
&= h_N^d \int_{\mathbb{R}^d} K_d(\mathbf{s})\psi_\ell(\mathbf{x}-\mathbf{s}h_n)\mathbf{ds} \\
&= O(h_N^d).
\end{aligned} \tag{2.10}$$

Combining (2.9) and (2.10), we derive

$$Var\left(\sqrt{Nh_N^d}\,\Xi_{1,\ell}(\mathbf{x})\right) = \frac{1}{h_N^d}\left\{h_N^d \psi_\lambda(\mathbf{x})\kappa - C h_N^{2d}\right\} \xrightarrow[N\to+\infty]{} \psi_\lambda(\mathbf{x})\kappa. \tag{2.11}$$

which completes the proof. □

Lemma 2.5. *Under Assumptions* ***K****(a),* ***K****(b) and* ***D****(a), we have*

$$Cov\left(\sqrt{Nh_N^d}\,\Xi_{1,1}(\boldsymbol{x}), \sqrt{Nh_N^d}\,\Xi_{1,2}(\boldsymbol{x})\right) \to \psi_3(\boldsymbol{x})\kappa.$$

Proof. By definition, we can express the covariance as

$$\begin{aligned}
Cov\left(\sqrt{Nh_N^d}\,\Xi_{1,1}(\mathbf{x}), \sqrt{Nh_N^d}\,\Xi_{1,2}(\mathbf{x})\right) &= Nh_N^d \mathbb{E}\left(\Xi_{1,1}(\mathbf{x})\Xi_{1,2}(\mathbf{x})\right) \\
&\quad -\sqrt{Nh_N^d}\left(\mathbb{E}\left(\Xi_{1,1}(\mathbf{x})\right)\mathbb{E}\left(\Xi_{1,2}(\mathbf{x})\right)\right) \\
&= \frac{1}{h_N^d}\left\{\mathbb{E}\left(Y_1^{-3}K_d^2\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_N}\right)\right)\right. \\
&\quad -\mathbb{E}\left(Y_1^{-1}K_d\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_N}\right)\right) \\
&\quad \left.\times\mathbb{E}\left(Y_1^{-2}K_d\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_N}\right)\right)\right\} \\
&=: \frac{1}{h_N^d}\left\{\mathcal{J}_1 - \mathcal{J}_2\mathcal{J}_3\right\}.
\end{aligned}$$

Following Assumptions **K**(b) and **D**(a), we find

$$
\begin{aligned}
\mathcal{J}_1 &= \mathbb{E}\left[Y_1^{-3}K_d^2\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_N}\right)\right] \\
&= \int_{\mathbb{R}^d}\int_{\mathbb{R}_+^*} t^{-3}K_d^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_N}\right) f_{\mathbf{X},Y}(\mathbf{u},t)\mathbf{du}dt \\
&= \int_{\mathbb{R}^d} K_d^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_N}\right)\psi_3(\mathbf{u})\mathbf{du} \\
&= \int_{\mathbb{R}^d} h_N^d K_d^2(\mathbf{s})\psi_3(\mathbf{x}-\mathbf{s}h_N)\mathbf{ds} \\
&= h_N^d\psi_3(\mathbf{x})\int_{\mathbb{R}^d} K_d^2(\mathbf{s})\mathbf{ds} \\
&= h_N^d\psi_3(\mathbf{x})\kappa.
\end{aligned}
$$

For $\mathcal{J}_2$ and $\mathcal{J}_3$, we apply the same methodology as in the proof of $\sqrt{\mathcal{I}_2}$ from Lemma 2.4 and at the end, we get

$$
Cov\left(\sqrt{Nh_N^d}\Xi_{1,1}(\mathbf{x}),\sqrt{Nh_N^d}\Xi_{1,2}(\mathbf{x})\right) = \frac{1}{h_N^d}\left\{h_N^d\psi_3(\mathbf{x})\kappa - Ch_N^{2d}\right\} \xrightarrow[N\to+\infty]{} \psi_3(\mathbf{x})\kappa. \quad (2.12)
$$

This completes the proof. □

We now turn our attention to demonstrate the asymptotic normality of the leading term. To accomplish this, we establish the following lemma.

Lemma 2.6. *Under Assumptions* ***K****(a),* ***K****(b),* ***D****(a), we have*

$$
\sqrt{Nh_N^d}\left(\Xi_{1,1}(\boldsymbol{x}),\Xi_{1,2}(\boldsymbol{x})\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma_{\boldsymbol{x}}\kappa),
$$

where $\Sigma_{\boldsymbol{x}} = \begin{pmatrix}\psi_2(\boldsymbol{x}) & \psi_3(\boldsymbol{x})\\ \psi_3(\boldsymbol{x}) & \psi_4(\boldsymbol{x})\end{pmatrix}$.

Proof. To prove this lemma, it suffices to show that for any given $a^T = (a_1, a_2)$

$$
\sqrt{Nh_N^d}\left(a_1\Xi_{1,1}(\mathbf{x}) + a_2\Xi_{1,2}(\mathbf{x})\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, a^T\Sigma_{\mathbf{x}}a\kappa).
$$

To this end, we apply Lyapounov's central limit theorem.
First, we define

$$
\begin{aligned}
\Xi(\mathbf{x}) &= \sqrt{Nh_N^d}(a_1\Xi_{1,1}(\mathbf{x}) + a_2\Xi_{1,2}(\mathbf{x})) \\
&= \sum_{i=1}^N\left\{a_1\frac{1}{\sqrt{Nh_N^d}}\left[Y_i^{-1}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right) - \mathbb{E}\left(Y_i^{-1}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right)\right)\right]\right. \\
&\quad \left. +a_2\frac{1}{\sqrt{Nh_N^d}}\left[Y_i^{-2}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right) - \mathbb{E}\left(Y_i^{-2}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right)\right)\right]\right\} \\
&=: \sum_{i=1}^N \Xi_i(\mathbf{x}).
\end{aligned}
$$

Then, we verify that for $\nu > 2$

$$\frac{1}{(Var(\Xi(\mathbf{x})))^{\frac{\nu}{2}}}\sum_{i=1}^{N}\mathbb{E}\left[|\Xi_i(\mathbf{x})|^{\nu}\right] \longrightarrow 0.$$

By applying C_r inequality (See, Loève (1965)), we obtain

$$\begin{aligned}
\mathbb{E}\left[|\Xi_i(\mathbf{x})|^{\nu}\right] &= \mathbb{E}\left|a_1\frac{1}{\sqrt{Nh_N^d}}\left[Y_i^{-1}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right)-\mathbb{E}\left(Y_i^{-1}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right)\right)\right]\right.\\
&\quad \left.+a_2\frac{1}{\sqrt{Nh_N^d}}\left[Y_i^{-2}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right)-\mathbb{E}\left(Y_i^{-2}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right)\right)\right]\right|^{\nu}\\
&\le 2^{\nu-1}\left(\frac{h_N^d}{N}\right)^{\frac{\nu}{2}}\left[a_1^{\nu}\mathbb{E}\left|\frac{1}{h_N^d}Y_i^{-1}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right)\right|^{\nu}\right.\\
&\quad \left.+a_2^{\nu}\mathbb{E}\left|\frac{1}{h_N^d}Y_i^{-2}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right)\right|^{\nu}\right].
\end{aligned}$$

Since

$$\mathbb{E}\left(Y_i^{-\ell}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_N}\right)\right) = O\left(h_N^d\right),$$

we obtain

$$\sum_{i=1}^{n}\mathbb{E}\left[|\Xi_i(\mathbf{x})|^{\nu}\right] = O\left(N^{1-\frac{\nu}{2}}h_N^{\frac{d\nu}{2}}\right), \tag{2.13}$$

which is $o(1)$, since $1-\frac{\nu}{2} < 0$.
On the other hand, we have

$$\begin{aligned}
Var\left(\Xi(\mathbf{x})\right) &= a_1^2 Var\left(\sqrt{Nh_N^d}\,\Xi_{1,1}(\mathbf{x})\right) + a_2^2 Var\left(\sqrt{Nh_N^d}\,\Xi_{1,2}(\mathbf{x})\right)\\
&\quad +2a_1a_2Cov\left(\sqrt{Nh_N^d}\,\Xi_{1,1}(\mathbf{x}),\sqrt{Nh_N^d}\,\Xi_{1,2}(\mathbf{x})\right).
\end{aligned}$$

Using Lemma 2.4 and Lemma 2.5, we get

$$Var\left(\Xi(\mathbf{x})\right) \xrightarrow[N\to+\infty]{} a_1^2\psi_2(\mathbf{x})\kappa + a_2^2\psi_4(\mathbf{x})\kappa + 2a_1a_2\psi_3(\mathbf{x})\kappa = a^T\Sigma_{\mathbf{x}}a\kappa > 0. \tag{2.14}$$

From (2.13) and (2.14), the Lyaponouv's condition is satisfied and then the proof of Lemma 2.6 is complete. □

Proof of Theorem 2.2. We define the function $\Phi:\mathbb{R}^2\to\mathbb{R}$ as follows: $\Phi(x,y)=\frac{x}{y}$, for $y\neq 0$. This allows us to express the difference between the estimator and the regression function as

$$\begin{aligned}
\widehat{m}_N(\mathbf{x}) - m(\mathbf{x}) &= \frac{\widehat{\psi}_{N,1}(\mathbf{x})}{\widehat{\psi}_{N,2}(\mathbf{x})} - \frac{\psi_1(\mathbf{x})}{\psi_2(\mathbf{x})}\\
&= \Phi\left(\widehat{\psi}_{N,1}(\mathbf{x}),\widehat{\psi}_{N,2}(\mathbf{x})\right) - \Phi\left(\psi_1(\mathbf{x}),\psi_2(\mathbf{x})\right).
\end{aligned}$$

From equation (2.8), Lemma 2.3, Lemma 2.6 and by applying the delta method (See, Shao (2003)), we obtain

$$\sqrt{nh_n^d}\left(\widehat{m}_n(x) - m(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \nabla\Phi^T\Sigma_x\nabla\Phi\kappa\right),$$

where the gradient $\nabla\Phi^T = \left(\frac{\partial\Phi}{\partial x}, \frac{\partial\Phi}{\partial y}\right)$ is evaluated at point $(\psi_1(\mathbf{x}), \psi_2(\mathbf{x}))$. Then an elementary calculation gives $\nabla\Phi^T\Sigma_{\mathbf{x}}\nabla\Phi\kappa = \sigma_N^2(\mathbf{x})$ which completes the proof. □

Chapter 3

Strong uniform consistency of the relative regression function estimator: LTRC and α-mixing data

In this chapter, we propose a nonparametric estimator for the regression function associated with left truncated and right censored (LTRC) data. The estimator is constructed by minimizing the mean squared relative error. Under the α-mixing condition, we establish the strong uniform convergence of the estimator with a rate over a compact set. To evaluate the estimator's performance, we conduct an extensive simulation study, comparing its efficiency against that of classical regression estimator using finite samples across various scenarios. Moreover, we illustrate the practical utility of the proposed estimator through a real world case study.

3.1 Introduction

In regression analysis, many results have been established under the assumption of complete and independent samples. However, it is often beneficial to consider and analyze incomplete and dependent samples to address practical situations where data are neither complete nor independent.

For instance, in a clinical trial investigating mortality among patients diagnosed with HIV, the dataset may exhibit both left truncation and right censoring. Left truncation occurs when only patients who have survived for a specified duration after diagnosis (e.g., one year) are enrolled in the study, thus excluding those who died shortly after diagnosis. Right censoring, on the other hand, arises when patients remain alive at the study's conclusion or are lost to follow-up, leaving their survival times incomplete. Additionally, the data may demonstrate dependence due to shared environmental factors, treatment protocols, or the timing of diagnoses among patients. This highlights the necessity of accounting for dependence structures, such as α-mixing, in the analytical framework.

The primary objective of this chapter is to introduce a nonparametric estimator for the regression function for LTRC data and to establish its strong uniform convergence under the assumption of α-mixing dependence. The organization of this chapter is as follows: Section 3.2 introduces the estimator under LTRC model. Section 3.3 presents the assumptions and main result. Section 3.4 describes the simulation study, followed by a real data application in Section 3.5. Finally, the proofs are provided in Section 3.6.

3.2 Definition of the estimator under LTRC model

Let $\{(Y_i, T_i, W_i);\ i = 1, \ldots, N\}$ be a sequence of random vectors from (Y, T, W), where Y denotes the lifetime with continuous distribution function (df) F. T and W are the variables of left truncation and right censoring times with continuous dfs L and G, respectively. In the random LTRC model, one observes (Z, T, δ) such that $Z \geq T$, where $Z = Y \wedge W$ and $\delta = \mathbb{1}_{\{Y \leq W\}}$, with $\wedge$ denoting the minimum operator and $\mathbb{1}_A$ being the indicator function of the event A. When $Z < T$, nothing is observed. All along this chapter, we suppose that Y, T and W are mutually independent, then Z has a df $H = 1 - (1 - F)(1 - G)$. Take $\eta := \mathbb{P}(Z \geq T)$, so it is necessary to assume $\eta > 0$ in order to have at least one observation available. As a consequence of truncation, n, the actual sample size, with $n \leq N$ and N fixed but not observable. Then without possible confusion, we still denote $\{(Z_i, T_i, \delta_i);\ i = 1, \ldots, n\}$ the observed sample.

Since N is unknown and n is known, our results would be stated with respect to conditional probability $\mathbf{P}$ (related to the n-sample) instead of the probability measure $\mathbb{P}$ (related to the N-sample). Similarly, $\mathbb{E}$ and $\mathbf{E}$ will denote the expectation operators related to $\mathbb{P}$ and $\mathbf{P}$, respectively. More generally, any operator or function related to the probability measure $\mathbf{P}$ will be denoted in bold.

Now, let $C(\cdot)$ be defined by

$$\begin{aligned} C(y) &:= \mathbb{P}(T \leq y \leq Z | Z \geq T) \\ &= \frac{1}{\eta} L(y)(1 - H(y)), \end{aligned} \tag{3.1}$$

with the empirical estimator

$$C_n(y) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{T_i \leq y \leq Z_i\}}.$$

The nonparametric estimator of the df F is the TJW product-limit estimator F_n, defined in Tsai et al. (1987) as

$$F_n(y) = 1 - \prod_{i=1}^{n} \left(1 - \frac{\mathbb{1}_{\{Z_i \leq y\}} \delta_i}{n C_n(Z_i)}\right).$$

For any df Q, we define

$$a_Q = \inf\{y : Q(y) > 0\} \text{ and } b_Q = \sup\{y : Q(y) < 1\}$$

as the endpoints of the Q support. Gijbels and Wang (1993) pointed out that the df F can be estimated only if

$$a_L \leq a_H \text{ and } b_L \leq b_H. \tag{3.2}$$

By the strong law of large numbers, as $N \to \infty$ we have

$$\frac{n}{N} \longrightarrow \eta, \ \mathbb{P}\text{- a.s.}$$

Note that, the ratio $\frac{n}{N}$ can not be used to estimate η since N is unknown. Indeed, from (3.1) we have

$$\eta = \frac{L(y)(1-F(y))(1-G(y))}{C(y)}, \tag{3.3}$$

where η is independent of y. Following the idea of He and Yang (1998), a feasible estimator for α is given by

$$\eta_n = \frac{L_n(y)(1-F_n(y))(1-G_n(y))}{C_n(y)}, \tag{3.4}$$

for any y such that $C_n(y) \neq 0$, where G_n is the TJW-type estimator of the df G, defined as

$$G_n(y) = 1 - \prod_{i=1}^{n} \left(1 - \frac{\mathbb{1}_{\{Z_i \leq y\}}(1-\delta_i)}{nC_n(Z_i)}\right) \tag{3.5}$$

and L_n is the Lynden-Bell (1971) estimator of the df L, given by

$$L_n(y) = \prod_{i=1}^{n} \left(1 - \frac{\mathbb{1}_{\{T_i > y\}}}{nC_n(T_i)}\right). \tag{3.6}$$

Henceforth, in addition to the triplet (Z, T, δ), assume that covariates are present and then one observes $\{(\mathbf{X}_i, Z_i, T_i, \delta_i);\ i = 1, \ldots, n\}$, with $Z_i \geq T_i$. Throughout this chapter, we suppose that condition (3.2) is satisfied and

$$(T, W) \text{ are independent of } (\mathbf{X}, Y). \tag{3.7}$$

Then, the sub-conditional df of $(\mathbf{X}, Z, \delta = 1)$ is given by

$$\begin{aligned}
\mathbf{H}_1(\mathbf{x}, y) &:= \mathbf{P}(\mathbf{X} \leq \mathbf{x}, Z \leq y, \delta = 1) \\
&= \mathbb{P}(\mathbf{X} \leq \mathbf{x}, Z \leq y, \delta = 1 | Z \geq T) \\
&= \frac{1}{\eta}\mathbb{P}(\mathbf{X} \leq \mathbf{x}, Y \leq y, Y \leq W, Y \geq T) \\
&= \frac{1}{\eta}\int_{-\infty}^{\mathbf{x}}\int_{a_H}^{y} L(t)(1-G(t)) f_{\mathbf{X},Y}(\mathbf{u}, t)\mathbf{du}dt, \qquad (3.8)
\end{aligned}$$

where, $f_{\mathbf{X},Y}(\cdot,\cdot)$ is the joint density function of $(\mathbf{X}, Y)$.
By differentiating, (3.8) becomes

$$d\mathbf{H}_1(\mathbf{x}, y) = \frac{1}{\eta} L(y)(1 - G(y)) f_{\mathbf{X},Y}(\mathbf{x}, y). \tag{3.9}$$

For any $\mathbf{x} \in \mathbb{R}^d$, the regression function given in (1.2) can be written as

$$m(\mathbf{x}) = \frac{\displaystyle\int_{\mathbb{R}_+^*} y^{-1} f_{\mathbf{X},Y}(\mathbf{x}, y) dy}{\displaystyle\int_{\mathbb{R}_+^*} y^{-2} f_{\mathbf{X},Y}(\mathbf{x}, y) dy} =: \frac{\psi_1(\mathbf{x})}{\psi_2(\mathbf{x})}, \tag{3.10}$$

with

$$\psi_\ell(\mathbf{x}) = \int_{\mathbb{R}_+^*} y^{-\ell} f_{\mathbf{X},Y}(\mathbf{x}, y) dy, \text{ for } \ell = 1, 2.$$

Under the random right censored model, Bouhadjera et al. (2019) used the so-called synthetic data to define a kernel estimator of (3.10) as

$$\widehat{m}(\mathbf{x}) =: \frac{\widehat{\psi}_1(\mathbf{x})}{\widehat{\psi}_2(\mathbf{x})},$$

with

$$\widehat{\psi}_\ell = \frac{1}{nh_n^d} \sum_{i=1}^{n} \frac{\delta_i Z_i^{-\ell}}{\bar{G}_n(Z_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right),$$

where $\bar{G}_n = 1 - G_n$ is the estimator of the survival function of the censored rv W, $K_d : \mathbb{R}^d \to \mathbb{R}$ is a kernel function, and h_n is a sequence of positive real numbers that approaches zero as $n \to \infty$.
Now, under the random LTRC model, following the idea introduced by Carbonez et al. (1995), an unbiased estimate for $\mathbb{E}[Y^{-\ell}|\mathbf{X}]$, $\ell = 1, 2$ is given by

$$\frac{1}{N} \sum_{i=1}^{N} \frac{\delta_i Z_i^{-\ell}}{L(Z_i)\bar{G}(Z_i)} \mathbb{1}_{\{Z_i \geq T_i\}}. \tag{3.11}$$

Indeed, using the conditional expectation property and condition (3.7), we have

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{N} \sum_{i=1}^{N} \frac{\delta_i Z_i^{-\ell}}{L(Z_i)\bar{G}(Z_i)} \mathbb{1}_{\{Z_i \geq T_i\}}\right] &= \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left(\mathbb{E}\left[\frac{\delta_i Z_i^{-\ell}}{L(Z_i)\bar{G}(Z_i)} \mathbb{1}_{\{Z_i \geq T_i\}} | \mathbf{X}_i, Y_i\right]\right) \\
&= \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}\left(\frac{Y_i^{-\ell}}{L(Y_i)\bar{G}(Y_i)} \mathbb{E}[\mathbb{1}_{\{Y_i \leq W_i\}} \mathbb{1}_{\{Y_i \geq T_i\}} | \mathbf{X}_i, Y_i]\right) \\
&= \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}[Y_i^{-\ell} | \mathbf{X}_i] \\
&= \mathbb{E}[Y^{-\ell} | \mathbf{X}].
\end{aligned}$$

Unfortunately, (3.11) can not be used in practice since N, L and $\bar{G}$ are unknown. Therefore, following the same reasoning as in Ould Saïd and Lemdani (2006), we define $\widehat{m}_n(\mathbf{x})$ as a kernel estimator for $m(\mathbf{x})$, given for any $\mathbf{x} \in \mathbb{R}^d$ by

$$\widehat{m}_n(\mathbf{x}) =: \frac{\widehat{\psi}_{n,1}(\mathbf{x})}{\widehat{\psi}_{n,2}(\mathbf{x})}, \tag{3.12}$$

with

$$\widehat{\psi}_{n,\ell}(\mathbf{x}) = \frac{\eta_n}{nh_n^d} \sum_{i=1}^{n} \frac{\delta_i Z_i^{-\ell}}{L_n(Z_i)\bar{G}_n(Z_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right), \text{ for } \ell = 1, 2.$$

3.3 Assumptions and main result

To establish the consistency result for the estimator $\widehat{m}_n(\cdot)$ based on the α-mixing sequence of random vectors $\{(X_i, Z_i, T_i, \delta_i); i = 1, \ldots, n\}$, we need to introduce some notations:

$$\widetilde{\psi}_n(\mathbf{x}) = \frac{\eta}{nh_n^d} \sum_{i=1}^{n} \frac{\delta_i Z_i^{-\ell}}{L(Z_i)\bar{G}(Z_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right), \text{ for } \ell = 1, 2$$

and

$$\mu_\lambda(\mathbf{u}) = \int_{t \geq a_H} \frac{t^{-\lambda}}{L(t)\bar{G}(t)} f_{\mathbf{X},Y}(\mathbf{u}, t)dt, \text{ for } \lambda = 2, 4.$$

Let $\mathcal{C}$ be a compact set in $\mathbb{R}^d$, and assume that $\inf_{\mathbf{x}\in\mathcal{C}} \widehat{\psi}_{n,2}(\mathbf{x}) > 0$. Throughout this chapter, when no confusion is possible, we will use C to represent any generic constant.

3.3.1 Assumptions

K. The kernel $K_d(\cdot)$ is a bounded probability density with compact support and satisfies:

(a) $\int_{\mathbb{R}^d} K_d(\mathbf{r})\mathbf{dr} = 1$, $\int_{\mathbb{R}^d} r_i K_d(\mathbf{r})\mathbf{dr} = 0$ and $\int_{\mathbb{R}^d} |r_i||r_j| K_d(\mathbf{r})\mathbf{dr} < \infty$, for $i, j = 1, \ldots, d$.

(b) $\int_{\mathbb{R}^d} K_d^2(\mathbf{r})\mathbf{dr} < \infty$ and $\int_{\mathbb{R}^d} r_i K_d^2(\mathbf{r})\mathbf{dr} < \infty$, for $i = 1, \ldots, d$.

(c) $\forall(\mathbf{t}, \mathbf{s}) \in \mathcal{C}^2$ $|K_d(\mathbf{t}) - K_d(\mathbf{s})| \leq \| \mathbf{t} - \mathbf{s} \|^\gamma$, for $\gamma > 0$.

H. The bandwidth h_n satisfies :

(a) $\lim_{n\to\infty} h_n = 0$, $\lim_{n\to\infty} nh_n^d = \infty$ and $\lim_{n\to\infty} \frac{\log n}{nh_n^d} = 0$.

(b) $\lim_{n\to\infty} h_n^{d(v-2)} \log n = 0$, for $v > 3$.

(c) $\exists\theta > 0, \exists C > 0$, such that

$$Cn^{\frac{\gamma(3-v)}{\gamma(v+1)+2\gamma+1}+\theta d} \leq h_n^d, \text{ for } v > 3 \text{ and } \gamma > 0.$$

D. (a) The function $\psi_\ell(\cdot)$ is twice continuously differentiable and $\sup_{\mathbf{x}\in\mathcal{C}} \left|\frac{\partial^2 \psi_\ell(\mathbf{x})}{\partial x_i \partial x_j}\right| < \infty$, for $\ell = 1, 2$ and $i, j = 1, \ldots, d$.

(b) $\forall Y > 0$, $\exists C > 0$, such that $Y^{-\ell} \leq C$, for $\ell = 1, 2$.

(c) The function $\mu_\lambda(\cdot)$ is continuously differentiable and $\sup_{\mathbf{x}\in\mathcal{C}} \left|\dfrac{\partial \mu_\lambda(\mathbf{x})}{\partial x_i}\right| < \infty$, for $\lambda = 2, 4$ and $i = 1, \ldots, d$.

(d) The joint density $\Upsilon_{i,j}(\cdot,\cdot)$ of $(\mathbf{X}_i, \mathbf{X}_j)$ exists and satisfies

$$\sup_{\mathbf{t},\mathbf{s}\in\mathcal{C}} |\Upsilon_{i,j}(\mathbf{t},\mathbf{s}) - \Upsilon_i(\mathbf{t})\Upsilon_j(\mathbf{s})| < \infty, \text{ for } i, j = 1, \ldots, n.$$

M. $\{(X_i, Z_i, T_i, \delta_i); i = 1, \ldots, n\}$ is an α-mixing sequence of random vectors, with strong mixing coefficient $\alpha(n)$. This coefficient satisfies:

(a) $\alpha(n) = O(n^{-v})$, for $v > 3$.

3.3.1.1 Comments on the Assumptions

Assumptions **K**, **H**(a), and **D**(a) are commonly used in nonparametric regression estimation in both independent and dependent cases. Assumption **D**(b) implies that the inverse of the variable of interest Y is bounded, which is specifically useful for proving Lemma 3.2 and Lemma 3.3. Assumptions **H**(b), **H**(c), **D**(c), **D**(d), and **M**(a) are technical and are necessary for studying the covariance term.

3.3.2 Main result

The following theorem presents the uniform almost sure convergence with a rate of the relative error regression estimator $\widehat{m}(\mathbf{x})$.

Theorem 3.1. *Under Assumptions **K**, **H**, **D**, and **M**, we have*

$$\sup_{\boldsymbol{x}\in\mathcal{C}} |\widehat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})| = O(h_n^2) + O\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(v-2)}\log n}\right) \ \boldsymbol{P}\text{- a.s. as } n \to \infty.$$

Remark 3.1. *The rate obtained here is identical to the one established for α-mixing and right censored data in the MSE framework by Guessoum and Ould Saïd (2010).*

3.4 Simulation study

To show how good our estimator $\widehat{m}_n(\cdot)$ is, a simulation study is performed for some particular cases of fixed size and different censoring, truncation and dependency rates, when the covariate $\mathbf{X}$ is one- and bi-dimensional random variable (i.e., d=1 and d=2).

3.4.1 One-dimensional case

3.4.1.1 Algorithm

1. Generate an α-mixing sequence $\{X_t; t \geq 1\}$ by the following AR(1) model:

$$X_{t+1} = \begin{cases} \rho X_t + 0.5e_{t+1}, & \text{if } \rho > 0.5 \\ \rho X_t + e_{t+1}, & \text{else,} \end{cases}$$

   where $0 < \rho < 1$ controls the degree of dependency, $X_1 = e_1$ and $e_t \rightsquigarrow N(0,1)$.

2. Calculate $Y_t = m(X_t) + \epsilon_t$, $t \geq 1$, where the white noise $\epsilon_t \rightsquigarrow N(0, 0.1)$.

3. Determine $Z_t = Y_t \wedge W_t$ and $\delta_t = \mathbb{1}_{\{Y_t \leq W_t\}}$, $t \geq 1$, where W_t is generated according to a exponential distribution with parameter a_0 which allows obtaining different censoring percentage (CP).

4. Generate $T_t = \rho T_{t-1} + b_0 + \xi_t$, $t \geq 2$, where $T_1 = \xi_1$, $\xi_t \rightsquigarrow N(0, 0.1)$ for $t \geq 1$ and b_0 is adapted to achieve different truncation percentage (TP).

5. Test if $Z_t \geq T_t$, $t \geq 1$. If true, the vector $(X_t, Z_t, T_t, \delta_t)$ is included in the final sample. Otherwise, reject the observation $(X_t, Z_t, T_t, \delta_t)$ and go back to step 1.

6. Repeat this procedure until the final simple size is n, i.e., $(X_i, Z_i, T_i, \delta_i)$; $i = 1, \ldots, n$.

7. Compute $\bar{G}_n(\cdot)$ and $L_n(\cdot)$ from (3.5) and (3.6), respectively.

8. Calculate the estimator $\widehat{m}_n(x)$ from (3.12) for $x \in \mathcal{C} = [-1.5, 1.5]$. The kernel $K(.)$ is taken as a standard normal distribution. The choice of the bandwidth is discussed immediately afterward.

3.4.1.2 Bandwidth selection

The bandwidth h_n is chosen as the minimizer of the global mean squared error (GMSE) criterion. We select this bandwidth from a grid of values denoted as $\mathcal{H}$. For each candidate bandwidth value $h_n \in \mathcal{H}$, we perform the following steps.

- Compute the MSE for the estimator $\widehat{m}_n(.)$ at the equidistant points $(x_i, i = 1, ..., A = 20)$ belonging to the compact set C. The MSE is calculated along $B = 50$ replications by:

$$MSE(x_i) = \frac{1}{B} \sum_{j=1}^{B} (\widehat{m}_{n,j}(x_i) - m(x_i))^2, i = 1, \ldots, A.$$

  Here $\widehat{m}_{n,j}(\cdot)$ is the value of the estimator $\widehat{m}_n(\cdot)$ at iteration j.

- Compute the GMSE by:

$$GMSE(h_n) = \frac{1}{A} \sum_{i=1}^{A} MSE(x_i). \tag{3.13}$$

Finally, the optimal bandwidth is determined by:

$$\arg\min_{h_n \in \mathcal{H}} GMSE(h_n).$$

3.4.1.3 Performance of the estimator $\widehat{m}_n(\cdot)$

In this part, we study the performance of our estimator when the theoretical function is of linear and nonlinear form.

1. Linear case

   We consider the following linear regression function $m(x) = x + 5$, then $Y_i = X_i + 5 + \epsilon_i$, $i = 1, \ldots, n$.

   (a) Effect of sample size and dependency: From Figure 3.1 and Figure 3.2 when CP and TP are fixed, we notice that the higher the sample size and smaller the rate of dependency, the better the quality of fit.

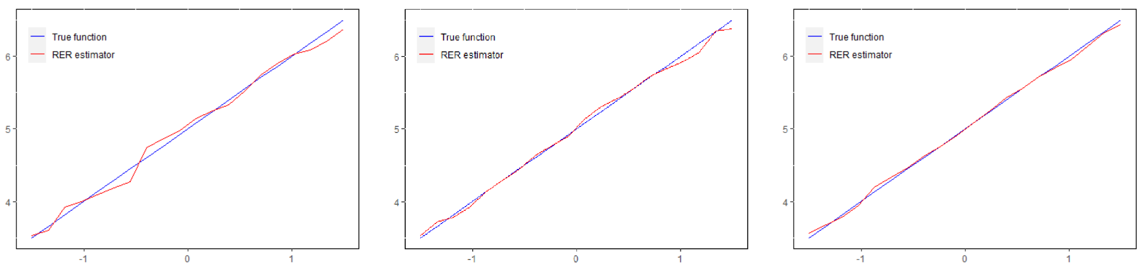

Figure 3.1: $m(\cdot)$ and $\widehat{m}_n(\cdot)$ with ρ=0.1, CP=20%, TP=20%, n=50,100, and 300 respectively.

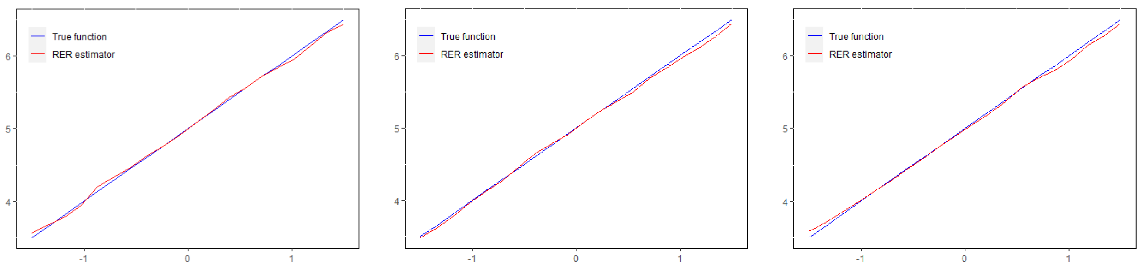

Figure 3.2: $m(\cdot)$ and $\widehat{m}_n(\cdot)$ with n=300, CP=20%, TP=20%, ρ=0.1, 0.5 and, 0.8 respectively.

(b) Effect of CP and TP: It is easy to see from Figure 3.3 that the estimator's quality is affected by CP, whereas it does not seem to be influenced by TP, as shown in Figure 3.4. In general, our estimator curve remains close to the theoretical curve even for a high CP.

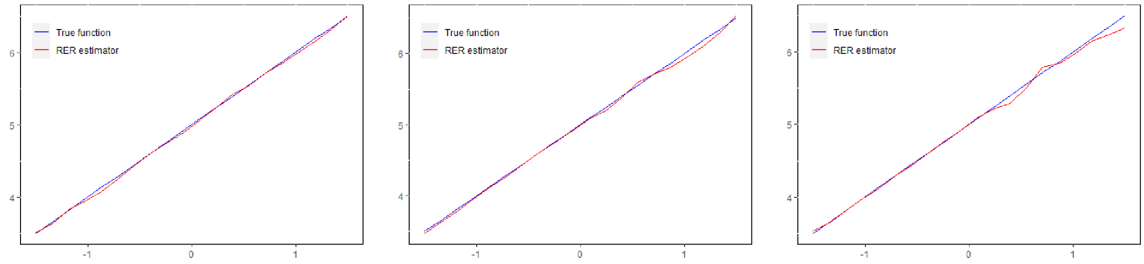

Figure 3.3: $m(\cdot)$ and $\widehat{m}_n(\cdot)$ with ρ=0.1, n=300, TP=20%, CP=10%, 30%, and 60% respectively.

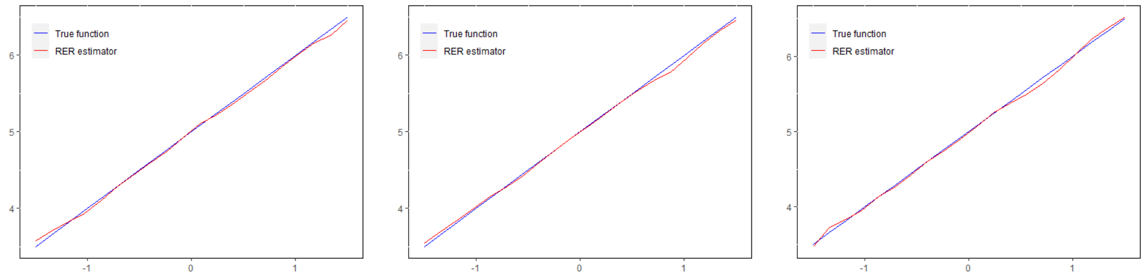

Figure 3.4: $m(\cdot)$ and $\widehat{m}_n(\cdot)$ with ρ=0.1, n=300, CP=20%, TP=10%, 30%, and 60% respectively.

(c) Effect of outliers: To show the robustness of our estimator, we create artificial outliers in the data; 4% of each sample is multiplied by a multiplier factor (MF). Then, from Figure 3.5, it is very clear that our estimator is resistant in the presence of outliers.

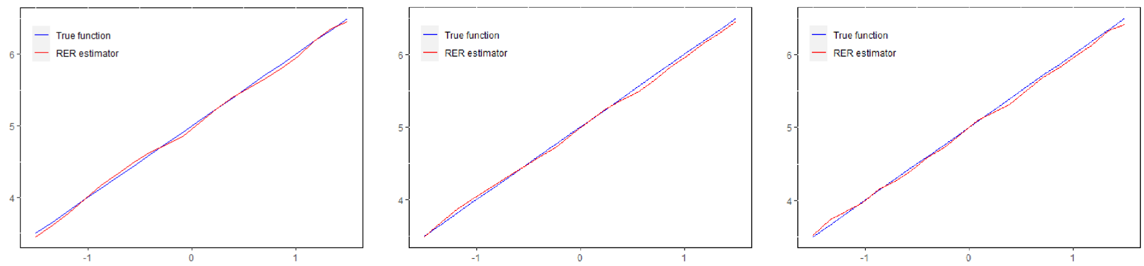

Figure 3.5: $m(\cdot)$ and $\widehat{m}_n(\cdot)$ with ρ=0.1, n=300, CP=20%, TP=20%, MF=50, 100, and 150 respectively.

2. Nonlinear case

Now, we consider the case of nonlinear regression by choosing the following three models:

Model 1:

$$Y = m_1(X) + \epsilon, \text{ with } m_1(x) = \exp(x) + 3,$$

Model 2:

$$Y = m_2(X) + \epsilon, \text{ with } m_2(x) = \sin(2x) + 4,$$

Model 3:

$$Y = m_3(X) + \epsilon, \text{ with } m_3(x) = x^2 + 5.$$

Figure 3.6 shows that the quality of fit for the nonlinear model is as good as for linear model.

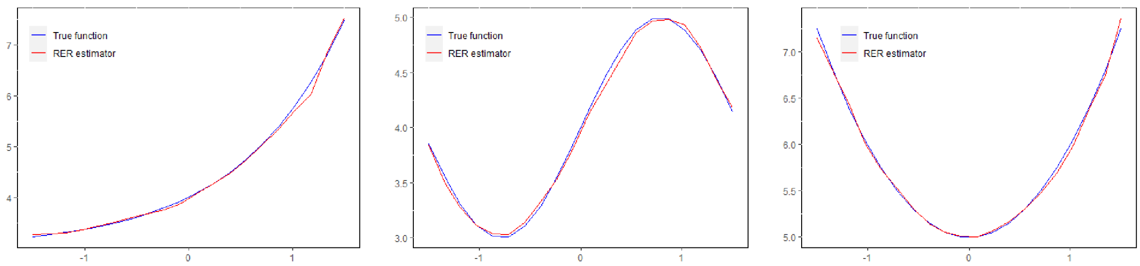

Figure 3.6: $m(\cdot)$ and $\widehat{m}_n(\cdot)$ with ρ=0.1, n=300, CP=20% and TP=20% for model 1, 2 and, 3 respectively.

3.4.1.4 Comparison study

Here, the goal is to compare the performance of the RER estimator with the classical regression (CR) estimator studied by Bey et al. (2022) and defined as:

$$\widehat{m}_{NW}(x) = \frac{\displaystyle\sum_{i=1}^{n} \frac{\delta_i Z_i}{L_n(Z_i)\bar{G}_n(Z_i)} K_d\left(\frac{x - X_i}{h_n}\right)}{\displaystyle\sum_{i=1}^{n} \frac{\delta_i}{L_n(Z_i)\bar{G}_n(Z_i)} K_d\left(\frac{x - X_i}{h_n}\right)},$$

in absence and presence of outliers. The performance of both estimators is evaluated via some graphic curves and the GMSE criterion under the linear regression function described in the one-dimensional case.

For the first case when there are no outliers in the observed samples, we can see from Figure 3.7 and Table 3.1 that there is no meaningful difference between the estimators $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$ and both of them share the following points

Table 3.1: GMSE's values of $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$.

		CP=20%				TP=20%			
		TP=10%		TP=40%		CP=10%		CP=40%	
ρ	n	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$
0.1	100	0.00228	0.00231	0.00234	0.00237	0.00203	0.00231	0.00417	0.00437
	300	0.00118	0.00105	0.00127	0.00116	0.00088	0.00091	0.00152	0.00150
0.8	100	0.00347	0.00325	0.00382	0.00333	0.00289	0.00250	0.00632	0.00594
	300	0.00181	0.00162	0.00224	0.00243	0.00095	0.00111	0.00217	0.00208

- The quality of fit becomes better when the sample size n increases.
- The estimator is affected by the degree of dependency and performs better for a small ρ.
- The accuracy of the estimator is influenced by CP and decreases with increasing CP, however it remains acceptable.
- The quality of estimation is slightly affected by TP for a small size of n, but this effect disappears with increasing sample size.

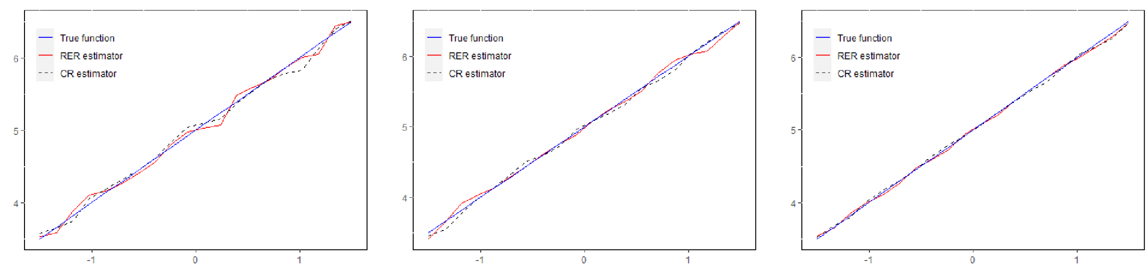

Figure 3.7: $m(\cdot)$, $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$ with ρ=0.1, CP=20%, TP= 20%, n=50, 100, and 300 respectively.

In the second case, we compare the estimators $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$ in the presence of outliers. For this purpose, we artificially introduce outliers by multiplying 4 % of each sample by a MF. As illustrated in Figure 3.8 and Table 3.2, the RER estimator is more stable than the CR estimator. This means that even if the quality of the estimation for both estimators decreases with increasing outliers, but this decrease in the quality is still not significant in the relative error regression compared to the classical one.

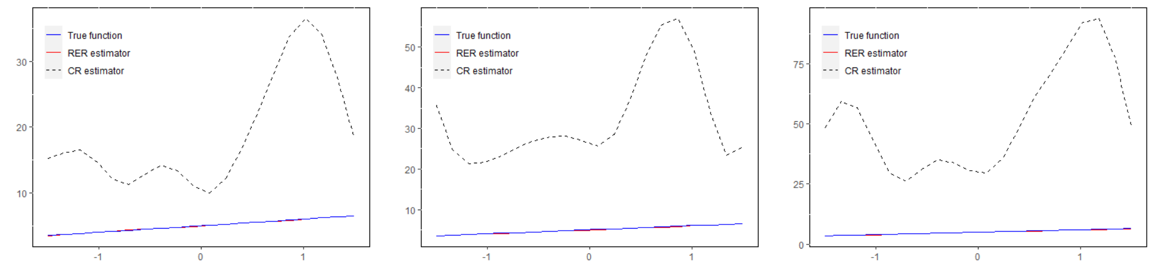

Figure 3.8: $m(\cdot)$, $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$ with n=300, ρ=0.1, CP=20%, TP= 20%, MF=50, 100, and 150 respectively.

Table 3.2: GMSE's values of $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$ with outliers for ρ=0.1.

		CP=20%				TP=20%			
		TP=10%		TP=40%		CP=10%		CP=40%	
n	MF	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$
	50	0.00245	3.11207×10^2	0.00266	4.18845×10^2	0.00237	2.48798×10^2	0.00463	5.62563×10^2
100	100	0.00316	1.52516×10^3	0.00358	1.92048×10^3	0.00263	1.00017×10^3	0.00549	2.53366×10^3
	150	0.00341	2.43604×10^3	0.00372	2.81923×10^3	0.00314	2.16226×10^3	0.00523	3.32406×10^3
	50	0.00107	2.30381×10^2	0.00107	3.05945×10^2	0.00092	2.26077×10^2	0.00164	4.51992×10^2
300	100	0.00121	9.30793×10^2	0.00132	1.17058×10^3	0.00102	1.39562×10^3	0.00183	1.92666×10^3
	150	0.00131	1.63935×10^3	0.00134	1.92157×10^3	0.00116	1.70041×10^3	0.00193	2.48458×10^3

3.4.2 Bi-dimensional case

In this second subsection, the aim is to study the performance of our estimator in the case of a bi-dimensional covariate under the following two models:

Model 1:

$$Y = m_1(X_1, X_2) + \epsilon, \text{ with } m_1(x_1, x_2) = x_1 + x_2 + 5,$$

Model 2:

$$Y = m_2(X_1, X_2) + \epsilon, \text{ with } m_2(x_1, x_2) = \cos(2x_1) + \cos(2x_2) + 4.$$

The data is generated using the same algorithm as for the one-dimensional case. In each model, We simulate $(X_{1,t}, X_{2,t})$, $t \geq 1$ as follows

$$\underset{j=1,2}{X_{j,t+1}} = \begin{cases} \rho X_{j,t} + 0.5 e_{j,t+1}, & \text{if } \rho > 0.5 \\ \rho X_{j,t} + e_{j,t+1}, & \text{else,} \end{cases}$$

where $X_{j,1} = e_{j,1}$ and $e_{j,t} \rightsquigarrow N(0,1)$. Then, we calculate $Y_t = m(X_{1,t}, X_{2,t}) + \epsilon_t$, $t \geq 1$. To compute the estimator $\widehat{m}_n(\cdot,\cdot)$, we use a standard multivariate normal kernel and a bandwidth that minimizes the GMSE defined in (3.13).

The results are presented in the following figures and tables. In general, the same comments that we gave in the one-dimensional case can be given here. More clearly, in the absence of outliers, we observe from Figures 3.9 and 3.10 (Model 1), Figure 3.11 (Model 2) and Table 3.3 that the estimators $\widehat{m}_n(\cdot)$ and $\widehat{m}_{NW}(\cdot)$ are almost equivalent and the quality of estimation for both of them becomes better for a large sample size and a small rate of censoring, truncation and dependency. However, in the presence of outliers, the RER estimator performs better than the CR estimator in all cases, as confirmed by Table 3.4. To conclude, the quality of fit for our estimator is good but better in the one-dimensional case.

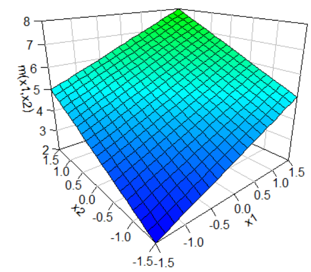

Figure 3.9: True surface for Model 1 with ρ=0.1, n=300, CP=20%, and TP=20%.

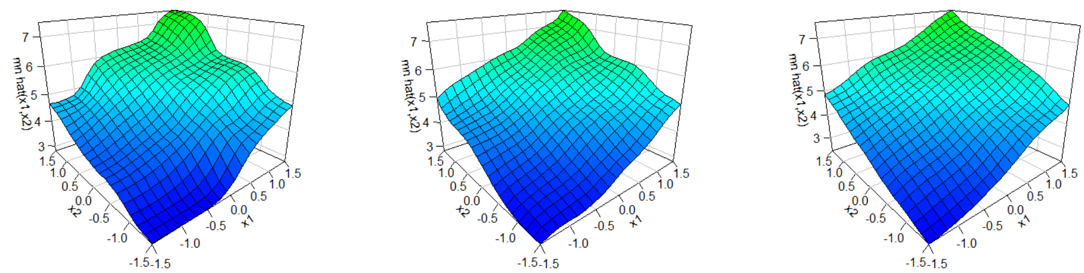

Figure 3.10: RER surface for Model 1 with ρ=0.1, CP=20%, TP=20%, n=50,100, and 300 respectively.

3.5 Real data application

In this section, we present a real data application where we assess the effectiveness of the RER estimator in the context of dependent data that is left truncated and right censored. The dataset

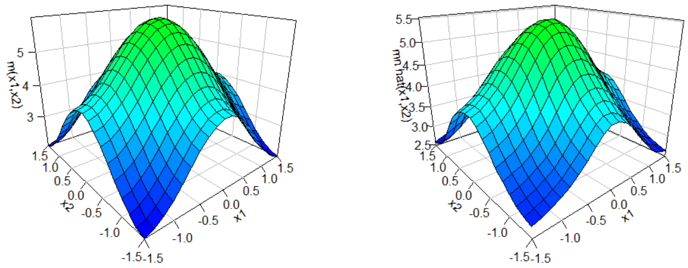

Figure 3.11: True and RER surfaces for Model 2 with ρ=0.1, n=300, CP=20%, and TP=20%.

Table 3.3: GMSE's values of $\widehat{m}_n(\cdot,\cdot)$ and $\widehat{m}_{NW}(\cdot,\cdot)$.

		CP=20%				TP=20%			
		TP=10%		TP=40%		CP=10%		CP=40%	
ρ	n	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$
0.1	100	0.06396	0.05726	0.11919	0.127636	0.05124	0.05513	0.09572	0.08368
	300	0.03566	0.03014	0.05049	0.05647	0.02567	0.02306	0.04346	0.04104
0.8	100	0.08249	0.09146	0.16514	0.15514	0.06196	0.07313	0.15381	0.14740
	300	0.05667	0.05126	0.07682	0.07011	0.04590	0.04653	0.06998	0.06332

used in our analysis contains information about patients diagnosed with AIDS in Australia before July 1^{st}, 1991, and it was obtained from Dr. Patty Solomon. For more detailed information about the dataset, we recommend consulting (Venables and Ripley (2002)).

The study focuses on patients who survived at least one year after diagnosis, comprising a total of 1276 individuals. The dataset includes various factors such as the dates of diagnosis, the dates of death, patients gender, age at diagnosis, survival status at the end of the study, as well as the state and transmission category they belong to.

Our main interest lies in analyzing the time to diagnosis and time to death. Upon examining the relationships within the data, we found a strong correlation (0.75) between these two factors, indicating their interdependence. Furthermore, the partial correlation plot suggests that our data follows an autoregressive process of order greater than 1.

To make predictions, we divided the dataset into two parts: 80% of the data served as the learning sample to calculate the estimator, while the remaining 20% formed the test sample to evaluate the predictions quality. We utilized the standard gaussian kernel function and employed the cross-

Table 3.4: GMSE's values of $\widehat{m}_n(\cdot,\cdot)$ and $\widehat{m}_{NW}(\cdot,\cdot)$ with outliers for ρ=0.1.

		CP=20%				TP=20%			
		TP=10%		TP=40%		CP=10%		CP=40%	
n	MF	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$	$\widehat{m}_n$	$\widehat{m}_{NW}$
	50	0.07123	7.81493×10^2	0.11381	1.38425×10^3	0.05571	6.68271×10^2	0.10518	9.13577×10^2
100	100	0.06744	2.64428×10^3	0.12730	4.41663×10^3	0.06166	2.11341×10^3	0.14459	3.78598×10^3
	150	0.07783	4.03922×10^3	0.12358	5.27606×10^3	0.06310	3.43390×10^3	0.13849	4.54624×10^3
	50	0.03231	5.48369×10^2	0.05916	8.24651×10^2	0.02884	3.85714×10^2	0.04664	7.21458×10^2
300	100	0.04076	2.02754×10^3	0.06805	3.68605×10^3	0.02610	1.77065×10^3	0.04454	3.22474×10^3
	150	0.03820	3.88562×10^3	0.06333	4.77619×10^3	0.02784	2.04465×10^3	0.05414	4.03390×10^3

validation method to determine the optimal bandwidth. It is important to note that around 43 % of the data was censored. Therefore, we excluded the censored data from the predicted values since it has no meaning to predict the survival time for such observations.

Figure 3.12 provides visual representations of our analysis. On the left-hand side of the figure, a scatter plot of the data is shown, distinguishing between censored and uncensored observations. On the right-hand side, the true and RER predicted values are displayed. It can be observed that the majority of the RER predicted values are close to the true values, demonstrating the robustness of our method and the accuracy of our predictor.

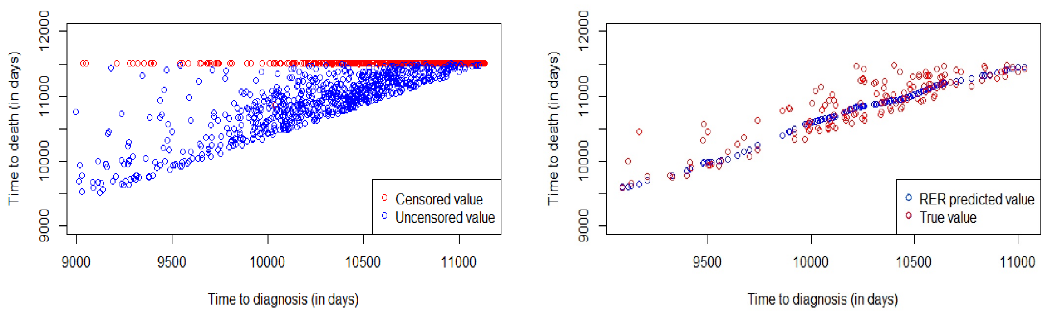

Figure 3.12: Scatter plot of time to diagnosis and time to death: censored vs. uncensored observations with true and RER predicted values for AIDS patients.

3.6 Auxiliary results and proofs

From (3.10) and (3.12), we have the following classical decomposition

$$
\begin{aligned}
&\widehat{m}_n(\mathbf{x}) - m(\mathbf{x})\\
= &\ \frac{\widehat{\psi}_{n,1}(\mathbf{x})}{\widehat{\psi}_{n,2}(\mathbf{x})} - \frac{\psi_1(\mathbf{x})}{\psi_2(\mathbf{x})}\\
= &\ \frac{1}{\widehat{\psi}_{n,2}(\mathbf{x})}\left\{\left[\widehat{\psi}_{n,1}(\mathbf{x}) - \psi_1(\mathbf{x})\right] - m(\mathbf{x})\left[\widehat{\psi}_{n,2}(\mathbf{x}) - \psi_2(\mathbf{x})\right]\right\}\\
= &\ \frac{1}{\widehat{\psi}_{n,2}(\mathbf{x})}\left\{\left[(\widehat{\psi}_{n,1}(\mathbf{x}) - \widetilde{\psi}_{n,1}(\mathbf{x})) + (\widetilde{\psi}_{n,1}(\mathbf{x}) - \mathbf{E}[\widetilde{\psi}_{n,1}(\mathbf{x})]) + (\mathbf{E}[\widetilde{\psi}_{n,1}(\mathbf{x})] - \psi_1(\mathbf{x}))\right]\right.\\
&\left. -m(\mathbf{x})\left[(\widehat{\psi}_{n,2}(\mathbf{x}) - \widetilde{\psi}_{n,2}(\mathbf{x})) + (\widetilde{\psi}_{n,2}(\mathbf{x}) - \mathbf{E}[\widetilde{\psi}_{n,2}(\mathbf{x})]) + (\mathbf{E}[\widetilde{\psi}_{n,2}(\mathbf{x})] - \psi_2(\mathbf{x}))\right]\right\}\\
=: &\ \frac{1}{\widehat{\psi}_{n,2}(\mathbf{x})}\left\{[\Lambda_{1,1}(\mathbf{x}) + \Lambda_{2,1}(\mathbf{x}) + \Lambda_{3,1}(\mathbf{x})] - m(\mathbf{x})\left[\Lambda_{1,2}(\mathbf{x}) + \Lambda_{2,2}(\mathbf{x}) + \Lambda_{3,2}(\mathbf{x})\right]\right\}.
\end{aligned}
\tag{3.14}
$$

In order to prove Theorem 3.1, some auxiliary results are needed and will be introduced in Lemma 3.1, Lemma 3.2, and Lemma 3.3 hereafter.

Lemma 3.1. *Under Assumptions **K**(a) and **D**(a), we have for $\ell = 1, 2$*

$$\sup_{\boldsymbol{x}\in\mathcal{C}} |\Lambda_{3,\ell}(\boldsymbol{x})| = O_{a.s.}(h_n^2) \text{ as } n \to \infty.$$

Proof. We have

$$
\begin{aligned}
|\Lambda_{3,\ell}(\mathbf{x})| &= \left|\mathbf{E}\left[\frac{\eta}{nh_n^d}\sum_{i=1}^n \frac{\delta_i Z_i^{-\ell}}{L(Z_i)(1-G(Z_i))} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\right] - \psi_\ell(\mathbf{x})\right|\\
&= \left|\mathbf{E}\left[\frac{\eta}{h_n^d}\frac{\delta_1 Z_1^{-\ell}}{L(Z_1)(1-G(Z_1))} K_d\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right] - \psi_\ell(\mathbf{x})\right|.
\end{aligned}
\tag{3.15}
$$

From (3.9) and using a change of variable, we get

$$
\begin{aligned}
&\mathbf{E}\left[\frac{\eta}{h_n^d}\frac{\delta_1 Z_1^{-\ell}}{L(Z_1)(1-G(Z_1))} K_d\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right]\\
= &\int_{\mathbb{R}^d}\int_{t\geq a_H} \frac{\eta t^{-\ell}}{h_n^d L(t)(1-G(t))} K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right) d\mathbf{H}_1(\mathbf{u}, t)\\
= &\int_{\mathbb{R}^d}\int_{t\geq a_H} \frac{1}{h_n^d} t^{-\ell} K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right) f_{\mathbf{X},Y}(\mathbf{u}, t)\mathbf{du}dt\\
= &\int_{\mathbb{R}^d} \frac{1}{h_n^d} K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right) \psi_\ell(\mathbf{u})\mathbf{du}\\
= &\int_{\mathbb{R}^d} K_d(\mathbf{r})\psi_\ell(\mathbf{x}-\mathbf{r}h_n)\mathbf{dr}.
\end{aligned}
$$

So, (3.15) becomes

$$\begin{aligned}
&\left|\mathbf{E}\left[\frac{\eta}{h_n^d}\frac{\delta_1 Z_1^{-\ell}}{L(Z_1)(1-G(Z_1))}K_d\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right]-\psi_\ell(\mathbf{x})\right|\\
&= \left|\int_{\mathbb{R}^d} K_d(\mathbf{r})\psi_\ell(\mathbf{x}-\mathbf{r}h_n)\mathbf{dr}-\psi_\ell(\mathbf{x})\right|\\
&= \left|\int_{\mathbb{R}^d} K_d(\mathbf{r})\left[\psi_\ell(\mathbf{x}-\mathbf{r}h_n)-\psi_\ell(\mathbf{x})\right]\mathbf{dr}\right|.
\end{aligned}$$

A Taylor expansion around $\mathbf{x}$ gives

$$\psi_\ell(\mathbf{x}-\mathbf{r}h_n)-\psi_\ell(\mathbf{x}) = -h_n\sum_{i=1}^d r_i\frac{\partial\psi_\ell(\mathbf{x})}{\partial x_i}+\frac{h_n^2}{2}\left\{\sum_{i=1}^d r_i^2\frac{\partial^2\psi_\ell(\mathbf{x}_0)}{\partial x_i^2}+2\sum_{i=1}^d\sum_{j=1}^d r_ir_j\frac{\partial^2\psi_\ell(\mathbf{x}_0)}{\partial x_i\partial x_j}\right\},$$

when $\mathbf{x}_0$ is between $\mathbf{x}-\mathbf{r}h_n$ and $\mathbf{x}$. Then

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{C}}|\Lambda_{3,\ell}(\mathbf{x})| &= \sup_{\mathbf{x}\in\mathcal{C}}\left|\int_{\mathbb{R}^d} K_d(\mathbf{r})\left[\psi_\ell(\mathbf{x}-\mathbf{r}h_n)-\psi_\ell(\mathbf{x})\right]\mathbf{dr}\right|\\
&= \sup_{\mathbf{x}\in\mathcal{C}}\left|\int_{\mathbb{R}^d} K_d(\mathbf{r})\left[-h_n\sum_{i=1}^d r_i\frac{\partial\psi_\ell(\mathbf{x})}{\partial x_i}+\frac{h_n^2}{2}\left\{\sum_{i=1}^d r_i^2\frac{\partial^2\psi_\ell(\mathbf{x}_0)}{\partial x_i^2}\right.\right.\right.\\
&\quad\left.\left.\left.+2\sum_{i=1}^d\sum_{j=1}^d r_ir_j\frac{\partial^2\psi_\ell(\mathbf{x}_0)}{\partial x_i\partial x_j}\right\}\right]\mathbf{dr}\right|\\
&\le \frac{h_n^2}{2}\left\{\sum_{i=1}^d\sup_{\mathbf{x}\in\mathcal{C}}\left|\frac{\partial^2\psi_\ell(\mathbf{x}_0)}{\partial x_i^2}\right|\int_{\mathbb{R}^d} r_i^2K_d(\mathbf{r})\mathbf{dr}\right.\\
&\quad\left.+2\sum_{i=1}^d\sum_{j=1}^d\sup_{\mathbf{x}\in\mathcal{C}}\left|\frac{\partial^2\psi_\ell(\mathbf{x}_0)}{\partial x_i\partial x_j}\right|\int_{\mathbb{R}^d}|r_i||r_j|K_d(\mathbf{r})\mathbf{dr}\right\}.
\end{aligned}$$

Assumptions $\mathbf{K}$(a) and $\mathbf{D}$(a) complete the proof of the lemma. □

Lemma 3.2. *Under Assumptions **K**(a)-**K**(c), **H**(c), **D**(a)-**D**(d), and **M**(a), we have for* $\ell = 1, 2$

$$\sup_{\boldsymbol{x}\in\mathcal{C}}|\Lambda_{2,\ell}(\boldsymbol{x})| = O_{a.s.}\left(\sqrt{\frac{\log n}{nh_n^d}}+\sqrt{h_n^{d(v-2)}\log n}\right)\quad \text{as } n\to\infty.$$

Proof. Since $\mathcal{C}$ is a compact set, then it can be covered by a finite number ω_n of balls $\mathcal{B}_k(\mathbf{x}_k, a_n^d)$ centred at $\mathbf{x}_k = (x_{1,k},\ldots,x_{d,k})$, $1\le k\le\omega_n$, where ω_n and a_n^d satisfy

$$\omega_n \le Ma_n^{-d} \text{ and } a_n^d = h_n^{d(1+\frac{1}{2\gamma})}n^{-\frac{1}{2\gamma}},$$

with M is a positive constant and γ is the Lipschitz condition in Assumption $\mathbf{K}$(c). Then for all $\mathbf{x}\in\mathcal{C}$, there exists a ball B_k that contains $\mathbf{x}$ such that

$$\parallel \mathbf{x}-\mathbf{x}_k \parallel\le a_n^d. \tag{3.16}$$

For $\ell = 1, 2$, we have

$$\begin{aligned}
&\sup_{\mathbf{x}\in\mathcal{C}} |\Lambda_{2,\ell}(\mathbf{x})| \\
= &\sup_{\mathbf{x}\in\mathcal{C}} \left|(\widetilde{\psi}_{n,\ell}(\mathbf{x}) - \widetilde{\psi}_{n,\ell}(\mathbf{x}_k)) + (\mathbf{E}[\widetilde{\psi}_{n,\ell}(\mathbf{x}_k)] - \mathbf{E}[\widetilde{\psi}_{n,\ell}(\mathbf{x})]) + (\widetilde{\psi}_{n,\ell}(\mathbf{x}_k) - \mathbf{E}[\widetilde{\psi}_{n,\ell}(\mathbf{x}_k)])\right| \\
\leq &\max_{1\leq k\leq\omega_n} \sup_{\mathbf{x}\in\mathcal{C}} \left|\widetilde{\psi}_{n,\ell}(\mathbf{x}) - \widetilde{\psi}_{n,\ell}(\mathbf{x}_k)\right| + \max_{1\leq k\leq\omega_n} \sup_{\mathbf{x}\in\mathcal{C}} \left|\mathbf{E}[\widetilde{\psi}_{n,\ell}(\mathbf{x}_k)] - \mathbf{E}[\widetilde{\psi}_{n,\ell}(\mathbf{x})]\right| \\
&+ \max_{1\leq k\leq\omega_n} \left|\widetilde{\psi}_{n,\ell}(\mathbf{x}_k) - \mathbf{E}[\widetilde{\psi}_{n,\ell}(\mathbf{x}_k)]\right| \\
=: &\; I_1 + I_2 + I_3.
\end{aligned}$$

We start by treating the first term I_1. Under Assumptions $\mathbf{D}$(b) and $\mathbf{K}$(c), we have

$$\begin{aligned}
\left|\widetilde{\psi}_\ell(\mathbf{x}) - \widetilde{\psi}_\ell(\mathbf{x}_k)\right| &= \left|\frac{\eta}{nh_n^d}\sum_{i=1}^n \frac{\delta_i Z_i^{-\ell}}{L(Z_i)G(\bar{Z}_i)}\left[K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) - K_d\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n}\right)\right]\right| \\
&= \left|\frac{\eta}{nh_n^d}\sum_{i=1}^n \frac{Y_i^{-\ell}}{L(Y_i)\bar{G}(Y_i)}\left[K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) - K_d\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n}\right)\right]\right| \\
&\leq \frac{C}{nh_n^d L(a_H)\bar{G}(b_H)}\sum_{i=1}^n \left|K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) - K_d\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n}\right)\right| \\
&\leq \frac{C}{h_n^d L(a_H)\bar{G}(b_H)}\left\|\frac{\mathbf{x}-\mathbf{X}_1}{h_n} - \frac{\mathbf{x}_k-\mathbf{X}_1}{h_n}\right\|^\gamma \\
&\leq \frac{C \parallel \mathbf{x}-\mathbf{x}_k \parallel^\gamma}{h_n^{d+\gamma}}.
\end{aligned}$$

Then, from (3.16), we get

$$\sup_{\mathbf{x}\in\mathcal{C}}\left|\widetilde{\psi}_\ell(\mathbf{x}) - \widetilde{\psi}_\ell(\mathbf{x}_k)\right| \leq \frac{Ca_n^{d\gamma}}{h_n^{d+\gamma}} = \frac{C}{\sqrt{nh_n^d}}h_n^{\gamma(d-1)}.$$

Hence,

$$I_1 = \max_{1\leq k\leq\omega_n}\sup_{\mathbf{x}\in\mathcal{C}}\left|\widetilde{\psi}_\ell(\mathbf{x}) - \widetilde{\psi}_\ell(\mathbf{x}_k)\right| = O\left(\frac{1}{\sqrt{nh_n^d}}\right).$$

In the same way as for I_1, we obtain

$$I_2 = \max_{1\leq k\leq\omega_n}\sup_{\mathbf{x}\in\mathcal{C}}\left|\mathbf{E}[\widetilde{\psi}_\ell(\mathbf{x})] - \mathbf{E}[\widetilde{\psi}_\ell(\mathbf{x}_k)]\right| = O\left(\frac{1}{\sqrt{nh_n^d}}\right).$$

For I_3, we use the Fuk-Nagaev exponential inequality given in Proposition 1.2. For that, we set

$$U_{i,\ell}(\mathbf{x}_k) = \frac{\eta Z_i^{-\ell}\delta_i}{L(Z_i)\bar{G}(Z_i)}K_d\left(\frac{\mathbf{x}_k-\mathbf{X}_i}{h_n}\right) - \mathbf{E}\left[\frac{\eta Z_1^{-\ell}\delta_1}{L(Z_1)\bar{G}(Z_1)}K_d\left(\frac{\mathbf{x}_k-\mathbf{X}_1}{h_n}\right)\right], \text{ for } \ell = 1, 2.$$

Clearly, we have

$$\Lambda_{2,\ell}(\mathbf{x}_k) = \frac{1}{nh_n^d}\sum_{i=1}^n U_{\ell,i}(\mathbf{x}_k).$$

Now, we have to calculate

$$\begin{aligned} S_n^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}|\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k),U_{j,\ell}(\mathbf{x}_k))| \\ &= n\mathbf{Var}(U_{1,\ell}(\mathbf{x}_k))+\sum_{\substack{i=1\\ i\neq j}}^{n}\sum_{j=1}^{n}|\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k),U_{j,\ell}(\mathbf{x}_k))| \\ &=: CV_1+CV_2. \end{aligned}$$

On the one hand, we have

$$\begin{aligned} &\mathbf{Var}(U_{1,\ell}(\mathbf{x}_k)) \\ &= \mathbf{Var}\left[\frac{\eta_1 Z_1^{-\ell}\delta_1}{L(Z_1)\bar{G}(Z_1)}K_d\left(\frac{\mathbf{x}_k-\mathbf{X}_1}{h_n}\right)\right] \\ &= \mathbf{E}\left[\frac{\eta^2 Z_1^{-2\ell}\delta_1}{L^2(Z_1)\bar{G}^2(Z_1)}K_d^2\left(\frac{\mathbf{x}_k-\mathbf{X}_1}{h_n}\right)\right]-\mathbf{E}^2\left[\frac{\eta Z_1^{-\ell}\delta_1}{L(Z_1)\bar{G}(Z_1)}K_d\left(\frac{\mathbf{x}_k-\mathbf{X}_1}{h_n}\right)\right] \\ &=: \mathcal{R}_1-\mathcal{R}_2. \end{aligned}$$

From (3.9) and using a change of variable, a Taylor expansion and under Assumptions **K**(b) and **D**(c), we obtain for $\lambda=2\ell$ with $\ell=1,2$

$$\begin{aligned} \mathcal{R}_1 &= \mathbf{E}\left[\frac{\eta^2 Z_1^{-2\ell}\delta_1}{L^2(Z_1)\bar{G}^2(Z_1)}K_d^2\left(\frac{\mathbf{x}_k-\mathbf{X}_1}{h_n}\right)\right] \\ &= \int_{\mathbb{R}^d}\int_{t\geq a_H}\frac{\eta^2 t^{-2\ell}}{L^2(t)\bar{G}^2(t)}K_d^2\left(\frac{\mathbf{x}_k-\mathbf{u}}{h_n}\right)d\mathbf{H}_1(\mathbf{u},t) \\ &= \int_{\mathbb{R}^d}\int_{t\geq a_H}\frac{\eta t^{-2\ell}}{L(t)\bar{G}(t)}K_d^2\left(\frac{\mathbf{x}_k-\mathbf{u}}{h_n}\right)f_{\mathbf{X},Y}(\mathbf{u},t)\mathbf{du}dt \\ &\leq \int_{\mathbb{R}^d}K_d^2\left(\frac{\mathbf{x}_k-\mathbf{u}}{h_n}\right)\mu_\lambda(\mathbf{u})\mathbf{du} \\ &\leq h_n^d\int_{\mathbb{R}^d}K_d^2(\mathbf{s})\mu_\lambda(\mathbf{x}_k-\mathbf{s}h_n)\mathbf{ds} \\ &= O(h_n^d). \end{aligned} \tag{3.17}$$

For $\mathcal{R}_2$, we have

$$\begin{aligned} \sqrt{\mathcal{R}_2} &= \mathbf{E}\left[\frac{\eta Z_1^{-\ell}\delta_1}{L(Z_1)\bar{G}(Z_1)}K_d\left(\frac{\mathbf{x}_k-\mathbf{X}_1}{h_n}\right)\right] \\ &= \int_{\mathbb{R}^d}\int_{t\geq a_H}\frac{\eta t^{-\ell}}{L(t)\bar{G}(t)}K_d\left(\frac{\mathbf{x}_k-\mathbf{u}}{h_n}\right)d\mathbf{H}_1(\mathbf{u},t) \\ &\leq \int_{\mathbb{R}^d}\int_{t\geq a_H}t^{-\ell}K_d\left(\frac{\mathbf{x}_k-\mathbf{u}}{h_n}\right)f_{\mathbf{X},Y}(\mathbf{u},t)\mathbf{du}dt \\ &\leq \int_{\mathbb{R}^d}K_d\left(\frac{\mathbf{x}_k-\mathbf{u}}{h_n}\right)\psi_\ell(\mathbf{u})\mathbf{du} \\ &\leq h_n^d\int_{\mathbb{R}^d}K_d(\mathbf{s})\psi_\ell(\mathbf{x}_k-\mathbf{s}h_n)\mathbf{ds}, \end{aligned}$$

by a Taylor expansion around $\mathbf{x}_k$ and under Assumptions $\mathbf{K}$(a), $\mathbf{D}$(a), we get

$$\mathcal{R}_2 = O(h_n^{2d}). \tag{3.18}$$

Then, from (3.17) and (3.18), we obtain

$$CV_1 = n(\mathcal{R}_1 - \mathcal{R}_2) = O(nh_n^d). \tag{3.19}$$

On the other hand, under Assumption $\mathbf{D}$(b) and by a change of variable, we have

$$\begin{aligned}
&|\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| \\
=\ & |\mathbf{E}(U_{i,\ell}(\mathbf{x}_k)U_{j,\ell}(\mathbf{x}_k))| \\
=\ & \Bigg|\mathbf{E}\left[\frac{\eta Z_i^{-\ell}\delta_i}{L(Z_i)\bar{G}(Z_i)}K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n}\right)\frac{\eta Z_j^{-\ell}\delta_j}{L(Z_j)\bar{G}(Z_j)}K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_j}{h_n}\right)\right] \\
& - \mathbf{E}\left[\frac{\eta Z_i^{-\ell}\delta_i}{L(Z_i)\bar{G}(Z_i)}K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_i}{h_n}\right)\right]\mathbf{E}\left[\frac{\eta Z_j^{-\ell}\delta_j}{L(Z_j)\bar{G}(Z_j)}K_d\left(\frac{\mathbf{x}_k - \mathbf{X}_j}{h_n}\right)\right]\Bigg| \\
\leq\ & Ch_n^{2d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K_d(\mathbf{s})K_d(\mathbf{t})|\Upsilon_{ij}(\mathbf{x}_k - \mathbf{s}h_n, \mathbf{x}_k - \mathbf{t}h_n) \\
& - \Upsilon_i(\mathbf{x}_k - \mathbf{s}h_n)\Upsilon_j(\mathbf{x}_k - \mathbf{t}h_n)|\mathbf{dsdt}.
\end{aligned}$$

Assumption $\mathbf{D}$(d) gives

$$|\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| = O(h_n^{2d}). \tag{3.20}$$

Then, to evaluate the term CV_2, following Masry (1986), we divide the set $\{(i,j)/1 \leq |i-j| \leq n\}$ into two sub-sets E_1 and E_2 by introducing a sequence of integer $\beta_n = o(n)$, such that

$$E_1 = \{(i,j)/1 \leq |i-j| \leq \beta_n\} \text{ and } E_2 = \{(i,j)/\beta_n + 1 \leq |i-j| \leq n-1\}.$$

That is

$$\begin{aligned}
CV_2 &= \sum_{i=1}^{n}\sum_{j\in E_1}|\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| + \sum_{i=1}^{n}\sum_{j\in E_2}|\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| \\
&=: CV_{21} + CV_{22}.
\end{aligned}$$

From (3.20), we get

$$CV_{21} = \sum_{i=1}^{n}\sum_{j\in E_1}|\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| = C\sum_{i=1}^{n}\sum_{1\leq|i-j|\leq\beta_n} h_n^{2d} = O(nh_n^{2d}\beta_n).$$

For CV_{22}, we use the modified inequality of Davydov for mixing processes (See, Rio (2000)). This leads, for all $i \neq j$, to

$$|\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| \leq C\alpha(|i-j|).$$

Then, from Assumption $\mathbf{M}$(a) we get

$$\begin{aligned} CV_{22} = \sum_{i=1}^{n}\sum_{j\in E_2} |\mathbf{Cov}(U_{i,\ell}(\mathbf{x}_k), U_{j,\ell}(\mathbf{x}_k))| &\leq C\sum_{i=1}^{n}\sum_{\beta_n<|i-j|<n} \alpha(|i-j|) \\ &\leq Cn^2\alpha(\beta_n) \\ &= O(n^2\beta_n^{-v}). \end{aligned}$$

By choosing $\beta_n = [h_n^{-d}] + 1$, we obtain

$$CV_2 = CV_{21} + CV_{22} = O(nh_n^d) + O(n^2h_n^{dv}). \tag{3.21}$$

Finally, from (3.19) and (3.21), we get

$$S_n^2 = CV_1 + CV_2 = O(nh_n^d) + O(n^2h_n^{dv}).$$

Now, we are ready to apply the Fuk-Nagaev exponential inequality. For $\varepsilon > 0$, we have

$$\begin{aligned} &\mathbf{P}\{|\widetilde{\psi}_\ell(\mathbf{x}_k) - \mathbf{E}[\widetilde{\psi}_\ell(\mathbf{x}_k)]| > \varepsilon\} \\ =\ & \mathbf{P}\left\{\left|\sum_{i=1}^{n} U_{i,\ell}(\mathbf{x}_k)\right| > nh_n^d\varepsilon\right\} \\ \leq\ & C\left(1 + C\frac{\varepsilon^2nh_n^d}{q(1+nh_n^{d(v-1)})}\right)^{-\frac{q}{2}} + nCq^{-1}\left(\frac{q}{\varepsilon nh_n^d}\right)^{v+1}. \end{aligned} \tag{3.22}$$

By taking $\varepsilon = \varepsilon_0\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(v-2)}\log n}\right) =: \varepsilon_n$ for all $\varepsilon_0 > 0$, (3.22) becomes

$$\begin{aligned} &\mathbf{P}\{|\widetilde{\psi}_\ell(\mathbf{x}_k) - \mathbf{E}[\widetilde{\psi}_\ell(\mathbf{x}_k)]| > \varepsilon_n\} \\ \leq\ & C\left\{\left(1 + C\frac{\varepsilon_0^2\log n}{q}\right)^{-\frac{q}{2}} + nq^{-1}\left(\frac{q}{\varepsilon_0\left(\sqrt{\frac{\log n}{nh_n^d}} + \sqrt{h_n^{d(v-2)}\log n}\right)nh_n^d}\right)^{v+1}\right\} \\ =:\ & C(\varepsilon_1 + \varepsilon_2). \end{aligned}$$

If we replace q by $(\log n)^{1+b}$, with $b > 0$ and use a Taylor expansion of $\log(1+x)$, we get

$$\begin{aligned} \varepsilon_1 &= (1 + C\varepsilon_0^2(\log n)^{-b})^{-\frac{(\log n)^{1+b}}{2}} \\ &= \exp\left(-\frac{(\log n)^{1+b}}{2}\log(1 + C\varepsilon_0^2(\log n)^{-b})\right) \\ &\simeq n^{-C\frac{\varepsilon_0^2}{2}}. \end{aligned}$$

For the same choice of ε and q, we have

$$\varepsilon_2 \simeq n(\log n)^{v(1+b)}\varepsilon_0^{-(v+1)}(nh_n^d\log n)^{-\frac{(v+1)}{2}}.$$

Now, we can write

$$\begin{aligned}
& \mathbf{P}\{\max_{1\leq k\leq \omega_n} |\widetilde{\psi}_\ell(\mathbf{x}_k) - \mathbf{E}[\widetilde{\psi}_\ell(\mathbf{x}_k)]| > \varepsilon_n\} \\
\leq & \sum_{i=1}^{\omega_n} \mathbf{P}\{|\widetilde{\psi}_\ell(\mathbf{x}_k) - \mathbf{E}[\widetilde{\psi}_\ell(\mathbf{x}_k)]| > \varepsilon_n\} \\
\leq & M a_n^{-d} \left\{ C n^{-C\frac{\varepsilon_0^2}{2}} + n(\log n)^{v(1+b)} \varepsilon_0^{-(v+1)} (n h_n^d \log n)^{-\frac{(v+1)}{2}} \right\} \\
\leq & M C h_n^{-d(1+\frac{1}{2\gamma})} n^{\frac{1}{2\gamma} - C\frac{\varepsilon_0^2}{2}} \\
& + M C \varepsilon_0^{-(v+1)} n^{1+\frac{1}{2\gamma}} h_n^{-d(1+\frac{1}{2\gamma})} (\log n)^{v(1+b)} (n h_n^d \log n)^{-\frac{v+1}{2}} \\
=: & M C A_1 + M C \varepsilon_0^{-(v+1)} A_2.
\end{aligned} \tag{3.23}$$

We have from Assumption $\mathbf{H}$(c)

$$\begin{aligned}
A_2 &= (\log n)^{v(1+b) - \frac{v+1}{2}} n^{1-\frac{v+1}{2}+\frac{1}{2\gamma}} h_n^{-d\left(1+\frac{1}{2\gamma}+\frac{v+1}{2}\right)} \\
&\leq C (\log n)^{v(1+b) - \frac{v+1}{2}} n^{1-\frac{v+1}{2}+\frac{1}{2\gamma}} n^{-\frac{(3-v)}{2} - \theta d\left(\frac{\gamma(v+1)+2\gamma+1}{2\gamma}\right)} \\
&\leq C (\log n)^{v(1+b) - \frac{v+1}{2}} n^{-1+\frac{1-\theta d(\gamma(v+3)+1)}{2\gamma}}.
\end{aligned}$$

Then, for an appropriate choice of ε_0 and θ, A_1 and A_2 are the general terms of a convergent series. Finally, applying Borel-Cantelli's lemma to (3.23) gives the result. □

Lemma 3.3. *Under Assumptions **K**, **H**(a), and **D**(b), we have for* $\ell = 1, 2$

$$\sup_{\boldsymbol{x}\in\mathcal{C}} |\Lambda_{1,\ell}(\boldsymbol{x})| = O_{a.s.}\left(\sqrt{\frac{\log\log n}{n}}\right) \quad \text{as } n \to \infty.$$

Proof. We have

$$\begin{aligned}
\Lambda_{1,\ell}(\mathbf{x}) &= \frac{1}{n h_n^d} \sum_{i=1}^n \left(\frac{\eta_n Z_i^{-\ell} \delta_i}{L_n(Z_i) G_n(\bar{Z}_i)} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) - \frac{\eta Z_i^{-\ell}\delta_i}{L(Z_i) G(\bar{Z}_i)} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\right) \\
&= \frac{1}{n h_n^d} \sum_{i=1}^n Y_i^{-\ell} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) \left(\frac{\eta_n}{L_n(Z_i)\bar{G}_n(Z_i)} - \frac{\eta}{L(Z_i)\bar{G}(Z_i)} \right).
\end{aligned}$$

By replacing η and η_n, by their expressions defined in (3.3) and (3.4), respectively, we get

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{C}}|\Lambda_{1,\ell}(\mathbf{x})| &\leq \sup_{a_H\leq t\leq b_H}\left|\frac{\bar{F}_n(t)}{C_n(t)}-\frac{\bar{F}(t)}{C(t)}\right|\times\sup_{\mathbf{x}\in\mathcal{C}}\left|\frac{1}{nh_n^d}\sum_{i=1}^{n}Y_i^{-\ell}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\right| \\
&\leq \left\{\sup_{a_H\leq t\leq b_H}\left|\frac{F_n(t)-F(t)}{C_n(t)}\right|+\sup_{a_H\leq t\leq b_H}\left|\frac{\bar{F}(t)}{C_n(t)C(t)}(C_n(t)-C(t))\right|\right\} \\
&\quad\times\sup_{x\in\mathcal{C}}\left|\frac{1}{nh_n^d}\sum_{i=1}^{n}Y_i^{-\ell}K_d\left(\frac{x-X_i}{h_n}\right)\right| \\
&\leq \left\{\frac{\sup_{a_H\leq t\leq b_H}|F_n(t)-F(t)|}{\inf_{a_H\leq t\leq b_H}|C(t)|-\sup_{a_H\leq t\leq b_H}|(C_n(t)-C(t))|}\right. \\
&\quad\left.+\frac{\sup_{a_H\leq t\leq b_H}|(C_n(t)-C(t))|}{\inf_{a_H\leq t\leq b_H}|C(t)|\left(\inf_{a_H\leq t\leq b_H}|C(t)|-\sup_{a_H\leq t\leq b_H}|(C_n(t)-C(t))|\right)}\right\} \\
&\quad\times\sup_{\mathbf{x}\in\mathcal{C}}\left|\frac{1}{nh_n^d}\sum_{i=1}^{n}Y_i^{-\ell}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\right|.
\end{aligned}$$

From (3.1), we have $C(t)\geq\alpha^{-1}L(a_H)\bar{H}(b_H)>0$ for all $a_H\leq t\leq b_H$, and following Chen and Dai (2003), we have

$$\sup_{a_H\leq t\leq b_H}|F_n(t)-F(t)|=O_{\text{a.s.}}\left(\sqrt{\frac{\log\log n}{n}}\right)\quad\text{as } n\to\infty$$

and

$$\sup_{a_H\leq t\leq b_H}|C_n(t)-C(t)|=O_{\text{a.s.}}\left(\sqrt{\frac{\log\log n}{n}}\right)\quad\text{as } n\to\infty.$$

Furthermore, under Assumptions **K**, **H**(a) and **D**(b) we get the result. □

Proof of Theorem 3.1. By triangle inequality, (3.14) becomes

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{C}}|\widehat{m}_n(\mathbf{x})-m(\mathbf{x})| &\leq \frac{1}{\inf_{\mathbf{x}\in\mathcal{C}}|\widehat{\psi}_2(\mathbf{x})|}\left\{\sup_{\mathbf{x}\in\mathcal{C}}\left[|\Lambda_{1,1}(\mathbf{x})|+|\Lambda_{2,1}(\mathbf{x})|+|\Lambda_{3,1}(\mathbf{x})|\right]\right. \\
&\quad\left.+\sup_{\mathbf{x}\in\mathcal{C}}|m(\mathbf{x})|\left[|\Lambda_{1,2}(\mathbf{x})|+|\Lambda_{2,2}(\mathbf{x})|+|\Lambda_{3,2}(\mathbf{x})|\right]\right\}.
\end{aligned}$$

Then, the proof of Theorem 3.1 is completed by Lemma 3.1, Lemma 3.2 and Lemma 3.3. □

Chapter 4

Asymptotic normality of the relative regression function estimator: LTRC and α-mixing data

In this chapter, we establish the asymptotic normality and construct confidence intervals of a relative regression function estimator for LTRC and α-mixing data. Additionally, we validate the theoretical findings through a simulation study, which is further supported by a real world data application.

4.1 Introduction

This chapter extends the work presented in Chapter 3 on nonparametric estimation of the relative error regression function for left truncated and right censored data. The asymptotic property we address here is asymptotic normality, a crucial topic in statistics. In particular, asymptotic normality allows for the construction of confidence intervals and the performance of hypothesis tests.

To recall our working framework, we consider a sequence of random vectors $\{(Y_i, \mathbf{X}_i, T_i, W_i);\ i = 1, \ldots, N\}$ drawn from $(Y, \mathbf{X}, T, W)$, representing the response variable, explanatory variable, left truncation variable, and right censoring variable, respectively. Under the random left truncation and right censoring model, our observations are $\{(Z_i, T_i, \delta_i);\ i = 1, \ldots, n\}$ such that $Z_i \geq T_i$, where $Z_i = Y_i \wedge W_i$ and $\delta_i = \mathbb{1}_{\{Y_i \leq W_i\}}$. Based on these observations, Bayarassou et al. (2023) proposed a nonparametric estimator for the regression function $m(\cdot)$ given by

$$\widehat{m}_n(\mathbf{x}) =: \frac{\widehat{\psi}_{n,1}(\mathbf{x})}{\widehat{\psi}_{n,2}(\mathbf{x})},$$

where

$$\widehat{\psi}_{n,\ell}(\mathbf{x}) = \frac{\eta_n}{n h_n^d} \sum_{i=1}^{n} \frac{\delta_i Z_i^{-\ell}}{L_n(Z_i)\bar{G}_n(Z_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right), \text{ for } \ell = 1, 2.$$

The structure of the rest of this chapter is organized as follows: Section 4.2 presents the assumptions and the main results. Section 4.3 contains a simulation study, with a real data application provided in Section 4.4. The proofs are included in Section 4.5.

4.2 Assumptions and main results

Before stating the asymtotic normality result for the estimator $\widehat{m}_n(\cdot)$ derived from the strictly α-mixing sequence of random vectors $\{(X_i, Z_i, T_i, \delta_i); i = 1, \ldots, n\}$, we introduce some necessary notations. We define

$$\widetilde{\psi}_{n,\ell}(\mathbf{x}) = \frac{\eta}{nh_n^d} \sum_{i=1}^{n} \frac{\delta_i Z_i^{-\ell}}{L(Z_i)\bar{G}(Z_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right), \text{ for } \ell = 1, 2,$$

and

$$\mu_\lambda(\mathbf{u}) = \int_{t \geq a_H} \frac{t^{-\lambda}}{L(t)\bar{G}(t)} f_{\mathbf{X},Y}(\mathbf{u}, t) dt, \text{ for } \lambda = 2, 3, 4.$$

Furthermore, let $\mathcal{C}$ be a compact set in $\mathbb{R}^d$, and when no confusion is possible we will denote by C any generic constant.

4.2.1 Assumptions

K. The kernel $K_d(\cdot)$ is a bounded probability density with compact support and satisfies:

(a) $\int_{\mathbb{R}^d} K_d(\mathbf{r})\mathbf{dr} = 1$, $\int_{\mathbb{R}^d} r_i K_d(\mathbf{r})\mathbf{dr} = 0$ and $\int_{\mathbb{R}^d} |r_i||r_j| K_d(\mathbf{r})\mathbf{dr} < \infty$, for $i, j = 1, \ldots, d$.

(b) $\int_{\mathbb{R}^d} r_i K_d^2(\mathbf{r})\mathbf{dr} < \infty$, for $i = 1, \ldots, d$.

H. The bandwidth h_n satisfies:

(a) $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n^d = \infty$.

(b) $\lim_{n\to\infty} h_n^d \text{loglog} n = 0$.

(c) $\lim_{n\to\infty} nh_n^{d+4} = 0$.

D. (a) The function $\psi_\ell(\cdot)$ is twice continuously differentiable and $\sup_{\mathbf{x}\in\mathcal{C}} \left|\frac{\partial^2 \psi_\ell(\mathbf{x})}{\partial x_i \partial x_j}\right| < \infty$, for $\ell = 1, 2$ and $i, j = 1, \ldots, d$.

(b) $\forall Y > 0$, $\exists C > 0$, such that $Y^{-\ell} \leq C$, for $\ell = 1, 2$.

(c) The function $\mu_\lambda(\cdot)$ is continuously differentiable and $\sup_{\mathbf{x}\in\mathcal{C}} \left|\frac{\partial \mu_\lambda(\mathbf{x})}{\partial x_i}\right| < \infty$, for $\lambda = 2, 3, 4$ and $i = 1, \ldots, d$.

(d) The joint density $\Upsilon_{i,j}(\cdot,\cdot)$ of $(\mathbf{X}_i, \mathbf{X}_j)$ exists and satisfies

$$\sup_{\mathbf{t},\mathbf{s}\in\mathcal{C}} |\Upsilon_{i,j}(\mathbf{t},\mathbf{s}) - \Upsilon_i(\mathbf{t})\Upsilon_j(\mathbf{s})| < \infty, \text{ for } i,j = 1,\ldots,n.$$

M. $\{(X_i, Z_i, T_i, \delta_i); i = 1,\ldots,n\}$ is a stationary α-mixing sequence of random vectors, with strong mixing coefficient $\alpha(n)$. This coefficient satisfies:

(a) $\alpha(n) = O(n^{-v})$, $v > 3$.

(b) $\lim_{n\to\infty} h_n^{-d/a_1} \sum_{s>\beta_n} \alpha^{1/a_1}(s) = 0$, for $0 < a_1 < 1$ and $d > 1$.
Here, β_n is a sequence of positive integers going to infinity which will be defind later.

B. Let $0 < p := p_n < n$ and $0 < q := q_n < n$ be sequences of positive integers tending to infinity as $n \to \infty$. These sequences satisfy:

(a) $\lim_{n\to\infty} \frac{k(p+q)}{n} \to 1$.

(b) $\lim_{n\to\infty} \frac{kp}{n} \to 1$.

(c) $\lim_{n\to\infty} ph_n^d \to 0$ and $\lim_{n\to\infty} qh_n^d \to 0$.

(d) $\lim_{n\to\infty} \frac{p^2}{nh_n^d} \to 0$.

4.2.1.1 Comments on the assumptions

Assumptions **K**, **H**(a), and **D**(a) are standard in the context of nonparametric estimation. Assumptions **H**(b) and **H**(b) are critical for ensuring that certain terms become negligible. Assumptions **D**(b), **D**(c), **D**(d), and **M** are technical in nature and play a crucial role in deriving our results. Assumptions **B** are employed in applying the large blocks and small blocks procedure to establish asymptotic normality.

4.2.2 Main results

The principal result regarding the asymptotic normality of the estimator $\widehat{m}_n(\cdot)$ is outlined in the following theorem.

Theorem 4.1. *Under Assumptions **K**, **H**, **D**, **M** and **B**, we have:*

$$\sqrt{nh_n^d}\left(\widehat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_n^2(\boldsymbol{x})) \text{ as } n \to \infty,$$

where $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution and*

$$\sigma_n^2(\boldsymbol{x}) = \frac{\psi_2^2(\boldsymbol{x})\mu_2(\boldsymbol{x}) - 2\psi_1(\boldsymbol{x})\psi_2(\boldsymbol{x})\mu_3(\boldsymbol{x}) + \psi_1^2(\boldsymbol{x})\mu_4(\boldsymbol{x})}{\psi_2^4(\boldsymbol{x})}\kappa,$$

with $\kappa = \int_{\mathbb{R}^d} K_d^2(\boldsymbol{s})d\boldsymbol{s}$.

Remark 4.1. *A calculable estimator $\widehat{\sigma}_n^2(\boldsymbol{x})$ for the asymptotic variance $\sigma_n^2(\boldsymbol{x})$ can be obtained using a plug-in method. This involves replacing $\psi_\ell(\cdot)$ with its estimator $\widehat{\psi}_\ell(\cdot)$ for $\ell = 1, 2$ and $\mu_\lambda(\cdot)$ with*

$$\widehat{\mu}_\lambda(\boldsymbol{x}) = \frac{\eta_n}{nh_n^d} \sum_{i=1}^{n} \frac{\delta_i Z_i^{-\lambda}}{L_n^2(Z_i)\bar{G}_n^2(Z_i)} K_d\left(\frac{\boldsymbol{x} - X_i}{h_n}\right), \text{ for } \lambda = 2, 3, 4.$$

As an immediate application of Theorem 4.1, we obtain the following corollary.

Corollary 4.1. *Under the assumptions of Theorem 4.1 and for each fixed $\zeta \in (0, 1)$, a confidence interval of level $(1 - \zeta)$ for $m(\boldsymbol{x})$ is*

$$\left[\widehat{m}_n(\boldsymbol{x}) - t_{1-\frac{\zeta}{2}} \frac{\widehat{\sigma}_n(\boldsymbol{x})}{\sqrt{nh_n^d}}, \widehat{m}_n(\boldsymbol{x}) + t_{1-\frac{\zeta}{2}} \frac{\widehat{\sigma}_n(\boldsymbol{x})}{\sqrt{nh_n^d}}\right],$$

where $t_{1-\frac{\zeta}{2}}$ denotes the $\left(1 - \frac{\zeta}{2}\right)$-quantile of the standard normal distribution.

4.3 Simulation study

To examine the asymptotic normality of the estimator $\widehat{m}_n(\cdot)$, we conduct a simulation study focusing on a one-dimensional covariate scenario (i.e., d=1). We generate a bivariate α-mixing process denoted by $\{(X_t, Y_t); t \geq 1\}$, as follows:

$$\begin{cases} X_1 = e_1, \\ X_{t+1} = \begin{cases} \rho X_t + 0.5 e_{t+1}, & \text{if } \rho > 0.5 \\ \rho X_t + e_{t+1}, & \text{else} \end{cases} \\ Y_t = m(X_t) + \epsilon_t, \end{cases}$$

where $e_t \rightsquigarrow N(0, 1)$, $0 < \rho < 1$ controls the degree of dependency, and the white noise $\epsilon_t \rightsquigarrow N(0, 0.1)$. We define $Z_t = Y_t \wedge W_t$ and $\delta_t = \mathbb{1}_{\{Y_t \leq W_t\}}$, $t \geq 1$, where W_t is drawn from an exponential distribution with parameter a_0, influencing the censoring percentage (CP). We also generate $T_t \rightsquigarrow N(b_0, 2)$ for $t \geq 1$, where b_0 controls the truncation percentage (TP). We then retain the observations $\{(X_i, Z_i, T_i, \delta_i); i = 1, \ldots, n\}$ satisfying $Z_i \geq T_i$. From these observations, we calculate the values of the estimator $\widehat{m}_n(x)$ for $x \in \mathcal{C} = [-1.5, 1.5]$. We use a Gaussian kernel $K(\cdot)$ and select the bandwidth from a grid of values $\mathcal{H}$ that minimizes the global mean squared error (GMSE):

$$\underset{h_n \in \mathcal{H}}{GMSE(h_n)} = \frac{1}{A \cdot B} \sum_{j=1}^{B} \sum_{i=1}^{A} (\widehat{m}_{n,j}(x_i) - m(x_i))^2,$$

for B = 50 replications, where $\widehat{m}_{n,j}(\cdot)$ represents the estimator at iteration j and A is the number of equidistant points x_i within the compact set $\mathcal{C} = [-1.5, 1.5]$.

First, we assess the asymptotic normality of the estimator by comparing the shape of the estimated density to that of the standard normal density. We estimate the regression function $m(x) = \exp(x)+3$ by $\widehat{m}_n(x)$, and calculate the normalized deviation between this estimator and its theoretical regression function for $x = 0$, as follows:

$$m_n^\star = m_n^\star(0) := \frac{\sqrt{nh_n}}{\widehat{\sigma}_n(0)}(\widehat{m}_n(0) - m(0)).$$

Following this procedure, we generate a sequence of size $B = 100$. We then estimate its density using the kernel method with a Gaussian kernel and the bandwidth $h_B = C \cdot B^{-\frac{1}{5}}$ (as suggested in Silverman (1986)), where the constant C is chosen appropriately. Subsequently, we plot the estimated density $\widehat{m}_n^\star$ along with the standard normal density for various values of n and for different censoring, truncation and dependency rates.

Figures 4.1 to 4.5 illustrate that the estimated density improves with larger sample sizes and lower dependency rates. It appears to be relatively unaffected by high truncation or censoring values and remains robust in the presence of outliers, introduced by multiplying 4 % of each sample by a multiplier factor (MF).

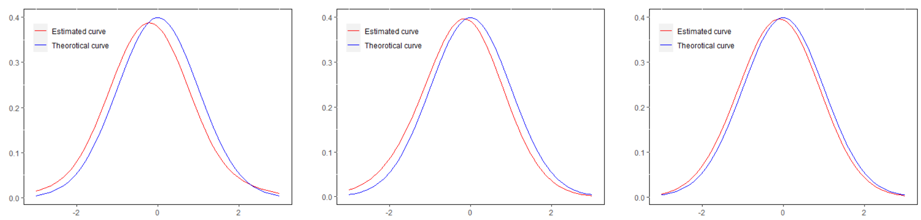

Figure 4.1: ρ=0.1, CP=20%, TP=20%, n=50,100, and 300 respectively.

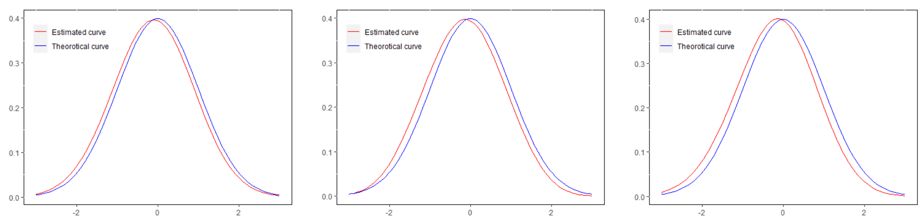

Figure 4.2: n=300, CP=20%, TP=20%, ρ=0.1, 0.5 and, 0.8 respectively.

Next, we simulate confidence intervals for $m(x)$ at a level of 95 %, as outlined in Corollary 4.1:

$$\widehat{m}_n(x) \pm 1.96 \times \frac{\widehat{\sigma}_n(x)}{\sqrt{nh_n^d}},$$

using the same data and arguments employed to calculate the estimator $\widehat{m}_n(x)$. Figure 4.6 clearly demonstrates the improvement in the interval's width for larger sample sizes.

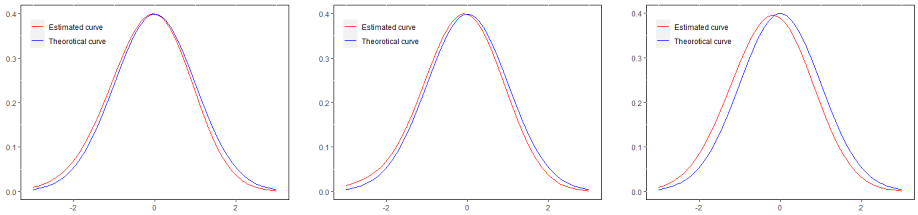

Figure 4.3: ρ=0.1, n=300, TP=20%, CP=10%, 30%, and 60% respectively.

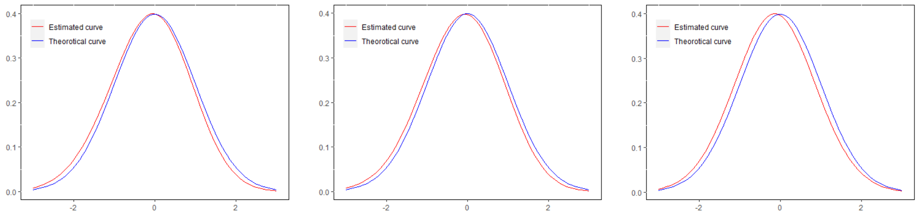

Figure 4.4: ρ=0.1, n=300, CP=20%, TP=10%, 30%, and 60% respectively.

Finally, we evaluate the effectiveness of the asymptotic normality by calculating the coverage probabilities of the 95 % confidence intervals. These are determined by calculating the percentage of true values that fall within the estimated confidence interval. Table 4.1 presents the results for 100 replications of observed sequences of size n. The coverage probability values suggest reasonable coverage for the confidence intervals, and they improve as the sample size n increases and the censoring, truncation, and dependency rates decrease. Additionally, the coverage appears to be relatively robust to the presence of outliers.

4.4 Real data application

This section evaluates the performance of the RER estimator for analyzing dependent data exhibiting left truncation and right censoring, using a real dataset reported by Hyde (1977). This dataset comprises the lifespans of 97 male residents at the Channing House retirement community in Palo Alto, California, from January 1964 to July 1975. It includes information on age at entry, age at death/departure, and survival status at the study's conclusion.

The dataset exhibits both left truncation and right censoring. Left truncation arises because residents needed to be at least 55 years old to enter the community. Right censoring occurs due to the study's end date or loss to follow-up. Our primary interest lies in exploring the relationship between the age of entry and the age of death.

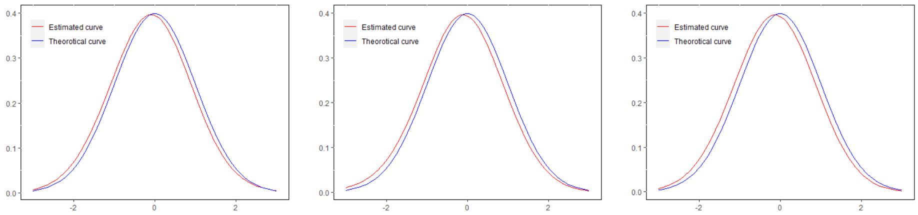

Figure 4.5: ρ=0.1, n=300, CP=20%, TP=20%, MF=50, 100, and 150 respectively.

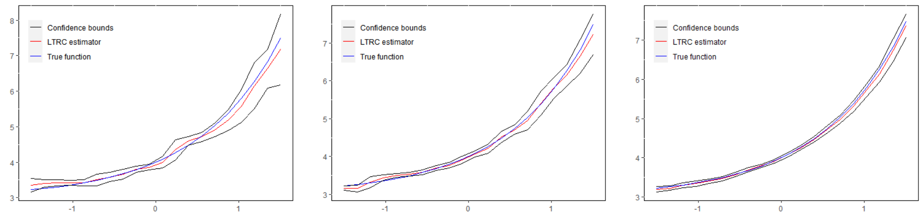

Figure 4.6: ρ=0.1, CP=20%, TP= 20%, n=50, 100, and 300 respectively.

Initial analysis revealed a strong positive correlation (0.79) between these two variables, indicating their dependence. Further analysis using partial correlation suggests an autoregressive process of order 1 (AR(1)), implying α-mixing dependence in the data.

For prediction purposes, the dataset was split, allocating 80% for estimator calculation and 20% for assessing prediction accuracy. We employed the Gaussian kernel and utilized the cross-validation method to determine the optimal bandwidth. Since approximately 53 % of the data was censored, these observations were excluded from the prediction process as they lack meaningful survival times.

Figure 4.7 presents the RER prediction of age at death based on age at entry for Channing House residents, along with pointwise confidence intervals. As evident from the figure, the predicted values are close to the true values within the upper and lower bands. This suggests the effectiveness of the RER approach and the precision of our estimator.

4.5 Auxiliary results and proofs

The objective of this part is to demonstrate the asymptotic normality result of the estimator $\widehat{m}_n(\cdot)$ stated in Theorem 4.1:

$$\sqrt{nh_n^d}(\widehat{m}_n(x) - m(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_n^2(x)) \text{ as } n \to \infty.$$

Table 4.1: Coverage probabilities for 100 replications.

		CP=20%				TP=20%			
		TP=10%		TP=40%		CP=10%		CP=40%	
n	MF	ρ=0.1	ρ=0.8	ρ=0.1	ρ=0.8	ρ=0.1	ρ=0.8	ρ=0.1	ρ=0.8
	1	0.949	0.933	0.940	0.927	0.955	0.924	0.932	0.903
100	50	0.945	0.927	0.935	0.920	0.952	0.922	0.929	0.900
	100	0.943	0.922	0.931	0.916	0.950	0.919	0.927	0.897
	150	0.938	0.919	0.929	0.901	0.947	0.915	0.918	0.896
	1	0.960	0.942	0.955	0.932	0.963	0.936	0.941	0.917
300	50	0.955	0.939	0.949	0.930	0.957	0.933	0.932	0.913
	100	0.955	0.933	0.947	0.926	0.956	0.933	0.930	0.909
	150	0.943	0.930	0.938	0.921	0.949	0.926	0.925	0.906

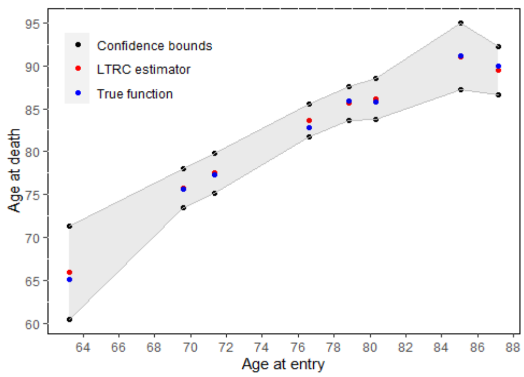

Figure 4.7: RER prediction of age at death based on age at entry for Channing House residents, along with pointwise confidence intervals.

To this end, we use the following decomposition

$$\begin{aligned}\sqrt{nh_n^d}\left(\widehat{\psi}_{n,\ell}(\mathbf{x})-\psi_\ell(\mathbf{x})\right) &= \sqrt{nh_n^d}\left\{\left(\widehat{\psi}_{n,\ell}(\mathbf{x})-\widetilde{\psi}_{n,\ell}(\mathbf{x})\right)+\left(\widetilde{\psi}_{n,\ell}(\mathbf{x})-\mathbf{E}[\widetilde{\psi}_{n,\ell}(\mathbf{x})]\right)\right.\\ &\quad\left.+\left(\mathbf{E}[\widetilde{\psi}_{n,\ell}(\mathbf{x})]-\psi_\ell(\mathbf{x})\right)\right\}\\ &=: \sqrt{nh_n^d}\left\{\Lambda_{1,\ell}(\mathbf{x})+\Lambda_{2,\ell}(\mathbf{x})+\Lambda_{3,\ell}(\mathbf{x})\right\},\ \text{for } \ell=1,2. \qquad (4.1)\end{aligned}$$

We start by proving in the upcoming lemma that the terms $\sqrt{nh_n^d}\Lambda_{1,\ell}(\mathbf{x})$ and $\sqrt{nh_n^d}\Lambda_{3,\ell}(\mathbf{x})$ are negligible. After that, we show that the leading term $\sqrt{nh_n^d}\Lambda_{2,\ell}(\mathbf{x})$ is asymptotically normal using the Bernstein's large blocks and small blocks procedure.

Lemma 4.1. *Under Assumptions **K**, **H**, **D**(a) and **D**(b), we have for $\ell = 1, 2$*

$$\begin{array}{ll} (i) & \sqrt{nh_n^d}|\Lambda_{1,\ell}(\boldsymbol{x})| = o_{a.s.}(1), \\ (ii) & \sqrt{nh_n^d}|\Lambda_{3,\ell}(\boldsymbol{x})| = o_{a.s.}(1) \text{ as } n \to \infty. \end{array}$$

Proof. After establishing Lemma 3.1 and Lemma 3.3 in Chapter 3 and using Assumptions **H**(b) and **H**(c), we find that

$$\begin{array}{ll} (i) & \sqrt{nh_n^d}|\Lambda_{1,\ell}(\mathbf{x})| = O_{a.s.}\left(\sqrt{h_n^d \log\log n}\right) = o_{a.s.}(1), \\ (ii) & \sqrt{nh_n^d}|\Lambda_{3,\ell}(\mathbf{x})| = O_{a.s.}\left(\sqrt{nh_n^{d+4}}\right) = o_{a.s.}(1), \end{array}$$

which conclude the proof. □

Next, before proceeding with the asymptotic normality of the dominant term, we begin by calculating the necessary variance and covariance in Lemma 4.2 and Lemma 4.3 hereafter.

Lemma 4.2. *Under Assumptions **K**(a), **K**(b), **D** and **M**(b), we have for $\lambda = 2\ell$ with $\ell = 1, 2$*

$$\boldsymbol{Var}\left(\sqrt{nh_n^d}\Lambda_{2,\ell}(\boldsymbol{x})\right) \to \mu_\lambda(\boldsymbol{x})\kappa.$$

Proof. Define $U_{i,\ell}(\mathbf{x})$ for $1 \le i \le n$ as

$$U_{i,\ell}(\mathbf{x}) = \frac{\eta Z_i^{-\ell}\delta_i}{L(Z_i)\bar{G}(Z_i)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_i}{h_n}\right) - \mathbf{E}\left[\frac{\eta Z_1^{-\ell}\delta_1}{L(Z_1)\bar{G}(Z_1)} K_d\left(\frac{\mathbf{x} - \mathbf{X}_1}{h_n}\right)\right], \text{ for } \ell = 1, 2.$$

It follows directly that

$$\Lambda_{2,\ell}(\mathbf{x}) = \frac{1}{nh_n^d}\sum_{i=1}^n U_{i,\ell}(\mathbf{x}).$$

We can then analyze the variance of the sum

$$\begin{aligned} \mathbf{Var}\left(\sqrt{nh_n^d}\Lambda_{2,\ell}(\mathbf{x})\right) &= \frac{1}{nh_n^d}\mathbf{Var}\left(\sum_{i=1}^n U_{i,\ell}(\mathbf{x})\right) \\ &= \frac{1}{h_n^d}\mathbf{Var}\left(U_{1,\ell}(\mathbf{x})\right) + \frac{1}{nh_n^d}\sum_{i=1}^n\sum_{\substack{j=1 \\ i\neq j}}^n \mathbf{Cov}\left(U_{i,\ell}(\mathbf{x}), U_{j,\ell}(\mathbf{x})\right) \\ &= \frac{1}{h_n^d}\mathbf{E}\left[U_{1,\ell}^2(\mathbf{x})\right] + \frac{1}{nh_n^d}\sum_{i=1}^n\sum_{\substack{j=1 \\ i\neq j}}^n \mathbf{E}\left[U_{i,\ell}(\mathbf{x})U_{j,\ell}(\mathbf{x})\right] \\ &=: \mathcal{I}_1 + \mathcal{I}_2. \end{aligned}$$

For the first term $\mathcal{I}_1$, we have

$$\begin{aligned}
\mathcal{I}_1 &= \frac{1}{h_n^d}\mathbf{E}[U_{1,\ell}^2(\mathbf{x})] \\
&= \frac{1}{h_n^d}\mathbf{E}\left[\left(\frac{\eta Z_1^{-\ell}\delta_1}{L(Z_1)\bar{G}(Z_1)}K_d\left(\frac{\mathbf{x}-X_1}{h_n}\right) - \mathbf{E}\left[\frac{\eta Z_1^{-\ell}\delta_1}{L(Z_1)\bar{G}(Z_1)}K_d\left(\frac{\mathbf{x}-X_1}{h_n}\right)\right]\right)^2\right] \\
&= \frac{1}{h_n^d}\left\{\mathbf{E}\left[\frac{\eta^2 Z_1^{-2\ell}\delta_1}{L^2(Z_1)\bar{G}^2(Z_1)}K_d^2\left(\frac{\mathbf{x}-X_1}{h_n}\right)\right] - \mathbf{E}^2\left[\frac{\eta Z_1^{-\ell}\delta_1}{L(Z_1)\bar{G}(Z_1)}K_d\left(\frac{\mathbf{x}-X_1}{h_n}\right)\right]\right\} \\
&=: \frac{1}{h_n^d}\left\{\mathcal{I}_{1,1} - \mathcal{I}_{1,2}\right\}.
\end{aligned}$$

Utilizing equation (3.9) and Assumptions $\mathbf{K}$(b) and $\mathbf{D}$(c), we obtain for $\lambda = 2\ell$ with $\ell = 1, 2$

$$\begin{aligned}
\mathcal{I}_{1,1} &= \mathbf{E}\left[\frac{\eta^2 Z_1^{-2\ell}\delta_1}{L^2(Z_1)\bar{G}^2(Z_1)}K_d^2\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right] \\
&= \int_{\mathbb{R}^d}\int_{t\geq a_H}\frac{\eta^2 t^{-2\ell}}{L^2(t)\bar{G}^2(t)}K_d^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)d\mathbf{H}_1(\mathbf{u},t) \\
&= \int_{\mathbb{R}^d}\int_{t\geq a_H}\frac{\eta t^{-2\ell}}{L(t)\bar{G}(t)}K_d^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)f_{\mathbf{X},Y}(\mathbf{u},t)\mathbf{du}dt \\
&\leq \int_{\mathbb{R}^d}K_d^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)\mu_\lambda(\mathbf{u})\mathbf{du} \\
&\leq h_n^d\int_{\mathbb{R}^d}K_d^2(\mathbf{s})\mu_\lambda(\mathbf{x}-\mathbf{s}h_n)\mathbf{ds} \\
&\leq h_n^d\mu_\lambda(\mathbf{x})\int_{\mathbb{R}^d}K_d^2(\mathbf{s})\mathbf{ds} \\
&\leq h_n^d\mu_\lambda(\mathbf{x})\kappa, \qquad (4.2)
\end{aligned}$$

Simularly, based on equation (3.9) and assumptions $\mathbf{K}$(a) and $\mathbf{D}$(a), we get

$$\begin{aligned}
\sqrt{\mathcal{I}_{1,2}} &= \mathbf{E}\left[\frac{\eta Z_1^{-\ell}\delta_1}{L(Z_1)\bar{G}(Z_1)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right] \\
&= \int_{\mathbb{R}^d}\int_{t\geq a_H}\frac{\eta t^{-\ell}}{L(t)\bar{G}(t)}K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)d\mathbf{H}_1(\mathbf{u},t) \\
&\leq \int_{\mathbb{R}^d}\int_{t\geq a_H}t^{-\ell}K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)f_{\mathbf{X},Y}(\mathbf{u},t)\mathbf{du}dt \\
&\leq \int_{\mathbb{R}^d}K_d\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)\psi_\ell(\mathbf{u})\mathbf{du} \\
&\leq h_n^d\int_{\mathbb{R}^d}K_d(\mathbf{s})\psi_\ell(\mathbf{x}-\mathbf{s}h_n)\mathbf{ds} \\
&= O(h_n^d). \qquad (4.3)
\end{aligned}$$

Then, by combining (4.2) and (4.3), we obtain

$$\mathcal{I}_1 = \frac{1}{h_n^d}\left\{h_n^d\mu_\lambda(\mathbf{x})\kappa - Ch_n^{2d}\right\} \xrightarrow[n\to+\infty]{} \mu_\lambda(\mathbf{x})\kappa. \qquad (4.4)$$

For the second term $\mathcal{I}_2$, we consider a sequence of integer β_n that approaches infinity as $n \to \infty$, such that

$$\mathcal{I}_2 = \frac{1}{nh_n^d}\left\{\sum_{1\leq|i-j|\leq\beta_n} \mathbf{E}\left[U_{i,\ell}(\mathbf{x})U_{j,\ell}(\mathbf{x})\right] + \sum_{\beta_n<|i-j|<n} \mathbf{E}\left[U_{i,\ell}(\mathbf{x})U_{j,\ell}(\mathbf{x})\right]\right\}.$$

On the one hand, under Assumption **D**(b) and **D**(d), we get

$$\begin{aligned}
&|\mathbf{E}(U_{i,\ell}(\mathbf{x})U_{j,\ell}(\mathbf{x}))| \\
&= \left|\mathbf{E}\left[\frac{\eta Z_i^{-\ell}\delta_i}{L(Z_i)\bar{G}(Z_i)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\frac{\eta Z_j^{-\ell}\delta_j}{L(Z_j)\bar{G}(Z_j)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\right]\right. \\
&\quad \left. - \mathbf{E}\left[\frac{\eta Z_i^{-\ell}\delta_i}{L(Z_i)\bar{G}(Z_i)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)\right]\mathbf{E}\left[\frac{\eta Z_j^{-\ell}\delta_j}{L(Z_j)\bar{G}(Z_j)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_j}{h_n}\right)\right]\right| \\
&\leq Ch_n^{2d}\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K_d(\mathbf{s})K_d(\mathbf{t})|\Upsilon_{ij}(\mathbf{x}-\mathbf{s}h_n, \mathbf{x}-\mathbf{t}h_n) \\
&\quad -\Upsilon_i(\mathbf{x}-\mathbf{s}h_n)\Upsilon_j(\mathbf{x}-\mathbf{t}h_n)|\mathbf{dsdt} \\
&= O(h_n^{2d}).
\end{aligned} \tag{4.5}$$

From (4.5), we have

$$\sum_{1\leq|i-j|\leq\beta_n} \mathbf{E}\left[U_{i,\ell}(\mathbf{x})U_{j,\ell}(\mathbf{x})\right] \leq Ch_n^{2d} n\beta_n,$$

and by choosing $\beta_n = [(1/h_n^d)^{1/b}]$ $b > 1$, we obtain

$$\frac{1}{nh_n^d}\sum_{1\leq|i-j|\leq\beta_n} \mathbf{E}\left[U_{i,\ell}(\mathbf{x})U_{j,\ell}(\mathbf{x})\right] \leq Ch_n^{d(b-1)/b} \to 0. \tag{4.6}$$

On the other hand, we use a Davydov covariance inequality given in Proposition 1.3. This leads to

$$\begin{aligned}
\frac{1}{nh_n^d}\sum_{\beta_n<|i-j|<n} |\mathbf{E}\left[U_{i,\ell}(\mathbf{x})U_{j,\ell}(\mathbf{x})\right]| &\leq \frac{1}{nh_n^d}\sum_{\beta_n<|i-j|<n} C\alpha^{\frac{1}{a_1}}(|i-j|) \\
&\quad \times \left(\mathbf{E}\left|K_d\left(\frac{\mathbf{x}-X_i}{h_n}\right)\right|^{a_2}\right)^{\frac{1}{a_2}}\left(\mathbf{E}\left|K_d\left(\frac{\mathbf{x}-X_i}{h_n}\right)\right|^{a_3}\right)^{\frac{1}{a_3}},
\end{aligned}$$

where a_1, a_2 and a_3 take values in $\mathbb{R}^+$. We have

$$\mathbf{E}\left|K_d\left(\frac{\mathbf{x}-X_i}{h_n}\right)\right|^{a_2} \leq Ch_n^d.$$

Then, from Assumption **M**(b), we get

$$\begin{aligned}
\frac{1}{nh_n^d}\sum_{\beta_n<|i-j|<n} |\mathbf{E}\left[U_{i,\ell}(\mathbf{x})U_{j,\ell}(\mathbf{x})\right]| &\leq \frac{C}{nh_n^d}\sum_{\beta_n<|i-j|<n} \alpha^{\frac{1}{a_1}}(|i-j|)(h_n^d)^{\frac{1}{a_2}}(h_n^d)^{\frac{1}{a_3}} \\
&\leq \frac{C}{nh_n^d}\sum_{\beta_n<|i-j|<n} \alpha^{\frac{1}{a_1}}(|i-j|)h_n^{\frac{d}{a_1}(a_1-1)} \\
&\leq \frac{C}{nh_n^{\frac{d}{a_1}}}\sum_{s>\beta_n} \alpha^{\frac{1}{a_1}}(s) \to 0.
\end{aligned} \tag{4.7}$$

Finally, by merging (4.4), (4.6) and (4.7), we arrive at the outcome of Lemma 4.2. □

Lemma 4.3. *Under Assumptions **K**(a), **K**(b), **D** and **M**(b), we have*

$$\boldsymbol{Cov}\left(\sqrt{nh_n^d}\Lambda_{2,1}(\boldsymbol{x}), \sqrt{nh_n^d}\Lambda_{2,2}(\boldsymbol{x})\right) \to \mu_3(\boldsymbol{x})\kappa.$$

Proof. We begin by decomposing the covariance term

$$\begin{aligned}
\mathbf{Cov}\left(\sqrt{nh_n^d}\Lambda_{2,1}(\mathbf{x}), \sqrt{nh_n^d}\Lambda_{2,2}(\mathbf{x})\right) &= \frac{1}{nh_n^d}\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{Cov}\left(U_{i,1}(\mathbf{x}), U_{j,2}(\mathbf{x})\right) \\
&= \frac{1}{nh_n^d}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\mathbf{E}\left(U_{i,1}(\mathbf{x})U_{j,2}(\mathbf{x})\right)\right. \\
&\quad \left. - \mathbf{E}\left(U_{i,1}(\mathbf{x})\right)\mathbf{E}\left(U_{j,2}(\mathbf{x})\right)\right] \\
&= \frac{1}{h_n^d}\mathbf{E}\left(U_{1,1}(\mathbf{x})U_{1,2}(\mathbf{x})\right) \\
&\quad + \frac{1}{nh_n^d}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ i\neq j}}^{n}\mathbf{E}\left(U_{i,1}(\mathbf{x})U_{j,2}(\mathbf{x})\right) \\
&=: \mathcal{J}_1 + \mathcal{J}_2.
\end{aligned}$$

On the one side, we have

$$\begin{aligned}
\mathcal{J}_1 &= \frac{1}{h_n^d}\mathbf{E}\left(U_{1,1}(\mathbf{x})U_{1,2}(\mathbf{x})\right) \\
&= \frac{1}{h_n^d}\left\{\mathbf{E}\left(\frac{\alpha^2 Z_1^{-3}\delta_1}{L^2(Z_1)\bar{G}^2(Z_1)}K_d^2\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right)\right. \\
&\quad \left. -\mathbf{E}\left(\frac{\alpha Z_1^{-1}\delta_1}{L(Z_1)\bar{G}(Z_1)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right)\mathbf{E}\left(\frac{\alpha Z_1^{-2}\delta_1}{L(Z_1)\bar{G}(Z_1)}K_d\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right)\right\} \\
&=: \frac{1}{h_n^d}\left\{\mathcal{J}_{1,1} - \mathcal{J}_{1,2}\mathcal{J}_{1,3}\right\}.
\end{aligned}$$

Utilizing equation (3.9) and Assumptions **K**(b) and **D**(c), we obtain

$$\begin{aligned}
\mathcal{J}_{1,1} &= \mathbf{E}\left[\frac{\eta^2 Z_1^{-3}\delta_1}{L^2(Z_1)\bar{G}^2(Z_1)}K_d^2\left(\frac{\mathbf{x}-\mathbf{X}_1}{h_n}\right)\right] \\
&= \int_{\mathbb{R}^d}\int_{t\geq a_H}\frac{\eta^2 t^{-3}}{L^2(Z_1)\bar{G}^2(Z_1)}K_d^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)d\mathbf{H}_1(\mathbf{u},t) \\
&= \int_{\mathbb{R}^d}\int_{t\geq a_H}\frac{\eta t^{-3}}{L(Z_1)\bar{G}(Z_1)}K_d^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)f_{\mathbf{X},Y}(\mathbf{u},t)\mathbf{du}dt \\
&\leq \int_{\mathbb{R}^d}K_d^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_n}\right)\mu_3(\mathbf{u})\mathbf{du} \\
&\leq \int_{\mathbb{R}^d}h_n^d K_d^2(\mathbf{s})\mu_3(\mathbf{x}-\mathbf{s}h_n)\mathbf{ds} \\
&\leq h_n^d\mu_3(\mathbf{x})\int_{\mathbb{R}^d}K_d^2(\mathbf{s})\mathbf{ds} \\
&\leq h_n^d\mu_3(\mathbf{x})\kappa.
\end{aligned}$$

For $\mathcal{J}_{1,2}$ and $\mathcal{J}_{1,3}$, we follow the same approach as in the proof of the term $\sqrt{\mathcal{I}_{1,2}}$ in Lemma 4.2 and at the end, we get

$$\mathcal{J}_1 \leq \frac{1}{h_n^d}\left\{h_n^d \mu_3(\mathbf{x})\kappa - Ch_n^{4d}\right\} \xrightarrow[n\to+\infty]{} \mu_3(\mathbf{x})\kappa. \tag{4.8}$$

On the other side, to show that $\mathcal{J}_2 = o(1)$, we work as for the proof of the term $\mathcal{I}_2$ and then we have the result of Lemma 4.3. □

Now, let us focus on proving the asymtotic normality of the leading term. For that, we establish the following lemma:

Lemma 4.4. *Under Assumptions **K**, **D**, **M** and **B**, we have*

$$\sqrt{nh_n^d}\left(\Lambda_{2,1}(\boldsymbol{x}), \Lambda_{2,2}(\boldsymbol{x})\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_{\boldsymbol{x}}\kappa).$$

where $\Sigma_{\boldsymbol{x}} = \begin{pmatrix} \mu_2(\boldsymbol{x}) & \mu_3(\boldsymbol{x}) \\ \mu_3(\boldsymbol{x}) & \mu_4(\boldsymbol{x}) \end{pmatrix}$.

Proof. In order to demonstrate the lemma, we need only to show that for any given $a^T = (a_1, a_2)$

$$\sqrt{nh_n^d}\left(a_1\Lambda_{2,1}(\mathbf{x}) + a_2\Lambda_{2,2}(\mathbf{x})\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, a^T\Sigma_{\mathbf{x}}a\kappa),$$

To do so, we first set

$$\Lambda_{2,n}(\mathbf{x}) := \sqrt{nh_n^d}(a_1\Lambda_{2,1}(\mathbf{x}) + a_2\Lambda_{2,2}(\mathbf{x}))$$

and

$$V_i(\mathbf{x}) := a_1 U_{i,1}(\mathbf{x}) + a_2 U_{i,2}(\mathbf{x}),$$

then

$$\Lambda_{2,n}(\mathbf{x}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}}.$$

This implies that our goal is to demonstrate that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}} \xrightarrow{\mathcal{D}} \mathcal{N}(0, a^T\Sigma_{\mathbf{x}}a\kappa). \tag{4.9}$$

For this purpose, as previously mentioned, we employ Bernstein's large blocks and small blocks procedure. According to the notations in Assumption (**B**), we partition the set $\{1, 2, \ldots, n\}$ into k large p-blocks separated by k small q-blocks, along with a set containing the remaining points which may be empty. This breakdown is illustrated as follows

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}} &= \frac{1}{\sqrt{n}}\left\{\sum_{j=1}^{k}\widetilde{V}_{1,j}(\mathbf{x}) + \sum_{j=1}^{k}\widetilde{V}_{2,j}(\mathbf{x}) + \widetilde{V}_{3,j}(\mathbf{x})\right\} \\ &=: \frac{1}{\sqrt{n}}\left\{\widetilde{V}_1(\mathbf{x}) + \widetilde{V}_2(\mathbf{x}) + \widetilde{V}_3(\mathbf{x})\right\},\end{aligned}$$

where

$$\widetilde{V}_{1,j}(\mathbf{x}) = \sum_{i=(j-1)(p+q)+1}^{(j-1)(p+q)+p} \frac{V_i(\mathbf{x})}{\sqrt{h_n^d}}, \widetilde{V}_{2,j}(\mathbf{x}) = \sum_{i=(j-1)(p+q)+p+1}^{j(p+q)} \frac{V_i(\mathbf{x})}{\sqrt{h_n^d}},$$

$$\widetilde{V}_{3,j}(\mathbf{x}) = \sum_{i=k(p+q)+1}^{n} \frac{V_i(\mathbf{x})}{\sqrt{h_n^d}}.$$

So, to establish the convergence indicated in (4.9), it is enough to verify that

$$\frac{1}{\sqrt{n}}\widetilde{V}_2(\mathbf{x}) \xrightarrow{\mathbf{P}} 0, \frac{1}{\sqrt{n}}\widetilde{V}_3(\mathbf{x}) \xrightarrow{\mathbf{P}} 0 \tag{4.10}$$

and

$$\frac{1}{\sqrt{n}}\widetilde{V}_1(\mathbf{x}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, a^T \Sigma_{\mathbf{x}} a \kappa). \tag{4.11}$$

We begin with (4.10). To prove that the two terms are **P**-negligible, we apply Tchebychev's inquality. We have:

$$\begin{aligned}
\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_2(\mathbf{x})\right) &= \frac{1}{n}\mathbf{Var}\left(\sum_{j=1}^{k}\widetilde{V}_{2,j}(\mathbf{x})\right) \\
&= \frac{1}{n}\sum_{j=1}^{k}\mathbf{Var}\left(\widetilde{V}_{2,j}(\mathbf{x})\right) + \frac{1}{n}\sum_{\substack{j=1 \\ j\neq m}}^{k}\sum_{m=1}^{k}\mathbf{Cov}\left(\widetilde{V}_{2,j}(\mathbf{x}), \widetilde{V}_{2,m}(\mathbf{x})\right) \\
&=: \mathcal{S}_1 + \mathcal{S}_2.
\end{aligned}$$

For $\mathcal{S}_1$, we have

$$\begin{aligned}
\mathcal{S}_1 &= \frac{1}{n}\sum_{j=1}^{k}\mathbf{Var}\left(\widetilde{V}_{2,j}(\mathbf{x})\right) \\
&= \frac{k}{n}\mathbf{Var}\left(\widetilde{V}_{2,1}(\mathbf{x})\right) \\
&= \frac{k}{n}\sum_{i=1}^{q}\mathbf{Var}\left(\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}}\right) + \frac{k}{n}\sum_{\substack{i=1 \\ i\neq s}}^{q}\sum_{s=1}^{q}\mathbf{Cov}\left(\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}}, \frac{V_s(\mathbf{x})}{\sqrt{h_n^d}}\right) \\
&=: \mathcal{S}_{1,1} + \mathcal{S}_{1,2}.
\end{aligned}$$

Using (4.4) and (4.8), we obtain

$$
\begin{aligned}
\mathcal{S}_{1,1} &= \frac{k}{n}\sum_{j=1}^{q}\mathbf{Var}\left(\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}}\right)\\
&= \frac{kq}{n}\mathbf{Var}\left(\frac{a_1U_{1,1}(\mathbf{x})+a_2U_{1,2}(\mathbf{x})}{\sqrt{h_n^d}}\right)\\
&= \frac{kq}{n}\left\{a_1^2\frac{1}{h_n^d}\mathbf{Var}\left(U_{1,1}(\mathbf{x})\right)+a_2^2\frac{1}{h_n^d}\mathbf{Var}\left(U_{1,2}(\mathbf{x})\right)\right.\\
&\quad\left.+a_1a_2\frac{1}{h_n^d}\mathbf{Cov}\left(U_{1,1}(\mathbf{x}),U_{1,2}(\mathbf{x})\right)\right\}\\
&= \frac{kq}{n}\left\{a_1^2\frac{1}{h_n^d}\mathbf{E}\left(U_{1,1}^2(\mathbf{x})\right)\right.\\
&\quad\left.+a_2^2\frac{1}{h_n^d}\mathbf{E}\left(U_{1,2}^2(\mathbf{x})\right)+a_1a_2\frac{1}{h_n^d}\mathbf{E}\left(U_{1,1}(\mathbf{x})U_{1,2}(\mathbf{x})\right)\right\}\\
&\leq C\left(\frac{k(p+q)}{n}-\frac{kp}{n}\right)h_n^d, \qquad (4.12)
\end{aligned}
$$

and by stationarity and (4.5), we get

$$
\begin{aligned}
\mathcal{S}_{1,2} &= \frac{k}{n}\sum_{\substack{i=1\\ i\neq s}}^{q}\sum_{s=1}^{q}\mathbf{Cov}\left(\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}},\frac{V_s(\mathbf{x})}{\sqrt{h_n^d}}\right)\\
&= \frac{2k}{n}\sum_{m=1}^{q-1}(q-m)\mathbf{Cov}\left(\frac{V_1(\mathbf{x})}{\sqrt{h_n^d}},\frac{V_{m+1}(\mathbf{x})}{\sqrt{h_n^d}}\right)\\
&\leq \frac{2kq}{n}\sum_{m=1}^{q-1}\mathbf{Cov}\left(\frac{a_1U_{1,1}(\mathbf{x})+a_2U_{1,2}(\mathbf{x})}{\sqrt{h_n^d}},\frac{a_1U_{m+1,1}(\mathbf{x})+a_2U_{m+1,2}(\mathbf{x})}{\sqrt{h_n^d}}\right)\\
&\leq \frac{2kq}{n}\sum_{m=1}^{q-1}\frac{1}{h_n^d}\left\{a_1^2\mathbf{E}\left(U_{1,1}(\mathbf{x})U_{m+1,1}(\mathbf{x})\right)+a_1a_2\mathbf{E}\left(U_{1,1}(\mathbf{x})U_{m+1,2}(\mathbf{x})\right)\right.\\
&\quad\left.+a_2a_1\mathbf{E}\left(U_{1,2}(\mathbf{x})U_{m+1,1}(\mathbf{x})\right)+a_2^2\mathbf{E}\left(U_{1,2}(\mathbf{x})U_{m+1,2}(\mathbf{x})\right)\right\}\\
&\leq C\left(\frac{k(p+q)}{n}-\frac{kp}{n}\right)qh_n^d. \qquad (4.13)
\end{aligned}
$$

From (4.12), (4.13) and by using Assumptions **B**(a), **B**(b) and **B**(c), the term $\mathcal{S}_1\to 0$.

For $\mathcal{S}_2$, using stationary, Assumptions $\mathbf{M}$(b), $\mathbf{B}$(a) and $\mathbf{B}$(b), we get

$$\begin{aligned}
\mathcal{S}_2 &= \frac{1}{n}\sum_{\substack{j=1 \\ j\neq m}}^{k}\sum_{m=1}^{k}\mathbf{Cov}\left(\widetilde{V}_{2,j}(\mathbf{x}),\widetilde{V}_{2,m}(\mathbf{x})\right) \\
&= \frac{2}{n}\sum_{s=1}^{k-1}(k-s)\mathbf{Cov}\left(\widetilde{V}_{2,1}(\mathbf{x}),\widetilde{V}_{2,s+1}(\mathbf{x})\right) \\
&\le \frac{2k}{nh_n^d}\sum_{s=1}^{k-1}\mathbf{Cov}\left(\sum_{i=1}^{q}V_i(\mathbf{x}),\sum_{r=s(p+q)+p+1}^{(s+1)(p+q)}V_r(\mathbf{x})\right) \\
&\le \frac{2k}{nh_n^d}\sum_{s=1}^{k-1}\left[\sum_{m=1}^{q}(q-m+1)\mathbf{Cov}\left(V_1(\mathbf{x}),V_{s(p+q)+m}(\mathbf{x})\right)\right. \\
&\quad \left.+\sum_{m=1}^{q-1}(q-m)\mathbf{Cov}\left(V_{m+1}(\mathbf{x}),V_{s(p+q)+1}(\mathbf{x})\right)\right] \\
&\le \frac{2k}{nh_n^d}\sum_{s=1}^{k-1}\left[q\sum_{m=s(p+q)+1}^{s(p+q)+q}\mathbf{Cov}\left(V_1(\mathbf{x}),V_m(\mathbf{x})\right)\right. \\
&\quad \left.+q\sum_{m=s(p+q)-(q-2)}^{s(p+q)}\mathbf{Cov}\left(V_1(\mathbf{x}),V_m(\mathbf{x})\right)\right] \\
&\le \frac{2k}{nh_n^d}\sum_{s=1}^{k-1}\left[q\sum_{m=s(p+q)-(q-2)}^{s(p+q)+q}\mathbf{Cov}\left(V_1(\mathbf{x}),V_m(\mathbf{x})\right)\right] \\
&\le \frac{2kq}{nh_n^d}\sum_{s=1}^{k-1}\left[\sum_{m=1}^{2q-1}\mathbf{Cov}\left(V_1(\mathbf{x}),V_{s(p+q)-q+m+1}(\mathbf{x})\right)\right] \\
&\le \frac{2kq}{nh_n^d}\sum_{s=1}^{k-1}\sum_{m=1}^{2q-1}\alpha^{\frac{1}{a_1}}(s(p+q)-q+m)h_n^{\frac{d}{a_1}(a_1-1)} \\
&\le \frac{2kq}{n}\frac{1}{h_n^{d/a_1}}\sum_{j=p+1}^{\infty}\alpha^{\frac{1}{a_1}}(j) \\
&\le \left(\frac{k(p+q)}{n}-\frac{kp}{n}\right)\frac{1}{h_n^{d/a_1}}\sum_{j=p+1}^{\infty}\alpha^{\frac{1}{a_1}}(j)\to 0.
\end{aligned}$$

Thus

$$\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_2(\mathbf{x})\right)\to 0.$$

Next, we move to term $\frac{1}{\sqrt{n}}\widetilde{V}_3(\mathbf{x})$ and we have

$$\begin{aligned}\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_3(\mathbf{x})\right) &= \frac{1}{n}\mathbf{Var}\left(\sum_{i=k(p_n+q_n)+1}^{n}\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}}\right)\\ &= \frac{1}{n}\sum_{i=k(p_n+q_n)+1}^{n}\mathbf{Var}\left(\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}}\right)\\ &\quad+\frac{1}{n}\sum_{\substack{i=k(p+q)+1\\ i\neq s}}^{n}\sum_{s=k(p+q)+1}^{n}\mathbf{Cov}\left(\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}},\frac{V_s(\mathbf{x})}{\sqrt{h_n^d}}\right)\\ &= \mathcal{R}_1+\mathcal{R}_2.\end{aligned}$$

In the same way as for $\mathcal{S}_{1,1}$, we obtain

$$\begin{aligned}\mathcal{R}_1 &= \frac{1}{n}\sum_{i=k(p_n+q_n)+1}^{n}\mathbf{Var}\left(\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}}\right)\\ &= \frac{n-k(p+q)}{n}\mathbf{Var}\left(\frac{a_1U_{1,1}(\mathbf{x})+a_2U_{1,2}(\mathbf{x})}{\sqrt{h_n^d}}\right)\\ &\leq C\left(1-\frac{k(p+q)}{n}\right)h_n^d,\end{aligned}$$

and in a manner similar to $\mathcal{S}_{1,2}$, we reach

$$\begin{aligned}\mathcal{R}_2 &= \frac{1}{n}\sum_{\substack{i=k(p+q)+1\\ i\neq s}}^{n}\sum_{s=k(p+q)+1}^{n}\mathbf{Cov}\left(\frac{V_i(\mathbf{x})}{\sqrt{h_n^d}},\frac{V_s(\mathbf{x})}{\sqrt{h_n^d}}\right)\\ &= \frac{2}{n}\sum_{m=1}^{p-1}(p-m)\mathbf{Cov}\left(\frac{V_1(\mathbf{x})}{\sqrt{h_n^d}},\frac{V_{m+1}(\mathbf{x})}{\sqrt{h_n^d}}\right)\\ &\leq C\frac{p}{n}ph_n^d.\end{aligned}$$

So, under Assumptions **B**(a) and **B**(c), we get

$$\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_3(\mathbf{x})\right)\to 0.$$

As for the verification of (4.11), it will be accomplished in three points:

- The rvs $\frac{1}{\sqrt{n}}\widetilde{V}_{1,j}(\mathbf{x})$ are asymptotically independent.
- $\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_1(\mathbf{x})\right)\to a^T\Sigma_{\mathbf{x}}a\kappa$.
- For all $\epsilon>0$, $\sum_{j=1}^{k}\mathbf{E}\left(\frac{1}{n}\widetilde{V}_{1,1}^2(\mathbf{x})\mathbb{1}_{\left(\left|\frac{1}{\sqrt{n}}\widetilde{V}_{1,1}(\mathbf{x})\right|>\epsilon\sqrt{\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_1(\mathbf{x})\right)}\right)}\right)\to 0$ (Lindeberg-Feller's condition).

For the first point, it suffices to prove that

$$\left|\mathbf{E}\left(e^{it\sum_{j=1}^{k}\frac{1}{\sqrt{n}}\widetilde{V}_{1,j}(\mathbf{x})}\right)-\prod_{j=1}^{k}\mathbf{E}\left(e^{it\frac{1}{\sqrt{n}}\widetilde{V}_{1,j}(\mathbf{x})}\right)\right|\to 0.$$

Using the inequality of Volkonskii and Rozanov (1959) given in Proposition 1.4 and under Assumption **M**(a), we achieve the intended target

$$\left|\mathbf{E}\left(e^{it\sum_{j=1}^{k}\frac{1}{\sqrt{n}}\widetilde{V}_{1,j}(\mathbf{x})}\right)-\prod_{j=1}^{k}\mathbf{E}\left(e^{it\frac{1}{\sqrt{n}}\widetilde{V}_{1,j}(\mathbf{x})}\right)\right|$$
$$\leq 16(k-1)\alpha(q+1)\leq Ck\alpha(q)\to 0.$$

Regarding the second point, we follow a similar approach as with the term $\frac{1}{\sqrt{n}}\widetilde{V}_2(\mathbf{x})$, we obtain

$$\begin{aligned}\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_1(\mathbf{x})\right) &= \frac{1}{n}\mathbf{Var}\left(\sum_{j=1}^{k}\widetilde{V}_{1,j}(\mathbf{x})\right)\\ &= \frac{1}{n}\sum_{j=1}^{k}\mathbf{Var}\left(\widetilde{V}_{1,j}(\mathbf{x})\right)+\frac{1}{n}\sum_{\substack{j=1\\ j\neq m}}^{k}\sum_{m=1}^{k}\mathbf{Cov}\left(\widetilde{V}_{1,j}(\mathbf{x}),\widetilde{V}_{1,m}(\mathbf{x})\right)\\ &\leq \left(C\frac{kp}{n}h_n^d+C\frac{kp}{n}ph_n^d\right)+\frac{kp}{n}\frac{1}{h_n^{d/a_1}}\sum_{j=q+1}^{\infty}\alpha^{\frac{1}{a_1}}(j).\end{aligned}$$

In light of Assumptions **B**(b), **B**(c) and **M**(b), we get the result.
For the last point, we have:

$$|\widetilde{V}_{1,1}(\mathbf{x})|\leq C\frac{p||K_d||_\infty}{\sqrt{h_n^d}L(a_H)\bar{G}(b_H)},$$

then, using Tchebychev's inequality and under Assumption **B**(d), we attain the desired aim

$$\begin{aligned}&k\mathbf{E}\left(\frac{1}{n}\widetilde{V}_{1,1}^2(\mathbf{x})\mathbb{1}_{\left(\left|\frac{1}{\sqrt{n}}\widetilde{V}_{1,1}(\mathbf{x})\right|>\epsilon\sqrt{\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_1(\mathbf{x})\right)}\right)}\right)\\ &\leq C\frac{kp^2||K_d||_\infty^2}{nh_n^dL^2(a_H)\bar{G}^2(b_H)}\mathbf{P}\left(\left|\frac{1}{\sqrt{n}}\widetilde{V}_{1,1}(\mathbf{x})\right|>\epsilon\sqrt{\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_1(\mathbf{x})\right)}\right)\\ &\leq C\frac{p^2||K_d||_\infty^2}{nh_n^dL^2(a_H)\bar{G}^2(b_H)}\frac{k\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_{1,1}(\mathbf{x})\right)}{\epsilon^2\mathbf{Var}\left(\frac{1}{\sqrt{n}}\widetilde{V}_1(\mathbf{x})\right)}\to 0.\end{aligned}$$

The proof of Lemma 4.4 is complete. □

Proof of Theorem 4.1. Consider the function $\Phi:\mathbb{R}^2\to\mathbb{R}$ defined as $\Phi(x,y)=\frac{x}{y}$, for $y\neq 0$, then

$$\begin{aligned}\widehat{m}_n(\mathbf{x})-m(\mathbf{x}) &= \frac{\widehat{\psi}_{n,1}(\mathbf{x})}{\widehat{\psi}_{n,2}(\mathbf{x})}-\frac{\psi_1(\mathbf{x})}{\psi_2(\mathbf{x})}\\ &= \Phi\left(\widehat{\psi}_{n,1}(\mathbf{x}),\widehat{\psi}_{n,2}(\mathbf{x})\right)-\Phi\left(\psi_1(\mathbf{x}),\psi_2(\mathbf{x})\right).\end{aligned}$$

Applying equation (4.1), Lemma 4.1, Lemma 4.4 and the delta method (see, Shao (2003)), we derive that

$$\sqrt{nh_n^d}\left(\widehat{m}_n(x)-m(x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\nabla\Phi^T\Sigma_x\nabla\Phi\kappa\right),$$

where the gradient $\nabla\Phi^T = \left(\frac{\partial\Phi}{\partial x}, \frac{\partial\Phi}{\partial y}\right)$ is evaluated at point $(\psi_1(\mathbf{x}), \psi_2(\mathbf{x}))$. The proof concludes by noting that

$$\begin{aligned}
\nabla\Phi^T\Sigma_x\nabla\Phi\kappa &= \begin{pmatrix} \frac{1}{\psi_2(\mathbf{x})} & -\frac{\psi_1(\mathbf{x})}{\psi_2^2(\mathbf{x})} \end{pmatrix} \begin{pmatrix} \mu_2(\mathbf{x}) & \mu_3(\mathbf{x}) \\ \mu_3(\mathbf{x}) & \mu_4(\mathbf{x}) \end{pmatrix} \begin{pmatrix} \frac{1}{\psi_2(\mathbf{x})} \\ -\frac{\psi_1(\mathbf{x})}{\psi_2^2(\mathbf{x})} \end{pmatrix} \kappa \\
&= \frac{\psi_2^2(\mathbf{x})\mu_2(\mathbf{x}) - 2\psi_1(\mathbf{x})\psi_2(\mathbf{x})\mu_3(\mathbf{x}) + \psi_1^2(\mathbf{x})\mu_4(\mathbf{x})}{\psi_2^4(\mathbf{x})}\kappa \\
&= \sigma_n^2(\mathbf{x}).
\end{aligned}$$

□

Conclusion and Perspectives

This thesis has investigated the nonparametric estimation of the relative error regression function, with a specific emphasis on left truncated and right censored data under the assumption of α-mixing dependency. The findings of this research contribute significantly to the field of statistical inference, particularly in scenarios where incomplete data are prevalent.

In Chapter 2, we examined the asymptotic properties of the estimator proposed by Jones et al. (2008) for complete data, demonstrating both its almost sure uniform convergence with a rate and its asymptotic normality. The rate of convergence was derived using Bernstein's inequality, while Lyapunov's central limit theorem was employed to establish asymptotic normality. These theoretical results, supported by simulation studies, highlight the robustness of the estimator.

Chapter 3 introduced a novel nonparametric estimator adapted for LTRC data. We established its almost sure uniform convergence with a rate, which was characterized using the Fuk-Nagaev inequality. Practical applications were provided to illustrate the effectiveness of the proposed estimator in real world contexts.

In Chapter 4, we derived the asymptotic normality of the proposed estimator for LTRC data, utilizing Bernstein's large blocks and small blocks technique. We constructed confidence intervals, and the theoretical results were validated through simulation studies and a real world data example, demonstrating the method's practical relevance.

To conclude, we propose several potential research directions to further enhance and expand the results outlined in this thesis:

- While this thesis focused exclusively on covariates in $\mathbb{R}^d$, it is possible to extend our methodology to LTRC and α-mixing data with functional covariates. The case of independent LTRC data with functional covariates has been recently explored by Boucetta et al. (2024) in the context of relative error regression.

- The dependence structure considered in this work is α-mixing, a specific form of weak dependency. The results obtained in Chapters 3 and 4 could be generalized to LTRC data with broader classes of weak dependence structures.

- Following the establishment of asymptotic normality in Chapters 3 and 4, a promising direction for further research would be to derive a Berry-Esseen-type result for the kernel estimator of the regression function. This would enable the quantification of the rate of convergence to the normal distribution, offering deeper insights into the estimator's finite sample properties.

Bibliographie

Adjoudj, L., and Tatachak, A. (2018), "Conditional quantile estimation for truncated and associated data", *Communications in Statistics-Theory and Methods,* 48, 4598 – 4641.

Aït Hennani, L., Lemdani, M., and Ould Saïd, E. (2019), "Robust regression analysis for a censored response and functional regressors", *Journal of Nonparametric Statistics,* 31, 221 – 243.

Altendji, B., Demongeot, J., Laksaci, A., and Rachdi, M. (2018), "Functional Data Analysis: Estimation of the Relative Error in Functional Under Left-truncated", *Journal of Nonparametric Statistics,* 30, 472 – 490.

Azzedine, N., Laksaci, A. and Ould Saïd, E. (2008), "On robust nonparametric regression estimation for a functional regressor", *Statistics* & *Probability Letters,* 78, 3216 – 3221.

Bayarassou, N., Hamrani, F. and Ould Saïd, E. (2023), "Nonparametric relative error estimation of the regression function for left runcated and right censored time series data ", *Journal of Nonparametric Statistics,* https://doi.org/10.1080/10485252.2023.2241572.

Benseradj, H., and Guessoum, Z. (2022), "Strong Uniform Consistency Rate of An M-estimator of Regression Function for Incomplete Data Under α-mixing Condition ", *Communications in Statistics -Theory and Methods,* 51, 2082 – 2115.

Bey, S., Guessoum, Z., and Tatachak, A. (2022), "Kernel regression estimation for LTRC and associated data", *Communications in Statistics-Theory and Method,* 52, 6381 – 6406.

Boente, G., and Rodriguez, D. (2006), "Robust estimators of high order derivatives of regression functions ", *Statistics* & *probability letters,* 76, 1335 – 1344.

Bouabsa, W. (2021), "Nonparametric Relative Error Estimation Via Functional Regressor by the K Nearest Neighbors Smoothing Under Truncation Random Data ", *Applications and Applied Mathematics,* 16, 97 – 116.

Boucetta, A., Guessoum, Z., and Ould-Said, E. (2024), "Strong convergence of a nonparametric relative error regression estimator under missing data with functional predictors ", *Journal of the Korean Statistical Society,* 1 – 33.

Bouhadjera, F., and Ould Saïd, E. (2019), "On the strong uniform consistency for relative error of the regression function estimator for censoring times series model", arXiv preprint arXiv:1910.01964.

Bouhadjera, F., Ould Saïd, E., and Remita, M.R. (2019), "Nonparametric relative error estimation of the regression function for censored data ", https://hal.archives-ouvertes.fr/hal-01994512.

Bouhadjera, F., Ould Saïd, E., and Remita, M.R. (2022), "Strong Consistency of the Nonparametric Local Linear Regression Estimation under Censorship Model ", *Communications in Statistics -Theory and Methods,* 51, 7056 – 7072.

Bouhadjera, F., Lemdani, M., and Ould Saïd, E. (2023), "Strong uniform consistency of the local linear relative error regression estimator under left truncation", *Statistical Papers,* 64, 421 – 447.

Bulinski, A., and Shashkin, A. (2007), "Limit theorems for associated random fields and related systems", *Advanced Series on Statistical Science* & *Applied Probability.*

Cai, Z. (2001), "Estimating a distribution function for censored time series data ", *Journal of Multivariate Analysis*, 78, 299 – 318.

Cai, Z., and Roussas, G. G. (1998), "Kaplan-Meier estimator under association", *Journal of Multivariate Analysis,* 67, 318 – 348.

Cao, R., Janssen, P., and Veraverbeke, N. (2005), "Relative hazard rate estimation for right censored and left truncated data", *Test,* 14, 257 – 280.

Carbonez, A., Gyorfi, L., and Van Der Meulen, E.C. (1995), "Partitioning Estimates of a Regression Function under Random Censoring", *Statistics & Risk Modeling,* 13, 21 – 38.

Chen, Q., and Dai, Y. (2003), "Kernel estimation of higher derivative of density and hazard rate function for truncated and censored dependent data ", *Acta Mathematica Scientia*, 23, 477 – 486.

Demongeot, J., Hamie, A., Laksaci, A., and Rachdi, M. (2016), "Relative-Error Prediction in Nonparametric Functional Statistics: Theory and Practice", *Journal of Multivariate Analysis,* 146, 261 – 268.

Doukhan, P. (1994), "Mixing: Properties and Examples, Lecture Notes in Statistics 85", *Springer-Verlag, New York.*

Doukhan, P., and Louhichi, S. (1999), "A new weak dependence condition and applications to moment inequalities", *Stochastic processes and their applications*, 84, 313 – 342.

Esary, J. D., Proschan, F., and Walkup, D. W. (1967), "Association of random variables, with applications", *The Annals of Mathematical Statistics,* 38, 1466 – 1474.

Derrar, S., Laksaci, A., and Ould Saïd, E. (2015), "On the nonparametric estimation of the functional ϕ-regression for a random left-truncation model", *Journal of statistical Theory and Practice,* 9, 823 – 849.

Fan, J., Hu, T. C., and Truong, Y. K. (1994), " Robust non-parametric function estimation", *Scandinavian journal of statistics,* 433 – 446.

Galton, F. (1890), "Kinship and correlation", *The North American Review,* 419 – 431.

Gauss, C. F. (1809), "Theoria motus corporum coelestium in sectionibus conicis solem ambientium auctore Carolo Friderico Gaus", *Frid. Perthes et IH Besser.*

Gijbels, I., and Wang, J.L. (1993), "Strong representations of the survival function estimator for truncated and censored data with applications", *Journal of Multivariate Analysis,* 47, 210 – 229.

Guessoum, Z., and Ould Saïd, E. (2009), "On nonparametric estimation of the regression function under random censorship model", *Statistics & Decision,* 26, 159 – 177.

Guessoum, Z., and Ould Saïd, E. (2010), "Kernel regression uniform rate estimation for censored data under α-mixing condition ", *Electronic Journal of Statistics,* 4, 117 – 132.

Guessoum, Z., and Ould Saïd, E. (2012), "Central limit theorem for the kernel estimator of the regression function for censored time series", *Journal of Nonparametric Statistics,* 24, 379 – 397.

Guessoum, Z., Ould Saïd, E., Sadki, O., and Tatachak, A. (2012), "A note on the Lynden-Bell estimator under association", *Statistics & Probability Letters,* 82, 1994 – 2000.

Hamrani, F., and Guessoum, Z. (2017), "Convergence rate of the kernel regression estimator for associated and truncated data", *Journal of Nonparametric Statistics,* 29, 425 – 446.

Härdle, W. (1984), "Robust regression function estimation", 14, 69 – 180.

Härdle, W. (1990), "Applied nonparametric regression", *Cambridge university press.*

He, S., and Yang, G.L. (1998), " Estimation of the truncation probability in the random truncation model ", *The Annals of Statistics,* 26, 1011 – 1027.

Huber, P.J. (1964), " Robust estimation of a location parameter", *Breakthroughs in statistics: Methodology and distribution,* 35, 73 – 101.

Hyde, J. (1977) "Testing survival under right censoring and left truncation ", *Biometrika,* 64, 225 – 230.

Jones, M.C., Park, H., Shin, K.I., Vines, S.K., and Jeong, S.O. (2008), "Relative Error Prediction Via Kernel Regression Smoothers ", *Journal of Statistical Planning and Inference,* 138, 2887 – 2898.

Kohler, M., Máthé, K., and Pintér, M. (2002), "Prediction from randomly right censored data", *Journal of Multivariate Analysis,* 80, 73 – 100.

Legendre, A. M. (1805), "Nouvelles méthodes pour la détermination des orbites des comètes", *F. Didot, Paris.*

Lemdani, M., and Ould Saïd, E. (2017), "Nonparametric robust regression estimation for censored data", *Statistical Papers,* 58, 505 – 525.

Liang, H. Y., Li, D., and Qi, Y. (2009), "Strong convergence in nonparametric regression with truncated dependent data", *Journal of multivariate analysis,* 100, 162 – 174.

Loève, M. (1963), "Probability Theory", *Springer-Verlag, New York.*

Lynden-Bell, D. (1971), "A method of allowing for known observational selection in small samples applied to 3CR quasars", *Monthly Notices of Royal Astronomical Society,* 155, 95 – 118.

Masry, E. (1986), "Recursive probability density estimation for weakly dependent stationary processes ", *IEEE Transactions on Information Theory*, 32, 254 – 267.

Masry, E. (2002), "Multivariate probability density estimation for associated processes: strong consistency and rates", *Statistics & Probability Letters,* 58, 205 – 219.

Menni, N., and Tatachak, A. (2016), "A note on estimating the conditional expectation under censoring and association: strong uniform consistency", *Statistical Papers,* 59, 1009 – 1030.

Molanes-López, E. M., and Cao, R. (2008), "Relative density estimation for left truncated and right censored data", *Journal of Nonparametric Statistics,* 20, 693 – 720.

Nadaraya, E. A. (1964), "On estimating regression", *Theory of Probability* & *Its Applications,* 9, 141 – 142.

Ould Saïd, E., and Lemdani, M. (2006), "Asymptotic Properties of a Nonparametric Regression Function Estimator with Randomly Data ", *Annals of the Institute of Statistical Mathematics,* 58, 357 – 378.

Park, H., and Stefanski, L.A. (1998), "Relative-error Prediction", *Statistics* & *Probability Letters,* 40, 227 – 236.

Révész, P. (1973), "Robbins-Monro procedure in a Hilbert space and its application in the theory of learning processes I", *Stud. Sci. Math. Hungar.*, 8, 391 – 398.

Rio, E. (2000), "Théorie asymptotique des processus aléatoires faiblement dépendants", *Collection Mathématiques* & *Applications, Springer.*

Shao, J. (2003), "Mathmetical Statistics", *Springer, New York.*

Silverman, B.W. (1986), "Density estimation for statistics and data analysis", *Monographs on Statistics and Applied Probability, Chapman* & *Hall/CRC.*

Stone, C. J. (1977), "Consistent nonparametric regression", *The annals of statistics,* 595 – 620.

Tsybakov, A. B. (2009), " Introduction to nonparametric estimation", *Springer.*

Venables, W. N., and Ripley, B. D. (2002), "Modern Applied Statistics with S ", *Springer.*

Volkonskii, V.A., and Rozanov, Y.A. (1959), "Some limit theorems for random functions I ", *Theory of Probability & Its Applications*, 4, 178 – 197.

Wang, J. F., and Liang, H. Y. (2012), "Asymptotic properties for an M-estimator of the regression function with truncation and dependent data", *Journal of the Korean Statistical Society,* 41, 351 – 367.

Watson, G. S. (1964), "Smooth regression analysis," *Sankhyā: The Indian Journal of Statistics,* 26, 359 – 372.

Tsai, W.Y., Jewell, N.P., and Wang, M.C. (1987), "A note on the product-limit estimator under right censoring and left truncation ", *Biometrika,* 74, 883 – 886.