People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research

University of Saida - Dr Moulay Tahar
Faculty of MIT
Department of Mathematics

Dissertation presented with a view to obtaining the diploma of

Academic Master

Discipline : MATHEMATICS
Specialty: Statistical Stochastic Analysis of Processes and Applications

by

Nadhira MEKKAOUI

Under the supervision of

Dr. Fatima BENZIADI

Theme:

Conditional models estimation in the single functional index with missing data

Defended on 15/06/2026 before the jury composed of

Dr. Chafiaa AYHAR	University Center of El Bayadh Nour El Bachir	President
Dr. Fatima BENZIADI	University of Saida Dr. Moulay Tahar	Supervisor
Dr. Imene BOUAZZA	University of Saida Dr. Moulay Tahar	Examiner

University year: 2025/2026

Acknowledgements

First and foremost, I express my sincere gratitude to ***Allah****, whose endless blessings, strength, and guidance have made this achievement possible.*
I extend my deepest gratitude to my supervisor, Dr. Fatima BENZIADI, for her unwavering support, and continuous encouragement.
I take this opportunity to thank the esteemed members of the committee, Dr. AYHAR Chafia and Dr. BOUAZZA Imene who generously dedicated their time and expertise to evaluate and appraise this work.
A special thanks to all professors of the Laboratory of Stochastic Models, Statistic and Applications of the Dr. Moulay Tahar University of Saida.

Dedication

To my beloved parents, whose constant support and sacrifices.

To my dear sisters and brothers for their constant encouragement and moral support.

To my beloved children.

To all my colleagues of Master 2 ASSPA.

And to all those who seek knowledge and aspire to grow.

Contents

Introduction

Statistical analysis of functional variables has considerably grown over the past two decades. Indeed, immense innovations in measuring devices have emerged, permitting the monitoring of numerous objects in a continuous manner, such as stock market indices, pollution levels, climatological patterns, satellite images, and many others. Consequently, a new branch of statistics, known as functional statistics, has developed to treat observations as functional random elements [17].

The study of statistical models for functional data has been the subject of several recent works and developments. The first results on conditional models were obtained by [14], where these authors established the almost complete convergence rate of kernel estimators for the conditional distribution function, the conditional density and its derivatives, the conditional mode, and the conditional quantiles. As a conditional nonparametric model, regression was one of the first predictive analysis tools. Conditional mode estimation is particularly useful in prediction settings, providing an alternative approach to classical regression estimation. For more recent advances on this topic, we refer to [12]. In functional statistics, this model was introduced by [6], and the nonparametric study of this model has been considered by [17].

Mode regression is a common way to describe the dependence structure between a response variable Y and some covariate X. Unlike the regression function (which is defined as the conditional mean) that relies only on the central tendency of the data, the conditional mode function allows analysts to estimate the functional dependence between variables for all portions of the conditional distribution of the response variable. Moreover, compared with the standard approach based on functional conditional mean prediction that is sensitive to outliers, functional conditional mode prediction can be seen as a reasonable alternative to conditional mean because of its robustness.

The estimation of conditional densities for functional data has a rich history in the statistical literature. Foundational work in [14] established the almost complete convergence of kernel estimators under independent and mixing conditions. Subsequent developments include asymptotic normality results in [12], and extensions to the single-index model in [2]. In [24, 25] nonparametric estimation methods for ergodic functional data have been developed.

Conditional quantiles are well known for their robustness to heavy-tailed error distributions and outliers, which allows them to be considered as a useful alternative to the regression function [7]. Conditional quantiles and conditional mode are used in finance and/or insurance to model the risks of extreme values. Furthermore, conditional quantiles can be used to detect outliers in the data as well as to establish probabilistic forecasts. For these theoretical and application reasons, the statistical community has placed great interest in estimating conditional quantiles; specifically, the conditional median function is an interesting alternative predictor to the conditional mean due to its robustness to the presence of outliers [8].

Estimation of the conditional mode of a scalar response given a functional covariate has attracted the attention of many researchers. [13] introduced a nonparametric estimator of the conditional mode when data are dependent and established its rate of almost complete consistency. [12] established the asymptotic normality of the kernel conditional mode estimator under an α-mixing assumption. In the censored case, [33] stated the uniform strong consistency with rates of the kernel estimator of the conditional mode function. In this context, we refer to [26] for the estimation of conditional quantiles. Other authors have been interested in the estimation of conditional models when the observations are censored or truncated [27, 21, 22, 23, 34, 39].

The ergodic theory has appeared in statistical mechanics, notably in Maxwell's and Gibbs's theories. It is necessary to make a sort of logical transition between the average behavior of the set of dynamic systems and the temporal average of the behaviors of a single dynamic system. It is derived from an ingenious hypothesis used for a long time without justifying it, and in various forms. In the context of ergodic functional data with censored observations, the literature is very restricted. We refer to [7], who studied the asymptotic properties of the kernel-type estimator of the conditional quantiles when the response variable is right-censored and the data are sampled from an underlying stationary ergodic process.

The single-index model represents one of the well-known semi-parametric models, which is very popular in the economics community as it allows reduction of the dimensionality of the covariate space while offering flexibility in describing the relationship between the response and the covariate through an unknown link function. The statistical study of these models, in the context of vectorial explanatory random variables, was initiated by [19]. [20] provides both new theoretical and bibliographic elements. Several authors have worked on single functional index models; we can cite [1], [3], and [5].

However, in many practical applications such as pharmaceutical tracing tests, reliability tests, and so on, some pairs of observations may be incomplete; in this case, we refer to missing data. Many examples of missing data and their statistical inferences for regression models can be found in the statistical literature when explanatory variables are of finite dimensionality [9, 31, 32, 40, 28, 10, 11]. When the explanatory variable is infinite dimensional or of a functional nature, very few studies have been reported to investigate the statistical properties of functional nonparametric regression models for missing data. Recently, [16] first proposed to estimate the mean of a scalar response based on an i.i.d. functional sample in which explanatory variables are observed for every subject, while part of the responses are missing at random (MAR) for some of them. This generalized the results of [9] to the case where the explanatory variables are of functional nature.

In practice, this study has great importance because it permits the construction of a prediction method based on the conditional mode estimator. Moreover, in the case where the functional single-index is unknown, our estimate can be used to estimate this parameter via the pseudo-maximum likelihood estimation method. To the best of our knowledge, the estimation of the nonparametric conditional density in the functional single-index structure combining missing data and stationary ergodic processes with functional nature has not been studied in the statistical literature. Thus, in the present work, we investigate conditional density estimation when the data are both MAR and ergodic. First, an estimator of the regression operator in the functional single-index model, with a scalar response and a functional covariate assumed to be sampled from a stationary and ergodic process, is constructed. Our aim is to develop a functional methodology for dealing with MAR samples in nonparametric problems (namely in conditional mode estimation). Then, the asymptotic properties of the estimator are obtained under some mild conditions.

Here, we consider a model in which the response variable is missing. Besides the infinite-dimensional character of the data, we avoid here the widely used strong mixing condition and its variants to measure dependency and the very involved probabilistic calculations that it implies. Therefore, we consider, in our setting, the ergodic property to allow the maximum possible generality with regard to the dependence setting. Further motivations for considering ergodic data are discussed in [24], where details defining the ergodic property of processes are also given. As far as we know, the estimation of conditional quantiles combining missing data, ergodic theory, and functional data with single-index structure has not been studied in the statistical literature. This work extends, to the functional single-index model case, the work of [29, 30].

This work is organized as follows. Chapter 1 provides the background, beginning with functional spaces such as Banach and Hilbert spaces, orthonormal bases, linear operators, and duality. It then introduces functional data analysis, semi-parametric models with emphasis on the single-index model, missing data mechanisms with focus on the MAR assumption, conditional density and distribution functions, ergodic processes, and convergence concepts. Chapter 2 develops kernel estimation of conditional density in the single-index model, first under complete data and then under missing at random data, including asymptotic results and conditional mode estimation. Chapter 3 addresses kernel estimation of conditional distribution functions in the single-index model, covering complete data and MAR data, with results for conditional quantiles. Chapter 4 presents the numerical study, detailing the simulation design, missing data mechanisms, estimation procedures, and numerical results, followed by discussion and conclusion. Finally, Appendix A contains technical lemmas and detailed proofs supporting the theoretical results.

Chapter 1

Background

1.1 Functional Spaces: Hilbert and Banach Spaces

The mathematical framework for functional data analysis relies on the theory of functional spaces. This section presents the essential concepts of Hilbert and Banach spaces that are necessary for understanding the theoretical developments in this work.

1.1.1 Banach Spaces

Definition 1.1 (Banach Space)**.** A Banach space is a complete normed vector space $(\mathcal{B}, \|\cdot\|)$. That is:

(i) $\mathcal{B}$ is a vector space over $\mathbb{R}$ or $\mathbb{C}$;

(ii) $\|\cdot\| : \mathcal{B} \to \mathbb{R}^+$ is a norm satisfying:

$$\|x\| = 0 \iff x = 0, \quad \|\lambda x\| = |\lambda|\|x\|, \quad \|x + y\| \leq \|x\| + \|y\|;$$

(iii) Every Cauchy sequence in $\mathcal{B}$ converges to an element of $\mathcal{B}$ (completeness property).

Banach spaces provide a natural framework for studying functional data as they allow for infinite-dimensional analysis while maintaining completeness. Common examples include:

- The space $L^p([0,1])$ of functions f such that $\int_0^1 |f(t)|^p dt < \infty$, equipped with the norm $\|f\|_p = (\int_0^1 |f(t)|^p dt)^{1/p}$;
- The space $C([0,1])$ of continuous functions on $[0,1]$ equipped with the supremum norm $\|f\|_\infty = \sup_{t\in[0,1]} |f(t)|$;
- The Sobolev spaces $W^{k,p}([0,1])$ of functions with derivatives up to order k in L^p.

1.1.2 Hilbert Spaces

Definition 1.2 (Hilbert Space)**.** A Hilbert space $\mathcal{H}$ is a Banach space whose norm is induced by an inner product. That is, there exists an inner product $\langle\cdot,\cdot\rangle : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ satisfying:

(i) $\langle x, x\rangle \geq 0$ and $\langle x, x\rangle = 0 \iff x = 0$;

(ii) $\langle x, y\rangle = \langle y, x\rangle$ (symmetry);

(iii) $\langle \lambda x + \mu y, z\rangle = \lambda\langle x, z\rangle + \mu\langle y, z\rangle$ (bilinearity);

(iv) The norm is given by $\|x\| = \sqrt{\langle x, x\rangle}$;

(v) $\mathcal{H}$ is complete with respect to this norm.

Hilbert spaces are special cases of Banach spaces where the geometry is richer due to the inner product structure. Key properties include:

- **Parallelogram law**: $\|x + y\|^2 + \|x - y\|^2 = 2(\|x\|^2 + \|y\|^2)$.
- **Pythagorean theorem**: If $\langle x, y\rangle = 0$, then $\|x + y\|^2 = \|x\|^2 + \|y\|^2$.
- **Cauchy-Schwarz inequality**: $|\langle x, y\rangle| \leq \|x\|\|y\|$.

The most common Hilbert space in functional data analysis is $L^2([0, 1])$ with inner product

$$\langle f, g\rangle = \int_0^1 f(t)g(t)dt.$$

1.1.3 Orthonormal Bases and Dimensionality

Definition 1.3 (Orthonormal Basis)**.** A sequence $\{e_k\}_{k=1}^{\infty}$ in a Hilbert space $\mathcal{H}$ is an orthonormal basis if:

(i) $\langle e_i, e_j\rangle = \delta_{ij}$ (Kronecker delta);

(ii) Every $x \in \mathcal{H}$ can be represented as $x = \sum_{k=1}^{\infty}\langle x, e_k\rangle e_k$.

Any separable Hilbert space admits an orthonormal basis. This allows the representation of functional data as sequences of coefficients:

$$X(t) = \sum_{k=1}^{\infty} \xi_k \phi_k(t), \quad \text{where } \xi_k = \langle X, \phi_k\rangle.$$

1.1.4 Linear Operators in Hilbert Spaces

Definition 1.4 (Bounded Linear Operator)**.** A linear operator $T : \mathcal{H} \to \mathcal{H}$ is bounded if there exists a constant $C > 0$ such that

$$\|Tx\| \leq C\|x\| \quad \forall x \in \mathcal{H}.$$

The operator norm is defined as $\|T\| = \sup_{\|x\|=1} \|Tx\|$.

In the context of the single-index model, the index $\theta \in \mathcal{H}$ defines a bounded linear functional via the inner product:

$$\varphi_\theta(x) = \langle x, \theta\rangle.$$

1.1.5 Duality and the Riesz Representation Theorem

Theorem 1.5 (Riesz Representation Theorem)**.** *Let $\mathcal{H}$ be a Hilbert space and let $f : \mathcal{H} \to \mathbb{R}$ be a continuous linear functional. Then there exists a unique vector $\theta \in \mathcal{H}$ such that*

$$f(x) = \langle x, \theta \rangle \quad \forall x \in \mathcal{H}.$$

Moreover, $\|f\| = \|\theta\|$.

This theorem is fundamental for the single-index model because it guarantees that any continuous linear functional on a Hilbert space can be represented as an inner product with a unique index parameter θ.

1.1.6 Relationship between Hilbert and Banach Spaces

The relationship between Hilbert and Banach spaces can be summarized as follows:

$$\text{Hilbert Space} \subsetneq \text{Banach Space} \subsetneq \text{Normed Space}.$$

- Every Hilbert space is a Banach space, but the converse is false. For example, $L^p([0, 1])$ with $p \neq 2$ is a Banach space but not a Hilbert space.
- Hilbert spaces have a natural notion of angle and orthogonality, which is absent in general Banach spaces.
- The norm in a Hilbert space satisfies the parallelogram law, which characterizes Hilbert spaces among Banach spaces.

1.1.7 Why Hilbert Spaces for Functional Data?

In this work, we assume that the functional covariate X takes values in a separable Hilbert space $\mathcal{H}$. This choice is motivated by several considerations:

1. **Inner product structure**: The single-index model relies on the inner product $\langle X, \theta \rangle$ to project the infinite-dimensional covariate onto a one-dimensional index.
2. **Riesz representation**: The Riesz representation theorem ensures that any continuous linear functional can be represented via an inner product, making the model identifiable.
3. **Orthonormal basis**: Separable Hilbert spaces admit orthonormal bases, allowing for dimension reduction techniques such as functional principal component analysis (FPCA).
4. **Computational tractability**: Many algorithms for functional data analysis (e.g., smoothing, FPCA) are naturally formulated in Hilbert spaces.
5. **Theoretical convenience**: The geometry of Hilbert spaces simplifies the derivation of asymptotic properties, particularly when using martingale techniques.

This functional analytical framework provides the mathematical rigor necessary for the asymptotic analysis developed in subsequent chapters.

1.2 Introduction to Functional Data Analysis

1.2.1 Historical Development

Functional data analysis (FDA) has emerged as a contemporary area of statistical research over the past two decades. The field was popularized by [35, 36, 37] through their seminal works on the applied aspects of functional data. Early theoretical developments were made by [4] for linear processes in function spaces, while [17] provided a comprehensive treatment of nonparametric functional data analysis.

The evolution of FDA can be traced through several key periods:

- **1990s**: Foundational work by Ramsay and Silverman establishing basic concepts and methodologies
- **2000s**: Development of nonparametric approaches by Ferraty, Vieu, and collaborators
- **2010s**: Extension to semi-parametric models and dependent functional data
- **2020s**: Integration with machine learning and high-dimensional statistics

1.2.2 Examples of Functional Data

Functional data appear naturally in numerous scientific domains where observations are collected continuously over time, space, or other continua. Common examples [35, 36, 37] include:

- **Spectrometric data**: Near-infrared spectra of food samples, where each observation is a curve representing absorbance at different wavelengths.
- **Annual electricity consumption**: Monthly electricity usage data where each year forms a functional observation.
- **Climatological curves**: Temperature or precipitation patterns recorded daily over years.
- **Growth curves**: Height or weight measurements of individuals over time.
- **Stock market indices**: Continuous price trajectories over trading days.

1.2.3 Challenges in Functional Data Analysis

The analysis of functional data presents unique challenges that distinguish it from classical multivariate statistics:

- **Infinite-dimensional nature**: Functional observations lie in infinite-dimensional spaces, requiring careful handling of topology and geometry.
- **Curse of dimensionality**: Traditional statistical methods deteriorate as dimension increases; functional data represent an extreme case where dimension is infinite.
- **Choice of metric**: Unlike finite dimensions where metrics are equivalent, in infinite dimensions the choice of metric fundamentally affects the analysis.

- **Dependence structures**: Functional data often exhibit temporal or spatial dependence that must be accounted for.
- **Missing data**: Functional observations may be incomplete, with missing segments or entirely missing curves.

1.3 Semi-parametric Models

1.3.1 Semi-parametric Approaches

Statistical modeling approaches can be classified along a spectrum from fully parametric to fully nonparametric [1].

1.3.2 The Single-Index Model

The single-index model is one of the most popular semi-parametric approaches in statistics and econometrics.

Definition 1.7 (Single-Index Model)**.** Let (X, Y) be a pair of random variables where $Y \in \mathbb{R}$ is a scalar response and X takes values in a Hilbert space $\mathcal{H}$. The single-index model assumes that the conditional distribution of Y given X depends only on a linear combination $\langle \theta, X \rangle$, where $\theta \in \mathcal{H}$ is an unknown index parameter. Specifically, the conditional density satisfies:

$$f(y \mid X = x) = f(y \mid \langle \theta, x \rangle)$$

for some unknown function f.

The motivation for the single-index model stems from dimension reduction, as the infinite-dimensional predictor X is projected onto a one-dimensional index, making the problem tractable. Many classical models are special cases, including linear models, generalized linear models, and transformation models.

1.3.3 Estimation of the Index Parameter θ

In the single-index model, the parameter θ plays a central role as it defines the projection $\langle X, \theta \rangle$ on which the conditional distribution of Y depends. Since θ is unknown, it must be estimated prior to constructing kernel estimators. We present two common approaches:

Method 1: Least Squares / Average Derivative Estimation (ADE). This method relies on minimizing squared residuals or exploiting average derivatives to recover the direction of θ. Specifically, one can consider the minimization problem

$$\widehat{\theta}_{LS} = \arg\min_{\theta \in \Theta} \sum_{i=1}^{n} \Big(Y_i - g(\langle X_i, \theta \rangle) \Big)^2,$$

[1]

Definition 1.6 (Semi-parametric Models)**.** Semi-parametric models combine parametric and nonparametric components, offering a compromise between flexibility and interpretability. They contain a finite-dimensional parameter of interest and an infinite-dimensional nuisance parameter.

where $g(\cdot)$ is a smooth link function. Identification requires a normalization constraint, typically $\|\theta\| = 1$. Alternatively, the ADE approach uses the identity

$$\mathbb{E}\left[\frac{\partial}{\partial x}\mathbb{E}[Y \mid X = x]\right] \propto \theta,$$

which allows recovery of the direction of θ from average derivatives.

Method 2: Maximum Likelihood / Pseudo-Likelihood Estimation. When a parametric assumption is made on the conditional distribution of Y given $\langle X, \theta\rangle$, θ can be estimated by maximizing the likelihood. For instance, if

$$Y \mid X \sim f(y \mid \langle X, \theta\rangle; \beta),$$

with parameter β, then

$$\widehat{\theta}_{ML} = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \log f(Y_i \mid \langle X_i, \theta\rangle; \beta).$$

In semiparametric settings, pseudo-likelihood approaches are often employed, where the conditional density f is replaced by a kernel-smoothed estimator $\widehat{f}_n$, yielding

$$\widehat{\theta}_{PL} = \arg\max_{\theta\in\Theta} \sum_{i=1}^{n} \log \widehat{f}_n(Y_i \mid \langle X_i, \theta\rangle).$$

These estimation procedures provide the parametric component of the model, which is then combined with nonparametric kernel methods to estimate conditional densities and distribution functions under both complete and missing data.

1.4 Missing Data Mechanisms

Missing data are ubiquitous in practical applications, arising from various causes such as non-response, dropout in longitudinal studies, equipment failure, or data corruption. Understanding the mechanism that generates missingness is crucial for valid statistical inference.

1.4.1 Complete Case Analysis

The simplest approach to handling missing data is complete case analysis (also called listwise deletion), where only observations with complete records are analyzed.

Definition 1.8 (Complete Case Analysis)**.** Complete case analysis uses only those observations for which all variables are observed. If the original sample size is n and m observations are complete, the analysis is based on these m observations.

While simple to implement, complete case analysis has serious drawbacks:

- Loss of efficiency due to reduced sample size
- Potential bias if the missingness is not completely at random
- Inefficient use of partially observed information

1.4.2 MCAR, MAR, MNAR: Definitions and Examples

Rubin (1976) introduced a taxonomy of missing data mechanisms that has become standard in the literature.

Definition 1.9 (Missing Completely At Random (MCAR)). Missingness is MCAR if the probability of missingness does not depend on either the observed or unobserved values:

$$\mathbb{P}(\delta = 1 \mid X, Y) = \mathbb{P}(\delta = 1)$$

where δ is the indicator that Y is observed (take 0 when the data are missing and 1 when the data are observed).

Example 1.10 (MCAR). In a survey, some respondents accidentally skip a question due to a printing error. The probability of missingness is the same for all respondents, regardless of their characteristics or answers.

Definition 1.11 (Missing At Random (MAR)). Missingness is MAR if the probability of missingness depends only on the observed variables and not on the missing values themselves:

$$\mathbb{P}(\delta = 1 \mid X, Y) = \mathbb{P}(\delta = 1 \mid X)$$

Example 1.12 (MAR). In a study of income, individuals with higher education levels may be less likely to report their income, but conditional on education, the probability of missing income does not depend on the actual income amount.

In this work, we focus on the MAR mechanism, specifically:

$$\mathbb{P}(\delta = 1 \mid \langle X, \theta \rangle = \langle x, \theta \rangle, Y = y) = \mathbb{P}(\delta = 1 \mid \langle X, \theta \rangle = \langle x, \theta \rangle) = p(\theta, x)$$

Definition 1.13 (Missing Not At Random (MNAR)). Missingness is MNAR if the probability of missingness depends on the missing values themselves, even after conditioning on observed variables:

$$\mathbb{P}(\delta = 1 \mid X, Y) \text{ depends on } Y$$

Example 1.14 (MNAR). In a survey about depression, individuals with severe depression may be less likely to respond to the survey. The probability of non-response depends on the depression level, which is partially unobserved.

1.4.3 Why MAR is Important in Practice

The MAR assumption occupies a central position in missing data analysis for several reasons:

1. **Realism**: MAR is more realistic than MCAR in many applications. Missingness often depends on observed characteristics (e.g., certain demographic groups are less likely to respond).

2. **Tractability**: Unlike MNAR, MAR allows valid inference without specifying a model for the missingness mechanism, provided the observed variables are properly accounted for.

3. **Identifiability**: Under MAR, the parameters of interest are identifiable from the observed data distribution.

4. **Methodological development**: Many sophisticated methods (multiple imputation, inverse probability weighting, likelihood-based methods) rely on the MAR assumption.

In the functional data context, Ferraty, Sued, and Vieu (2013) first considered MAR for functional covariates, establishing the foundation for subsequent work including this work.

1.5 Conditional Density and Distribution Functions

The conditional density and distribution functions are fundamental objects in statistics, characterizing the entire distribution of a response variable given covariates.

Definition 1.15 (Conditional Density Function)**.** Let (X, Y) be a pair of random variables where $Y \in \mathbb{R}$ and X takes values in a semi-metric space $(\mathcal{F}, d)$. The conditional density function $f^x(y)$ of Y given $X = x$ is defined by:

$$f^x(y) = \frac{d}{dy}\mathbb{P}(Y \leq y \mid X = x)$$

for all $y \in \mathbb{R}$ where the derivative exists.

Definition 1.16 (Conditional Mode)**.** For a fixed $x \in \mathcal{H}$ and index $\theta \in \mathcal{H}$, the conditional mode is defined as

$$M_\theta(x) = \arg\sup_{y \in \mathcal{S}_\mathbb{R}} f(\theta, y, x),$$

where $\mathcal{S}_\mathbb{R}$ is a fixed compact subset of $\mathbb{R}$.

Definition 1.17 (Conditional Distribution Function)**.** The conditional distribution function $F^x(y)$ is defined as:

$$F^x(y) = \mathbb{P}(Y \leq y \mid X = x) = \int_{-\infty}^{y} f^x(t)dt$$

Definition 1.18. (conditional quantile) Let $\tau \in (0, 1)$. The conditional quantile of order τ is defined as

$$q_\tau(\theta, x) = \inf\big\{y \in \mathbb{R} : F(\theta, y, x) \geq \tau\big\},$$

where $F(\theta, y, x)$ is the conditional distribution function.

1.5.1 Kernel Estimation Method

The kernel method (also known as the Parzen-Rosenblatt method) is the most common non-parametric approach for density estimation.

Definition 1.19 (Kernel Density Estimator)**.** For a univariate random sample $X_1, \ldots, X_n$, the kernel density estimator is:

$$\widehat{f}(x) = \frac{1}{nb}\sum_{i=1}^{n} H\left(\frac{x - X_i}{b}\right)$$

where H is a kernel function (typically a symmetric probability density) and $b > 0$ is a bandwidth parameter.

For conditional density estimation with functional covariate $(X_i, Y_i) \in (H \times \mathbb{R})$, the kernel estimator takes the form:

$$\widehat{f^x}(y) = \frac{\sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{b_K}\right) H\left(\frac{y-Y_i}{b_H}\right)}{b_H \sum_{i=1}^{n} K\left(\frac{d(x,X_i)}{b_K}\right)} \tag{1.1}$$

where:

- K is a kernel function for the functional covariate

- H is a kernel function for the response
- b_K and b_H are bandwidth parameters
- d is a semi-metric on the functional space

1.6 Ergodic Processes

In this work, we adopt the ergodic property, which provides maximum generality with regard to dependence.

Definition 1.20 (Ergodic Process)**.** A stationary process $\{Z_i, i \geq 1\}$ is ergodic if for any measurable function h with $\mathbb{E}|h(Z_1)| < \infty$:

$$\frac{1}{n}\sum_{i=1}^{n} h(Z_i) \xrightarrow{a.s.} \mathbb{E}[h(Z_1)] \quad \text{as } n \to \infty$$

Ergodicity implies that time averages converge to ensemble averages. Advantages of the ergodic framework include generality, applicability to many real-world processes, and the ability to use martingale techniques.

1.7 Convergence Concepts

Definition 1.21 (Almost Sure Convergence)**.** A sequence of random variables $\{Z_n\}$ converges almost surely to a random variable Z, denoted $Z_n \xrightarrow{a.s.} Z$, if

$$\mathbb{P}\left(\lim_{n\to\infty} Z_n = Z\right) = 1$$

Definition 1.22 (Almost Complete Convergence)**.** A sequence $\{Z_n\}$ converges almost completely to Z, denoted $Z_n \xrightarrow{a.co.} Z$, if for every $\epsilon > 0$:

$$\sum_{n=1}^{\infty} \mathbb{P}(|Z_n - Z| > \epsilon) < \infty$$

Almost complete convergence implies almost sure convergence by the Borel-Cantelli lemma and is particularly useful in nonparametric functional data analysis.

Chapter 2

Kernel Estimation of Conditional Density in the Single-Index Model

This chapter establishes the asymptotic properties of the kernel estimator for the conditional density function in the functional single-index model under stationary ergodic dependence. We first treat the case of completely observed data, then extend the results to the missing-at-random (MAR) setting. Finally, we derive the asymptotic behaviour of the conditional quantile estimator under MAR, which is obtained as the generalized inverse of the conditional distribution estimator studied in Chapter 3.

2.1 Under Complete Data

2.1.1 Model and Assumptions

Consider a random pair (X, Y) where X takes values in $\mathcal{H}$ and Y takes values in $\mathbb{R}$. Let $(X_i, Y_i)_{i=1,\ldots,n}$ be a sequence of stationary and ergodic functional samples, identically distributed as (X, Y).

We assume the existence of a fixed but unknown functional index $\theta \in \mathcal{H}$ such that the conditional distribution of Y given X depends only on the projection $\langle X, \theta \rangle$. Specifically,

$$f(y \mid X = x) = f(y \mid \langle x, \theta \rangle) =: f(\theta, y, x).$$

The index θ is identifiable up to a multiplicative constant; we impose the normalization $\|\theta\| = 1$.

Define the semi-metric associated with the single index θ as

$$d_\theta(x_1, x_2) := |\langle x_1 - x_2, \theta \rangle|, \quad \forall x_1, x_2 \in \mathcal{H}.$$

Let $\mathcal{F}_i$ be the σ-fields generated by $((\langle X_1, \theta \rangle, Y_1), \ldots, (\langle X_i, \theta \rangle, Y_i))$. Define the ball

$$B_\theta(x, h) = \{\chi \in \mathcal{H} : |\langle x - \chi, \theta \rangle| \leq h\}.$$

The small ball probability function is

$$\psi_{\theta,x}(u) = \mathbb{P}(d_\theta(x, X_i) \leq u) = \mathbb{P}(X_i \in B_\theta(x, u)).$$

The conditional version given past information is

$$\psi_{\theta,x}^{\mathcal{F}_{i-1}}(u) = \mathbb{P}\left(d_\theta(x, X_i) \leq u \mid \mathcal{F}_{i-1}\right).$$

The kernel estimator under complete data is

$$\widehat{f}_n(\theta, y, x) = \frac{b_H^{-1} \sum_{i=1}^n K\left(b_K^{-1} d_\theta(x, X_i)\right) H\left(b_H^{-1}(y - Y_i)\right)}{\sum_{i=1}^n K\left(b_K^{-1} d_\theta(x, X_i)\right)},$$

where $K : \mathbb{R}^+ \to \mathbb{R}^+$ is a kernel function for the functional covariate, $H : \mathbb{R} \to \mathbb{R}^+$ is a kernel function for the response, and b_K, b_H are bandwidth parameters satisfying $b_K \to 0$, $b_H \to 0$ as $n \to \infty$.

This estimator can be expressed as a ratio:

$$\widehat{f}_n(\theta, y, x) = \frac{\widehat{f}_N(\theta, y, x)}{\widehat{f}_D(\theta, x)}$$

with

$$\widehat{f}_N(\theta, y, x) = \frac{1}{nb_H} \sum_{i=1}^n K_i(\theta, x) H_i(\theta, y, x), \quad \widehat{f}_D(\theta, x) = \frac{1}{n} \sum_{i=1}^n K_i(\theta, x),$$

where $K_i(\theta, x) = K(b_K^{-1} d_\theta(x, X_i))$ and $H_i(\theta, y, x) = H(b_H^{-1}(y - Y_i))$.

2.1.2 Assumptions

Assumption [A1] K is a nonnegative bounded kernel function over its support $[0, 1]$ with $K(1) > 0$, and the derivative K' exists on $[0, 1]$ with $K'(t) < 0$ for all $t \in [0, 1]$. Moreover, $\int_0^1 (K^j)'(t)\, dt < \infty$ for $j = 1, 2$.

Comment: This assumption is classical in functional estimation for finite or infinite dimension spaces. It ensures the kernel used in the smoothing procedure has appropriate properties (boundedness, monotonicity, and integrability of its derivatives) to guarantee the consistency and asymptotic normality of the kernel-based estimators. The decreasing nature of K gives more weight to observations near the point of interest, which is essential for local smoothing methods.

Assumption [A2]

(i) H is a positive bounded function with

$$\int_{\mathbb{R}} |t|^{b_2} H(t)\, dt < \infty \quad \text{and} \quad \int_{\mathbb{R}} tH(t)\, dt = 0,$$

and for all $(t_1, t_2) \in \mathbb{R}^2$, $|H(t_1) - H(t_2)| \leq C|t_1 - t_2|$.

(ii) $H^{(1)}(t)$ and $H^{(2)}(t)$ are bounded with $\int (H^{(1)}(t))^2\, dt < \infty$.

Comment: This assumption concerns the second kernel function H used in the estimation procedure (likely for the conditional density or regression function). The conditions ensure that H is a valid kernel of order 1 (zero first moment), Lipschitz continuous, and has bounded derivatives. These properties are crucial for controlling the bias and variance terms in the asymptotic expansion, particularly when dealing with the response variable Y. The finite moment condition of order b_2 aligns with the Hölder regularity of the conditional density in Assumption A4.

Assumption [A3] For $x \in \mathcal{H}$:

(i) There exist a sequence of nonnegative bounded random functions $(f_{i,1})_{i\geq 1}$, a sequence of random functions $(g_{i,\theta,x})_{i\geq 1}$, a deterministic nonnegative bounded function f_1 and a nonnegative real function $\phi_\theta(\cdot)$ tending to zero as its argument tends to zero, such that

$$\psi_{\theta,x}(b) = \phi_\theta(b) f_1(\theta, x) + o(\phi_\theta(b)) \quad \text{as } b \to 0.$$

(ii) For any $i \in \mathbb{N}$,

$$\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b) = \phi_\theta(b) f_{i,1}(\theta, x) + g_{i,\theta,x}(b)$$

with $g_{i,\theta,x} = o_{a.s.}(\phi(t))$ as $b \to 0$, $\frac{g_{i,\theta,x}(b)}{\phi_\theta(b)}$ a.s. bounded, and

$$n^{-1} \sum_{i=1}^{n} g_{i,\theta,x}^{j}(b) = o_{a.s.}(\phi_\theta^{j}(b)) \quad \text{as } n \to \infty \text{ for } j = 1, 2.$$

(iii) $n^{-1} \sum_{i=1}^{n} f_{i,1}^{j}(\theta, x) \to f_1^{j}(\theta, x)$ almost surely as $n \to \infty$ for $j = 1, 2$.

(iv) There exists a nondecreasing bounded function τ_0 such that, uniformly in $t \in [0, 1]$,

$$\phi_\theta(ht)/\phi_\theta(h) = \tau_0(t) + o(1), \quad \text{as } h \downarrow 0,$$

and $\int_0^1 (K^j)' \tau_0(t)\, dt < \infty$ for $j \geq 1$.

(v) $n^{-1} \sum_{i=1}^{n} b_i(\theta, x) \to D(\theta, x) < \infty$ as $n \to \infty$.

Comment: This is the core structural assumption for the functional time series or functional data setting. It establishes the asymptotic expansion of the conditional moments $\psi_{\theta,x}(b)$ around zero, which is essential for deriving the asymptotic bias and variance of the estimator. The decomposition into a dominant term $\phi_\theta(b) f_1(\theta, x)$ and a negligible remainder $o(\phi_\theta(b))$ allows us to isolate the leading behavior. The conditions on the random sequences $(f_{i,1})$ and $(g_{i,\theta,x})$ ensure that the expansion holds uniformly and that the averaging over observations converges appropriately. This assumption is particularly important because it handles the infinite-dimensional nature of the functional covariate X through the projection $\langle \theta, X_i \rangle$.

Assumption [A4] The conditional density $f(\theta, y, x)$ satisfies the Hölder condition: for all $(x_1, x_2) \in N_x \times N_x$, $(y_1, y_2) \in \mathcal{S}_{\mathbb{R}}^2$, and $b_1 > 0$, $b_2 > 0$,

$$|f(\theta, y_1, x_1) - f(\theta, y_2, x_2)| \leq C_{\theta,x}\big(\|x_1 - x_2\|^{b_1} + |y_1 - y_2|^{b_2}\big).$$

Comment: This assumption presents a regularity condition which characterizes the functional space of the model and is needed to evaluate the bias term of the asymptotic results. The Hölder continuity ensures that the conditional density does not vary too rapidly, allowing for precise control of the bias when approximating the density locally. The exponents b_1 and b_2 govern the smoothness in the functional and scalar directions, respectively, and directly influence the rate of convergence of the estimator.

Assumption [A5] For $j = 0, 1, 2$, and any $k \geq 1$,

$$\mathbb{E}\big[(b_H^{-1} H^{(j)}(b_H^{-1}(t - Y_i)))^k \mid \mathcal{G}_{i-1}\big] = \mathbb{E}\big[(b_H^{-1} H^{(j)}(b_H^{-1}(t - Y_i)))^k \mid \langle \theta, X_i \rangle\big],$$

where $\mathcal{G}_i$ are the σ-fields generated by $((\langle X_1, \theta \rangle, Y_1), \ldots, (\langle X_i, \theta \rangle, Y_i), \langle X_{i+1}, \theta \rangle)$.

Comment: This assumption is of Markov's nature. It is employed to evaluate the conditional bias. Essentially, it states that the conditional expectation of the kernel weights given the past depends only on the current projected covariate $\langle \theta, X_i \rangle$, not on the entire history. This Markov-type property simplifies the computation of conditional moments and is crucial for handling the dependence structure in the time series setting, allowing us to apply martingale difference arguments.

Assumption [A6] The bandwidths satisfy:

1. $b_K \to 0$ and $b_H \to 0$ as $n \to \infty$,
2. $\exists \xi > 0$ such that $n^{\xi} b_H^2 \to \infty$,
3. $\frac{\log n}{n b_H^2 \phi_\theta(b_K)} \to 0$ as $n \to \infty$.

Comment: These are standard technical conditions on the bandwidth sequences b_K (for the kernel K) and b_H (for the kernel H). Condition (i) ensures that the smoothing windows shrink to zero as the sample size increases, which is necessary for consistency. Condition (ii) guarantees that the effective sample size for the H-kernel grows fast enough to control the variance of the estimator. Condition (iii) is a crucial rate condition that ensures the uniform convergence of the estimator over the support, balancing the trade-off between the bias (controlled by the bandwidths) and the variance (controlled by the effective sample size). This condition is typical in nonparametric functional estimation to achieve strong consistency.

2.1.3 Decomposition of the Estimator

To study the asymptotic behavior of $\widehat{f}_n(\theta, y, x)$, we introduce the following decomposition:

$$\widehat{f}_n(\theta, y, x) - f(\theta, y, x) = \frac{Q_n(\theta, y, x) + R_n(\theta, y, x)}{\widehat{f}_D(\theta, x)} + B_n(\theta, y, x),$$

where

$$Q_n(\theta, y, x) = \big(\widehat{f}_N(\theta, y, x) - \bar{f}_N(\theta, y, x)\big) - f(\theta, y, x)\big(\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x)\big),$$

$$R_n(\theta, y, x) = -B_n(\theta, y, x)\big(\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x)\big),$$

$$B_n(\theta, y, x) = \frac{\bar{f}_N(\theta, y, x)}{\bar{f}_D(\theta, x)} - f(\theta, y, x),$$

with

$$\bar{f}_N(\theta, y, x) = \frac{1}{n b_H} \sum_{i=1}^{n} \mathbb{E}[K_i(\theta, x) H_i(\theta, y, x) \mid \mathcal{F}_{i-1}],$$

$$\bar{f}_D(\theta, x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}[K_i(\theta, x) \mid \mathcal{F}_{i-1}].$$

2.1.4 Technical Lemmas

Lemma 2.1 (Convergence of the Denominator)**.** *Under assumptions (A1) and (A3),*

$$\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x) = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n \phi_\theta(b_K)}}\right) = o_{a.s.}(1)$$

and

$$\lim_{n\to\infty} \bar{f}_D(\theta, x) = f_1(\theta, x) M_1 \quad a.s.,$$

where $M_1 = K(1) - \int_0^1 K'(t)\tau_0(t)\,dt$*.*

Proof. Define the martingale differences

$$M_{ni} = K_i(\theta, x) - \mathbb{E}[K_i(\theta, x) \mid \mathcal{F}_{i-1}].$$

By Lemma A.1 (see Appendix), $|M_{ni}| \le C$ almost surely. The conditional variance satisfies

$$\mathbb{E}[M_{ni}^2 \mid \mathcal{F}_{i-1}] \le C\phi_\theta(b_K) f_{i,1}(\theta, x) + o_{a.s.}(\phi_\theta(b_K)).$$

Summing over i and using (A3)(iii) gives

$$\sum_{i=1}^{n} \mathbb{E}[M_{ni}^2 \mid \mathcal{F}_{i-1}] = \mathcal{O}_{a.s.}(n\phi_\theta(b_K)).$$

Applying the exponential inequality for martingale differences (Lemma A.2), for any $\epsilon > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} M_{ni}\right| > \epsilon\sqrt{n\phi_\theta(b_K)\log n}\right) \le 2\exp\left(-\frac{\epsilon^2 \log n}{C}\right).$$

Choosing ϵ sufficiently large yields summability by the Borel–Cantelli lemma, establishing the rate. For the limit,

$$\bar{f}_D(\theta, x) = \frac{1}{n}\sum_{i=1}^{n} \phi_\theta(b_K) f_{i,1}(\theta, x) M_1 + o_{a.s.}(\phi_\theta(b_K)).$$

The result follows from (A3)(iii) and $\phi_\theta(b_K) \to 0$. □

Lemma 2.2 (Bias Term)**.** *Under assumptions (A2), (A3), and (A4),*

$$\sup_{y\in\mathcal{S}_\mathbb{R}} |B_n(\theta, y, x)| = \mathcal{O}(b_K^{b_1} + b_H^{b_2}).$$

Proof. Write $B_n(\theta, y, x) = \frac{\bar{f}_N(\theta, y, x)}{\bar{f}_D(\theta, x)} - f(\theta, y, x)$. Using the conditional expectation,

$$\bar{f}_N(\theta, y, x) = \frac{1}{nb_H}\sum_{i=1}^{n} \int_{\mathcal{H}} \int_{\mathbb{R}} K\left(\frac{d_\theta(x, \chi)}{b_K}\right) H\left(\frac{y - t}{b_H}\right) f(\theta, t, \chi)\,dt\,d\psi_{\theta,x}^{\mathcal{F}_{i-1}}(\chi).$$

By the change of variables $u = d_\theta(x, \chi)/b_K$ and $v = (y - t)/b_H$, and the Hölder condition (A4),

$$\bar{f}_N(\theta, y, x) = \frac{1}{n}\sum_{i=1}^{n} \phi_\theta(b_K) f_{i,1}(\theta, x) \int_0^1 K(u)\tau_0'(u)\,du\big(f(\theta, y, x) + \mathcal{O}(b_K^{b_1} + b_H^{b_2})\big).$$

Similarly, $\bar{f}_D(\theta, x) = \frac{1}{n}\sum_{i=1}^{n} \phi_\theta(b_K) f_{i,1}(\theta, x) M_1 + o_{a.s.}(\phi_\theta(b_K))$. Taking the ratio yields the stated bias bound. □

Lemma 2.3 (Variance Term)**.** *Under assumptions (A1)–(A5),*

$$\sup_{y\in\mathcal{S}_\mathbb{R}} |\widehat{f}_N(\theta,y,x) - \bar{f}_N(\theta,y,x)| = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{nb_H^2\phi_\theta(b_K)}}\right).$$

Proof. Consider a discretization of the compact set $\mathcal{S}_\mathbb{R}$ into points $y_1,\dots,y_{m_n}$ with mesh $\delta_n = (nb_H^2\phi_\theta(b_K)/\log n)^{-1/2}$. For any y, there exists y_k such that $|y-y_k| \le \delta_n$. Using the Lipschitz property of H,

$$|\widehat{f}_N(\theta,y,x) - \widehat{f}_N(\theta,y_k,x)| \le C\delta_n b_H^{-1}\widehat{f}_D(\theta,x).$$

Define for each fixed y,

$$L_{ni}(y) = \frac{1}{b_H}\left[K_i(\theta,x)H_i(\theta,y,x) - \mathbb{E}[K_i(\theta,x)H_i(\theta,y,x) \mid \mathcal{F}_{i-1}]\right].$$

These form martingale differences. The conditional variance satisfies

$$\mathbb{E}[L_{ni}^2(y) \mid \mathcal{F}_{i-1}] \le Cb_H^{-2}\phi_\theta(b_K)f_{i,1}(\theta,x)\int H^2(u)\,du + o_{a.s.}(b_H^{-2}\phi_\theta(b_K)).$$

Applying the exponential inequality for each y_k and using the union bound with $m_n \sim \sqrt{nb_H^2\phi_\theta(b_K)/\log n}$ yields the result. □

2.1.5 Results for Complete Data

Theorem 2.4 (Uniform Almost Complete Convergence)**.** *Under assumptions (A1)–(A6),*

$$\sup_{y\in\mathcal{S}_\mathbb{R}} |\widehat{f}_n(\theta,y,x) - f(\theta,y,x)| = \mathcal{O}_{a.s.}(b_K^{b_1} + b_H^{b_2}) + \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{nb_H^2\phi_\theta(b_K)}}\right).$$

Proof. Using the decomposition and Lemmas 2.1, 2.2, and 2.3,

$$\sup_y |\widehat{f}_n(\theta,y,x) - f(\theta,y,x)| \le \frac{\sup_y|Q_n(\theta,y,x)| + \sup_y|R_n(\theta,y,x)|}{|\widehat{f}_D(\theta,x)|} + \sup_y|B_n(\theta,y,x)|.$$

From Lemma 2.1, $\widehat{f}_D(\theta,x)$ is bounded away from zero almost surely for sufficiently large n. Lemma 2.2 gives the bias rate. For Q_n,

$$|Q_n(\theta,y,x)| \le |\widehat{f}_N - \bar{f}_N| + f(\theta,y,x)|\widehat{F}_D - \bar{F}_D|.$$

Lemma 2.3 controls the first term, while Lemma 2.1 controls the second. The term R_n is of smaller order. This completes the proof. □

Theorem 2.5 (Asymptotic Normality)**.** *Under assumptions (A1)–(A6) and if additionally*

$$\sqrt{nb_H\phi_\theta(b_K)}(b_K^{b_1} + b_H^{b_2}) \to 0 \quad \text{as } n\to\infty,$$

then

$$\sqrt{nb_H\phi_\theta(b_K)}(\widehat{f}_n(\theta,y,x) - f(\theta,y,x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2(\theta,y,x)),$$

where

$$\sigma^2(\theta,y,x) = \frac{M_2}{M_1^2}\frac{f(\theta,y,x)}{f_1(\theta,x)}\int_\mathbb{R} H^2(u)\,du,$$

with $M_j = K^j(1) - \int_0^1 (K^j)'(u)\tau_0(u)\,du$ *for* $j = 1,2$.

Proof. From the decomposition and the additional condition, the bias term B_n and the remainder term R_n are asymptotically negligible. The main contribution comes from Q_n, which can be written as

$$\sqrt{nb_H\phi_\theta(b_K)}Q_n(\theta, y, x) = \sum_{i=1}^{n} \xi_{ni},$$

where

$$\xi_{ni} = \left(\frac{\phi_\theta(b_K)}{nb_H}\right)^{1/2} \Big[K_i(\theta, x)H_i(\theta, y, x) - \mathbb{E}[K_i(\theta, x)H_i(\theta, y, x) \mid \mathcal{F}_{i-1}] \\ - f(\theta, y, x)\big(K_i(\theta, x) - \mathbb{E}[K_i(\theta, x) \mid \mathcal{F}_{i-1}]\big)\Big].$$

The sequence $\{\xi_{ni}, \mathcal{F}_{i-1}\}$ forms a martingale difference array. The conditional variance converges a.s. to $\sigma^2(\theta, y, x)$ by direct computation using the small-ball probabilities and kernel properties. The Lindeberg condition is verified using moment bounds. Applying the martingale central limit theorem see [18] gives the result. □

2.2 Under Missing Data at Random (MAR) – Conditional Density

2.2.1 Model and Estimator

We observe the triplets $(X_i, Y_i, \delta_i)_{1\le i\le n}$, where $\delta_i = 1$ if Y_i is observed. The missing mechanism is assumed to be Missing At Random (MAR):

Assumption [A7 (MAR)]

$$\mathbb{P}(\delta = 1 \mid \langle X, \theta\rangle = \langle x, \theta\rangle, Y = y) = \mathbb{P}(\delta = 1 \mid \langle X, \theta\rangle = \langle x, \theta\rangle) = p(\theta, x),$$

with $0 < p_{\min} \le p(\theta, x) \le 1$.

Comment: This assumption formalizes the missing at random (MAR) mechanism, which is crucial for handling incomplete data in our functional setting. It ensures that the probability of observing the response Y depends only on the observed projected covariate $\langle X, \theta\rangle$, not on the missing value itself. This allows us to apply inverse probability weighting or other missing data techniques consistently. The lower bound $p_{\min} > 0$ ensures that all regions of the covariate space have a non-negligible chance of being observed, which is necessary for uniform convergence results.

The weighted kernel estimator of the conditional density under MAR is

$$\widehat{f}_n^{\mathrm{MAR}}(\theta, y, x) = \frac{\sum_{i=1}^{n} \delta_i K(b_K^{-1} d_\theta(x, X_i))H(b_H^{-1}(y - Y_i))}{b_H \sum_{i=1}^{n} \delta_i K(b_K^{-1} d_\theta(x, X_i))} = \frac{\widehat{f}_N^{\mathrm{MAR}}(\theta, y, x)}{\widehat{f}_D^{\mathrm{MAR}}(\theta, x)}.$$

with

$$\widehat{f}_N^{\mathrm{MAR}}(\theta, y, x) = \frac{1}{nb_H} \sum_{i=1}^{n} \delta_i K_i(\theta, x)H_i(\theta, y, x), \quad \widehat{f}_D^{\mathrm{MAR}}(\theta, x) = \frac{1}{n} \sum_{i=1}^{n} \delta_i K_i(\theta, x).$$

2.2.2 Technical Lemmas under MAR

Lemma 2.6 (Convergence of the Denominator – MAR Density)**.** *Under assumptions (A1), (A3) and (A7),*

$$\widehat{f}_D^{MAR}(\theta, x) - \bar{f}_D^{MAR}(\theta, x) = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right) = o_{a.s.}(1),$$

and

$$\lim_{n\to\infty} \bar{f}_D^{MAR}(\theta, x) = p(\theta, x) \quad a.s.$$

Proof. Define the martingale differences

$$M_{ni} = \frac{\delta_i K_i(\theta, x)}{\mathbb{E}[K_1(\theta, x)]} - \mathbb{E}\left[\frac{\delta_i K_i(\theta, x)}{\mathbb{E}[K_1(\theta, x)]} \mid \mathcal{F}_{i-1}\right].$$

By (A1) and (A3), there exists $C_0 > 0$ such that $|M_{ni}| \le C_0$ a.s. Under the MAR assumption (A7) and continuity of $p(\theta, \cdot)$,

$$\mathbb{E}[\delta_i \mid \mathcal{F}_{i-1}] = p(\theta, X_i) = p(\theta, x) + o(1).$$

Using the small-ball probability properties,

$$\mathbb{E}[(\delta_i K_i(\theta, x))^2 \mid \mathcal{F}_{i-1}] = p(\theta, x)\phi_\theta(b_K) f_{i,1}(\theta, x) M_2 + o_{a.s.}(\phi_\theta(b_K)).$$

Hence,

$$\mathbb{E}[M_{ni}^2 \mid \mathcal{F}_{i-1}] = \frac{p(\theta, x)M_2}{M_1^2 f_1^2(\theta, x)} \frac{f_{i,1}(\theta, x)}{\phi_\theta(b_K)} + o_{a.s.}(\phi_\theta(b_K)^{-1}).$$

By (A3)(iii), $n^{-1}\sum f_{i,1}(\theta, x) \to f_1(\theta, x)$ a.s., so there exists $C_1 > 0$ such that

$$\sum_{i=1}^n \mathbb{E}[M_{ni}^2 \mid \mathcal{F}_{i-1}] \le C_1 \frac{n}{\phi_\theta(b_K)} \quad a.s.$$

Let $D_n = C_1 n/\phi_\theta(b_K)$. The exponential inequality for martingale differences with a suitable ϵ gives

$$\mathbb{P}\Big(\Big|\sum_{i=1}^n M_{ni}\Big| > \epsilon\sqrt{D_n \log n}\Big) \le 2n^{-2}.$$

The Borel–Cantelli lemma yields the stated rate. The limit of $\bar{F}_D^{\text{MAR}}$ follows by taking conditional expectations and using (A3)(iii) together with $\mathbb{E}[K_1(\theta, x)] \sim \phi_\theta(b_K) f_1(\theta, x) M_1$. □

Lemma 2.7 (Bias Term – MAR Density)**.** *Under assumptions (A2), (A3), (A4) and (A7),*

$$\sup_{y\in\mathcal{S}_\mathbb{R}} |B_n^{MAR}(\theta, y, x)| = \mathcal{O}(b_K^{b_1} + b_H^{b_2}).$$

Proof. By the MAR assumption,

$$\mathbb{E}[\delta_i K_i(\theta, x)H_i(\theta, y, x) \mid \mathcal{F}_{i-1}] = p(\theta, x)\mathbb{E}[K_i(\theta, x)H_i(\theta, y, x) \mid \mathcal{F}_{i-1}].$$

The conditional expectation inside is the same as in the complete-data case. Therefore,

$$\bar{f}_N^{\text{MAR}}(\theta, y, x) = p(\theta, x) \cdot \big(\text{complete-data expression}\big) + o_{a.s.}(1).$$

Applying the same change-of-variable and Hölder arguments as in Lemma 2.2 yields

$$\bar{f}_N^{\text{MAR}}(\theta, y, x) = p(\theta, x)\phi_\theta(b_K) f_{i,1}(\theta, x) M_1\big(f(\theta, y, x) + \mathcal{O}(b_K^{b_1} + b_H^{b_2})\big) + o_{a.s.}(1).$$

Since $\bar{f}_D^{\text{MAR}}(\theta, x) \to p(\theta, x) > 0$ a.s. (Lemma 2.6), the ratio B_n^{MAR} satisfies the stated bound uniformly in y. □

Lemma 2.8 (Variance Term – MAR Density)**.** *Under assumptions (A1)–(A5) and (A7),*

$$\sup_{y\in\mathcal{S}_{\mathbb{R}}}\left|\widehat{f}_N^{MAR}(\theta,y,x)-\bar{f}_N^{MAR}(\theta,y,x)\right|=\mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{nb_H^2\phi_\theta(b_K)}}\right).$$

Proof. Define the martingale differences

$$L_{ni}(y)=\frac{\delta_i K_i(\theta,x)}{\mathbb{E}[K_1(\theta,x)]}\Big(H(b_H^{-1}(y-Y_i))-\mathbb{E}\big[H(b_H^{-1}(y-Y_i))\mid\langle\theta,X_i\rangle\big]\Big)-\mathbb{E}\Big[\cdots\;\Big|\;\mathcal{F}_{i-1}\Big].$$

By (A1) and (A2), $|L_{ni}(y)|\le C$ a.s. The conditional variance is bounded by

$$\mathbb{E}[L_{ni}(y)^2\mid\mathcal{F}_{i-1}]\le C\frac{p(\theta,x)M_2}{M_1^2f_1^2(\theta,x)}\frac{f_{i,1}(\theta,x)}{\phi_\theta(b_K)}b_H^{-2}\int H^2(u)\,du+o_{a.s.}(b_H^{-2}\phi_\theta(b_K)^{-1}).$$

Thus $\sum\mathbb{E}[L_{ni}(y)^2\mid\mathcal{F}_{i-1}]\le C_1n/(b_H^2\phi_\theta(b_K))$ a.s. Discretize $\mathcal{S}_{\mathbb{R}}$ into $m_n\sim\sqrt{nb_H^2\phi_\theta(b_K)/\log n}$ points. Using the Lipschitz property of H and the exponential inequality with union bound, the Borel–Cantelli lemma yields the uniform rate. □

2.2.3 Asymptotic Results Under MAR

Theorem 2.9 (Uniform Almost Complete Convergence – MAR Density)**.** *Under assumptions (A1)–(A6) and (A7),*

$$\sup_{y\in\mathcal{S}_{\mathbb{R}}}\left|\widehat{f}_n^{MAR}(\theta,y,x)-f(\theta,y,x)\right|=\mathcal{O}_{a.s.}(b_K^{b_1}+b_H^{b_2})+\mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{nb_H^2\phi_\theta(b_K)}}\right).$$

Theorem 2.10 (Asymptotic Normality – MAR Density)**.** *Under assumptions (A1)–(A6), (A7) and*

$$\sqrt{nb_H\phi_\theta(b_K)}(b_K^{b_1}+b_H^{b_2})\to 0\quad \textit{as } n\to\infty,$$

then

$$\sqrt{nb_H\phi_\theta(b_K)}\big(\widehat{f}_n^{MAR}(\theta,y,x)-f(\theta,y,x)\big)\xrightarrow{\mathcal{D}}\mathcal{N}\big(0,\sigma^2_{MAR}(\theta,y,x)\big),$$

where

$$\sigma^2_{MAR}(\theta,y,x)=\frac{M_2}{M_1^2}\frac{f(\theta,y,x)}{p(\theta,x)f_1(\theta,x)}\int_{\mathbb{R}}H^2(u)\,du.$$

2.2.4 Proof of results

Proof of Theorem 2.9. We use the same decomposition as in the complete data case:

$$\widehat{f}_n^{\mathrm{MAR}}(\theta,y,x)-f(\theta,y,x)=\frac{Q_n^{\mathrm{MAR}}(\theta,y,x)+R_n^{\mathrm{MAR}}(\theta,y,x)}{\widehat{f}_D^{\mathrm{MAR}}(\theta,x)}+B_n^{\mathrm{MAR}}(\theta,y,x),$$

where

$$Q_n^{\mathrm{MAR}}(\theta,y,x)=\big(\widehat{f}_N^{\mathrm{MAR}}-\bar{f}_N^{\mathrm{MAR}}\big)-f(\theta,y,x)\big(\widehat{F}_D^{\mathrm{MAR}}-\bar{F}_D^{\mathrm{MAR}}\big),$$
$$R_n^{\mathrm{MAR}}(\theta,y,x)=-B_n^{\mathrm{MAR}}(\theta,y,x)\big(\widehat{F}_D^{\mathrm{MAR}}-\bar{F}_D^{\mathrm{MAR}}\big),$$
$$B_n^{\mathrm{MAR}}(\theta,y,x)=\frac{\bar{f}_N^{\mathrm{MAR}}(\theta,y,x)}{\bar{f}_D^{\mathrm{MAR}}(\theta,x)}-f(\theta,y,x).$$

Taking the supremum over $y \in \mathcal{S}_{\mathbb{R}}$, we have

$$\sup_y \left| \widehat{f}_n^{\mathrm{MAR}} - f \right| \leq \frac{\sup_y |Q_n^{\mathrm{MAR}}| + \sup_y |R_n^{\mathrm{MAR}}|}{|\widehat{F}_D^{\mathrm{MAR}}|} + \sup_y |B_n^{\mathrm{MAR}}|.$$

By Lemma 2.6, $\widehat{f}_D^{\mathrm{MAR}}(\theta, x) \to p(\theta, x) > 0$ almost surely as $n \to \infty$. Hence, there exists a random variable $c > 0$ and N such that for all $n \geq N$, $|\widehat{f}_D^{\mathrm{MAR}}(\theta, x)| \geq c$ a.s.

From Lemma 2.7,

$$\sup_y |B_n^{\mathrm{MAR}}(\theta, y, x)| = \mathcal{O}_{a.s.}(b_K^{b_1} + b_H^{b_2}).$$

For the stochastic term Q_n^{MAR}, we decompose it as

$$|Q_n^{\mathrm{MAR}}(\theta, y, x)| \leq |\widehat{f}_N^{\mathrm{MAR}} - \bar{f}_N^{\mathrm{MAR}}| + f(\theta, y, x) \cdot |\widehat{F}_D^{\mathrm{MAR}} - \bar{F}_D^{\mathrm{MAR}}|.$$

Since $0 \leq f(\theta, y, x) \leq \|f\|_\infty < \infty$, taking the supremum over y yields

$$\sup_y |Q_n^{\mathrm{MAR}}| \leq \sup_y |\widehat{f}_N^{\mathrm{MAR}} - \bar{f}_N^{\mathrm{MAR}}| + \| f \|_\infty \cdot |\widehat{F}_D^{\mathrm{MAR}} - \bar{F}_D^{\mathrm{MAR}}|.$$

Applying Lemma 2.8 and Lemma 2.6, we obtain

$$\sup_y |Q_n^{\mathrm{MAR}}| = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n b_H^2 \phi_\theta(b_K)}} \right).$$

For the remainder term,

$$\sup_y |R_n^{\mathrm{MAR}}| \leq \sup_y |B_n^{\mathrm{MAR}}| \cdot |\widehat{F}_D^{\mathrm{MAR}} - \bar{F}_D^{\mathrm{MAR}}| = \mathcal{O}_{a.s.}(b_K^{b_1} + b_H^{b_2}) \cdot \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n b_H^2 \phi_\theta(b_K)}} \right),$$

which is of lower order than the main stochastic term. Combining all bounds and using that the denominator is bounded away from zero almost surely, we conclude

$$\sup_{y \in \mathcal{S}_{\mathbb{R}}} \left| \widehat{f}_n^{\mathrm{MAR}}(\theta, y, x) - f(\theta, y, x) \right| = \mathcal{O}_{a.s.}(b_K^{b_1} + b_H^{b_2}) + \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n b_H^2 \phi_\theta(b_K)}} \right).$$

This completes the proof. □

Proof of Theorem 2.10. From the decomposition and the additional bandwidth condition $\sqrt{n b_H \phi_\theta(b_K)}(b_K^{b_1} + b_H^{b_2}) \to 0$, the bias term B_n^{MAR} and the remainder R_n^{MAR} satisfy

$$\sqrt{n b_H \phi_\theta(b_K)} B_n^{\mathrm{MAR}} = o_{a.s.}(1), \quad \sqrt{n b_H \phi_\theta(b_K)} \frac{R_n^{\mathrm{MAR}}}{\widehat{F}_D^{\mathrm{MAR}}} = o_{a.s.}(1).$$

Thus,

$$\sqrt{n b_H \phi_\theta(b_K)} \big(\widehat{f}_n^{\mathrm{MAR}} - f \big) = \frac{\sqrt{n b_H \phi_\theta(b_K)} Q_n^{\mathrm{MAR}}(\theta, y, x)}{\widehat{f}_D^{\mathrm{MAR}}(\theta, x)} + o_{a.s.}(1).$$

By Lemma 2.6, $\widehat{f}_D^{\mathrm{MAR}}(\theta, x) \to p(\theta, x)$ almost surely. Therefore, it is sufficient to show the asymptotic normality of $\sqrt{n b_H \phi_\theta(b_K)} Q_n^{\mathrm{MAR}}$.

Define the martingale difference array

$$\xi_{ni} = \left(\frac{\phi_\theta(b_K)}{nb_H}\right)^{1/2} \Big[\delta_i K_i(\theta,x)H_i(\theta,y,x) - \mathbb{E}[\delta_i K_i H_i \mid \mathcal{F}_{i-1}] - f(\theta,y,x)\big(\delta_i K_i(\theta,x) - \mathbb{E}[\delta_i K_i \mid \mathcal{F}_{i-1}]\big)\Big].$$

Then

$$\sqrt{nb_H\phi_\theta(b_K)}Q_n^{\text{MAR}}(\theta,y,x) = \sum_{i=1}^{n} \xi_{ni} + o_{a.s.}(1).$$

The sequence $\{\xi_{ni}, \mathcal{F}_{i-1}\}$ is a martingale difference array. Using the MAR assumption (A7) and the small-ball probability properties (A3), the conditional variance satisfies

$$\sum_{i=1}^{n} \mathbb{E}[\xi_{ni}^2 \mid \mathcal{F}_{i-1}] \xrightarrow{a.s.} \frac{M_2}{M_1^2}\frac{f(\theta,y,x)}{p(\theta,x)f_1(\theta,x)} \int_{\mathbb{R}} H^2(u)\,du = \sigma^2_{\text{MAR}}(\theta,y,x).$$

The Lindeberg condition holds because

$$|\xi_{ni}| \le C\left(\frac{\phi_\theta(b_K)}{nb_H}\right)^{1/2} \cdot 2\sup_i \left|\frac{\delta_i K_i H_i}{\mathbb{E}[K_1]}\right| = o(1) \quad a.s.,$$

under the bandwidth assumptions. Therefore, by the martingale central limit theorem (Hall and Heyde, 1980),

$$\sum_{i=1}^{n} \xi_{ni} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2_{\text{MAR}}(\theta,y,x)).$$

Finally, since $\widehat{f}_D^{\text{MAR}}(\theta,x) \to p(\theta,x)$ a.s., Slutsky's theorem implies

$$\sqrt{nb_H\phi_\theta(b_K)}\big(\widehat{f}_n^{\text{MAR}}(\theta,y,x) - f(\theta,y,x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma^2_{\text{MAR}}(\theta,y,x)\big).$$

This completes the proof. □

2.3 Conditional Mode Estimation under MAR

The conditional mode is a robust alternative to the conditional mean, particularly useful when the conditional distribution is multimodal or heavy-tailed. In this section, we study the asymptotic properties of the conditional mode estimator derived from the kernel conditional density estimator under the missing-at-random (MAR) setting.

The natural estimator under MAR is

$$\widehat{M}_\theta^{\text{MAR}}(x) = \arg\sup_{y \in \mathcal{S}_{\mathbb{R}}} \widehat{f}_n^{\text{MAR}}(\theta,y,x).$$

Assumption [A11] (Mode regularity) There exists $\epsilon_0 > 0$ such that $f(\theta,\cdot,x)$ is strictly increasing on $(M_\theta(x) - \epsilon_0, M_\theta(x))$ and strictly decreasing on $(M_\theta(x), M_\theta(x) + \epsilon_0)$. Moreover, $f(\theta,\cdot,x)$ is twice continuously differentiable in a neighborhood of $M_\theta(x)$ with

$$f^{(1)}(\theta, M_\theta(x), x) = 0, \quad f^{(2)}(\theta, M_\theta(x), x) \neq 0.$$

Comment: This assumption guarantees that the conditional mode $M_\theta(x)$ is uniquely defined and well-behaved in a neighborhood. The monotonicity conditions ensure that the mode is a genuine maximum with no flat regions, while the differentiability conditions with $f^{(1)} = 0$

and $f^{(2)} \neq 0$ are standard regularity conditions that allow for quadratic approximation of the density near the mode. These properties are essential for deriving the asymptotic behavior of the mode estimator, particularly its rate of convergence and asymptotic normality. The non-zero second derivative ensures that the mode is identifiable and that the estimator will have the usual parametric convergence rate.

2.3.1 Rate of Convergence for the Conditional Mode under MAR

Theorem 2.11 (Rate of Convergence for Conditional Mode under MAR)**.** *Under assumptions (A1)–(A6), (A7) and (A11),*

$$|\widehat{M}_\theta^{MAR}(x) - M_\theta(x)| = \mathcal{O}_{a.s.}\left((b_K^{b_1} + b_H^{b_2})^{1/2} + \left(\frac{\log n}{nb_H^2\phi_\theta(b_K)}\right)^{1/4}\right).$$

Proof. Let $m = M_\theta(x)$ and $\widehat{m} = \widehat{M}_\theta^{\mathrm{MAR}}(x)$. By definition of the mode, $f^{(1)}(\theta, m, x) = 0$ and $f^{(2)}(\theta, m, x) \neq 0$.

Since $\widehat{m}$ maximizes the estimated density under MAR, we have $\widehat{f}_n^{\mathrm{MAR}(1)}(\theta, \widehat{m}, x) = 0$. Apply a second-order Taylor expansion of the true conditional density around m:

$$f(\theta, \widehat{m}, x) = f(\theta, m, x) + \frac{1}{2}f^{(2)}(\theta, m^*, x)(\widehat{m} - m)^2,$$

where m^* lies between $\widehat{m}$ and m.

On the other hand, by the uniform almost complete convergence of the MAR conditional density estimator (Theorem 2.9),

$$\left|\widehat{f}_n^{\mathrm{MAR}}(\theta, \widehat{m}, x) - f(\theta, \widehat{m}, x)\right| \leq \sup_{y \in \mathcal{S}_\mathbb{R}}\left|\widehat{f}_n^{\mathrm{MAR}}(\theta, y, x) - f(\theta, y, x)\right|.$$

Hence,

$$f(\theta, \widehat{m}, x) - f(\theta, m, x) \geq -2\sup_y\left|\widehat{f}_n^{\mathrm{MAR}} - f\right|.$$

Combining both expressions,

$$\frac{1}{2}|f^{(2)}(\theta, m^*, x)|\,|\widehat{m} - m|^2 \leq 2\sup_y\left|\widehat{f}_n^{\mathrm{MAR}} - f\right| + o_{a.s.}(1).$$

By Assumption (A11) and continuity of the second derivative, there exists $c > 0$ such that $|f^{(2)}(\theta, m^*, x)| \geq c$ for all sufficiently large n. Therefore,

$$|\widehat{m} - m|^2 \leq C\sup_y\left|\widehat{f}_n^{\mathrm{MAR}}(\theta, y, x) - f(\theta, y, x)\right|.$$

Substituting the uniform convergence rate under MAR,

$$\sup_y\left|\widehat{f}_n^{\mathrm{MAR}} - f\right| = \mathcal{O}_{a.s.}(b_K^{b_1} + b_H^{b_2}) + \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{nb_H^2\phi_\theta(b_K)}}\right),$$

we immediately obtain

$$|\widehat{M}_\theta^{\mathrm{MAR}}(x) - M_\theta(x)| = \mathcal{O}_{a.s.}\left((b_K^{b_1} + b_H^{b_2})^{1/2} + \left(\frac{\log n}{nb_H^2\phi_\theta(b_K)}\right)^{1/4}\right).$$

This completes the proof. □

2.3.2 Asymptotic Normality for the Conditional Mode under MAR

Theorem 2.12 (Asymptotic Normality for Conditional Mode under MAR)**.** *Under assumptions (A1)–(A6), (A7), (A11), and the additional bandwidth conditions*

$$\sqrt{nb_H\phi_\theta(b_K)}(b_K^{b_1}+b_H^{b_2}) \to 0, \quad nb_H^3\phi_\theta(b_K)\to 0 \quad \text{as } n\to\infty,$$

then

$$\sqrt{nb_H^3\phi_\theta(b_K)}\big(\widehat{M}_\theta^{MAR}(x)-M_\theta(x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0,\rho^2_{MAR}(\theta,x)\big),$$

where

$$\rho^2_{MAR}(\theta,x)=\frac{M_2}{M_1^2}\frac{f(\theta,M_\theta(x),x)}{p(\theta,x)f_1(\theta,x)[f^{(2)}(\theta,M_\theta(x),x)]^2}\int_{\mathbb{R}}(H^{(1)}(u))^2\,du.$$

Proof. Let $m=M_\theta(x)$ and $\widehat{m}=\widehat{M}_\theta^{\text{MAR}}(x)$. Since $\widehat{m}$ maximizes the estimated density, we have

$$\widehat{f}_n^{\text{MAR}(1)}(\theta,\widehat{m},x)=0.$$

Apply a first-order Taylor expansion of the first derivative around m:

$$0=\widehat{f}_n^{\text{MAR}(1)}(\theta,m,x)+\widehat{f}_n^{\text{MAR}(2)}(\theta,m^*,x)(\widehat{m}-m),$$

where m^* lies between $\widehat{m}$ and m. Rearranging gives

$$\widehat{m}-m=-\frac{\widehat{f}_n^{\text{MAR}(1)}(\theta,m,x)}{\widehat{f}_n^{\text{MAR}(2)}(\theta,m^*,x)}.$$

Multiplying by the normalizing factor yields

$$\sqrt{nb_H^3\phi_\theta(b_K)}(\widehat{m}-m)=-\frac{\sqrt{nb_H^3\phi_\theta(b_K)}\,\widehat{f}_n^{\text{MAR}(1)}(\theta,m,x)}{\widehat{f}_n^{\text{MAR}(2)}(\theta,m^*,x)}.$$

By the uniform consistency of $\widehat{f}_n^{\text{MAR}}$ and its derivatives (under the bandwidth conditions), we have

$$\widehat{f}_n^{\text{MAR}(2)}(\theta,m^*,x)\xrightarrow{\mathbb{P}} f^{(2)}(\theta,m,x)\neq 0.$$

For the numerator, observe that

$$\widehat{f}_n^{\text{MAR}(1)}(\theta,m,x)=\frac{b_H^{-2}\sum_{i=1}^n\delta_iK_i(\theta,x)H^{(1)}\big(b_H^{-1}(m-Y_i)\big)}{\sum_{i=1}^n\delta_iK_i(\theta,x)}.$$

This is a ratio of weighted kernel sums with the derivative kernel $H^{(1)}$ and an extra b_H^{-1} scaling. Applying the asymptotic normality result for the MAR conditional density estimator (Theorem 2.10), but with kernel $H^{(1)}$ instead of H, we obtain that

$$\sqrt{nb_H^3\phi_\theta(b_K)}\,\widehat{f}_n^{\text{MAR}(1)}(\theta,m,x)$$

is asymptotically normal with mean zero and variance

$$\frac{M_2}{M_1^2}\frac{f(\theta,m,x)}{p(\theta,x)f_1(\theta,x)}\int_{\mathbb{R}}(H^{(1)}(u))^2\,du.$$

By Slutsky's theorem (the denominator converges in probability to the non-zero limit $f^{(2)}(\theta,m,x)$), we conclude that

$$\sqrt{nb_H^3\phi_\theta(b_K)}\big(\widehat{M}_\theta^{\text{MAR}}(x)-M_\theta(x)\big)\xrightarrow{\mathcal{D}}\mathcal{N}\big(0,\rho^2_{\text{MAR}}(\theta,x)\big),$$

with the explicit variance given above.

This completes the proof. □

2.4 Conclusion

This chapter provides a complete asymptotic theory for kernel estimation of the conditional density and the conditional mode estimator in the functional single-index model under ergodic dependence. Both the completely observed case and the missing-at-random (MAR) case have been treated rigorously. The results include sharp rates of uniform almost complete convergence and explicit asymptotic normality with closed-form variance expressions. These findings lay the theoretical foundation for the study of the conditional distribution function and the conditional quantiles developed in the subsequent chapter.

Chapter 3

Kernel Estimation of Conditional Distribution Function in the Single-Index Model

This chapter extends the methodology developed in Chapter 2 to the estimation of the conditional distribution function in the functional single-index model under stationary ergodic dependence. We first establish the asymptotic properties under completely observed data, then extend the results to the case where the response is missing at random (MAR). Finally, we derive the asymptotic behaviour of the conditional quantile estimator under the MAR mechanism, as it is directly obtained from the conditional distribution function.

Unlike the conditional density, the distribution function does not require smoothing in the response direction, leading to simpler bias terms and faster rates of convergence. The MAR assumption is particularly relevant in functional data analysis, where missingness often depends on the observed functional covariate. This chapter generalizes previous results (Ferraty et al., 2013; Ling et al., 2015, 2016) to the single-index framework under ergodic dependence.

3.1 Under Complete Data

3.1.1 Model and Estimator

We observe the sample $(X_i, Y_i)_{i=1}^n$, where $X_i \in \mathcal{H}$ and $Y_i \in \mathbb{R}$. The conditional distribution function is defined as

$$F(\theta, y, x) = \mathbb{P}(Y \leq y \mid \langle X, \theta \rangle = \langle x, \theta \rangle) = \int_{-\infty}^{y} f(\theta, t, x)\, dt.$$

The kernel estimator under complete data is

$$\widehat{F}_n(\theta, y, x) = \frac{\sum_{i=1}^n K(b_K^{-1} d_\theta(x, X_i)) \mathbb{1}_{\{Y_i \leq y\}}}{\sum_{i=1}^n K(b_K^{-1} d_\theta(x, X_i))}.$$

This can be written as the ratio $\widehat{F}_n = \widehat{F}_N / \widehat{F}_D$, with obvious notation.

3.1.2 Additional Assumptions

In addition to assumptions (A1), (A3), and (A4) from Chapter 2, we assume:

Assumption [A8] The conditional distribution function $F(\theta, \cdot, x)$ is continuously differentiable with respect to y, and its derivative is the conditional density $f(\theta, \cdot, x)$ satisfying (A4).

Assumption [A9] The bandwidth b_K satisfies $b_K \to 0$ and $\frac{\log n}{n\phi_\theta(b_K)} \to 0$ as $n \to \infty$.

3.1.3 Results under Complete Data

Theorem 3.1 (Uniform Almost Complete Convergence – Complete Data). *Under assumptions (A1), (A3), (A4), (A8), (A9),*

$$\sup_{y \in \mathcal{S}_\mathbb{R}} |\widehat{F}_n(\theta, y, x) - F(\theta, y, x)| = \mathcal{O}_{a.s.}(b_K^{b_1}) + \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right).$$

Theorem 3.2 (Asymptotic Normality – Complete Data). *Under the conditions of the previous theorem and if additionally $\sqrt{n\phi_\theta(b_K)}b_K^{b_1} \to 0$ as $n \to \infty$,*

$$\sqrt{n\phi_\theta(b_K)}\big(\widehat{F}_n(\theta, y, x) - F(\theta, y, x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_F^2(\theta, y, x)),$$

where

$$\sigma_F^2(\theta, y, x) = \frac{M_2}{M_1^2} \frac{F(\theta, y, x)(1 - F(\theta, y, x))}{f_1(\theta, x)}.$$

(The proofs follow the same structure as the MAR case below, with $\delta_i \equiv 1$, and are therefore omitted here for brevity but can be obtained by setting $p(\theta, x) \equiv 1$.)

3.2 Under Missing at Random Data

3.2.1 Model and Assumptions for MAR Data

The missingness mechanism is assumed to be Missing At Random (MAR):

Assumption [MAR]

$$\mathbb{P}(\delta = 1 \mid \langle X, \theta \rangle = \langle x, \theta \rangle, Y = y) = \mathbb{P}(\delta = 1 \mid \langle X, \theta \rangle = \langle x, \theta \rangle) = p(\theta, x)$$

This assumption implies that conditional on the projected covariate, the probability of observing Y does not depend on the value of Y itself.

Assumption [A7] $p(\theta, x)$ is continuous in a neighborhood of x and bounded away from zero:

$$0 < p_{\min} \leq p(\theta, x) \leq 1$$

Conditional Distribution Function

The conditional distribution function is defined as:

$$F(\theta, y, x) = \mathbb{P}(Y \leq y \mid \langle X, \theta \rangle = \langle x, \theta \rangle) = \int_{-\infty}^{y} f(\theta, t, x) dt$$

Kernel Estimator for Conditional Distribution

For the MAR case, we define the weighted kernel estimator of the conditional distribution function as:

$$\widehat{F}_n(\theta, y, x) = \frac{\sum_{i=1}^n \delta_i K\left(b_K^{-1} d_\theta(x, X_i)\right) \mathbb{1}_{\{Y_i \leq y\}}}{\sum_{i=1}^n \delta_i K\left(b_K^{-1} d_\theta(x, X_i)\right)}$$

This estimator can be written as:

$$\widehat{F}_n(\theta, y, x) = \frac{\widehat{F}_N(\theta, y, x)}{\widehat{f}_D(\theta, x)}$$

where

$$\widehat{F}_N(\theta, y, x) = \frac{1}{n\mathbb{E}(K_1(\theta, x))} \sum_{i=1}^n \delta_i K_i(\theta, x) \mathbb{1}_{\{Y_i \leq y\}}$$

$$\widehat{f}_D(\theta, x) = \frac{1}{n\mathbb{E}(K_1(\theta, x))} \sum_{i=1}^n \delta_i K_i(\theta, x)$$

3.2.2 Additional Assumptions for Distribution Function Estimation

Assumption [A8] The conditional distribution function $F(\theta, \cdot, x)$ is continuously differentiable with respect to y, and its derivative is the conditional density $f(\theta, \cdot, x)$ satisfying (A4).

Comment: This assumption ensures the existence and regularity of the conditional density function, which is the primary object of interest in our functional mode estimation problem. The continuous differentiability of the conditional distribution function guarantees that the density $f(\theta, \cdot, x)$ is well-defined and smooth enough for our theoretical developments. By linking this assumption to the Hölder condition in (A4), we ensure that the density possesses sufficient smoothness properties to control the bias in the kernel-based estimation of the conditional mode. This regularity is fundamental for applying Taylor expansions and for deriving the asymptotic properties of the mode estimator.

Assumption [A9] The bandwidth b_K satisfies $b_K \to 0$ and $\frac{\log n}{n\phi_\theta(b_K)} \to 0$ as $n \to \infty$.

Comment: This assumption imposes the necessary conditions on the bandwidth sequence b_K used in the functional kernel K for estimating the conditional density in infinite-dimensional spaces. The condition $b_K \to 0$ ensures that the smoothing window shrinks as the sample size increases, which is essential for reducing the bias of the estimator. The second condition, $\frac{\log n}{n\phi_\theta(b_K)} \to 0$, is a standard rate condition in functional nonparametric estimation that ensures the uniform almost sure convergence of the estimator over the support. This condition is more stringent than in finite-dimensional settings due to the curse of dimensionality inherent in functional data analysis, and it balances the trade-off between the bias (controlled by b_K) and the variance (controlled by the effective sample size).

Note that unlike density estimation, no bandwidth is required in the response direction for distribution function estimation.

3.2.3 Decomposition for MAR Estimator

Similar to Chapter 2, we use the decomposition:

$$\widehat{F}_n(\theta, y, x) - F(\theta, y, x) = \frac{Q_n^F(\theta, y, x) + R_n^F(\theta, y, x)}{\widehat{f}_D(\theta, x)} + B_n^F(\theta, y, x)$$

where

$$Q_n^F(\theta, y, x) = \left(\widehat{F}_N(\theta, y, x) - \bar{F}_N(\theta, y, x)\right) - F(\theta, y, x)\left(\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x)\right)$$

$$R_n^F(\theta, y, x) = -B_n^F(\theta, y, x)\left(\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x)\right)$$

$$B_n^F(\theta, y, x) = \frac{\bar{F}_N(\theta, y, x)}{\bar{f}_D(\theta, x)} - F(\theta, y, x)$$

with

$$\bar{F}_N(\theta, y, x) = \frac{1}{n\mathbb{E}(K_1(\theta, x))}\sum_{i=1}^{n}\mathbb{E}[\delta_i K_i(\theta, x)\mathbb{1}_{\{Y_i \leq y\}} \mid \mathcal{F}_{i-1}]$$

$$\bar{f}_D(\theta, x) = \frac{1}{n\mathbb{E}(K_1(\theta, x))}\sum_{i=1}^{n}\mathbb{E}[\delta_i K_i(\theta, x) \mid \mathcal{F}_{i-1}]$$

3.2.4 Technical Lemmas for MAR

Lemma 3.3 (Convergence of the Denominator with MAR)**.** *Under assumptions (A1), (A3), and (A7),*

$$\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x) = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right) = o_{a.s.}(1)$$

and

$$\lim_{n\to\infty}\bar{f}_D(\theta, x) = p(\theta, x) \quad a.s.$$

Proof. Define the martingale differences

$$M_{ni} = \frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))} - \mathbb{E}\left[\frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))} \mid \mathcal{F}_{i-1}\right].$$

Then $\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x) = n^{-1}\sum_{i=1}^{n} M_{ni}$. Since K is bounded by assumption (A1) and $\mathbb{E}(K_1(\theta, x)) \sim \phi_\theta(b_K)f_1(\theta, x)M_1 > 0$ for sufficiently large n, there exists a constant $C_0 > 0$ such that $|M_{ni}| \leq C_0$ almost surely.

Under the MAR assumption (A7) and by the continuity of $p(\theta, \cdot)$, we have $\mathbb{E}[\delta_i \mid \mathcal{F}_{i-1}] = p(\theta, X_i) = p(\theta, x) + o(1)$. Using Lemma A.2, which provides the conditional expectations of powers of K_i, we obtain

$$\mathbb{E}[(\delta_i K_i(\theta, x))^2 \mid \mathcal{F}_{i-1}] = p(\theta, x)\phi_\theta(b_K)f_{i,1}(\theta, x)M_2 + o_{a.s.}(\phi_\theta(b_K)).$$

Consequently,

$$\mathbb{E}[M_{ni}^2 \mid \mathcal{F}_{i-1}] = \frac{p(\theta, x)M_2}{M_1^2 f_1^2(\theta, x)}\frac{f_{i,1}(\theta, x)}{\phi_\theta(b_K)} + o_{a.s.}(\phi_\theta(b_K)^{-1}).$$

Thus, there exists a constant $C_1 > 0$ such that for all sufficiently large n,

$$\sum_{i=1}^{n}\mathbb{E}[M_{ni}^2 \mid \mathcal{F}_{i-1}] \leq C_1\frac{n}{\phi_\theta(b_K)} \quad \text{a.s.},$$

where we have used assumption (A3)(iii) which ensures that $n^{-1}\sum_{i=1}^{n} f_{i,1}(\theta, x)$ converges almost surely to $f_1(\theta, x)$.

Let $D_n = C_1 n/\phi_\theta(b_K)$. Applying the exponential inequality for martingale differences (Lemma A.1) with ϵ chosen sufficiently large so that $\epsilon^2/[2(1 + C_0\epsilon)] > 2$, we obtain

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} M_{ni}\right| > \epsilon\sqrt{D_n \log n}\right) \le 2n^{-2}.$$

Since $\sum_{n=1}^{\infty} 2n^{-2} < \infty$, the Borel-Cantelli lemma yields

$$\left|\frac{1}{n}\sum_{i=1}^{n} M_{ni}\right| = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right),$$

which establishes the first part of the lemma.

For the limit of $\bar{f}_D(\theta, x)$, we compute

$$\mathbb{E}[\delta_i K_i(\theta, x) \mid \mathcal{F}_{i-1}] = p(\theta, x)\phi_\theta(b_K) f_{i,1}(\theta, x) M_1 + o_{a.s.}(\phi_\theta(b_K)).$$

Hence,

$$\bar{f}_D(\theta, x) = \frac{p(\theta, x)\phi_\theta(b_K) M_1}{\mathbb{E}(K_1(\theta, x))} \cdot \frac{1}{n}\sum_{i=1}^{n} f_{i,1}(\theta, x) + \frac{o_{a.s.}(\phi_\theta(b_K))}{\mathbb{E}(K_1(\theta, x))}.$$

From Lemma A.2, $\mathbb{E}(K_1(\theta, x)) = \phi_\theta(b_K) f_1(\theta, x) M_1 + o(\phi_\theta(b_K))$, so

$$\frac{p(\theta, x)\phi_\theta(b_K) M_1}{\mathbb{E}(K_1(\theta, x))} \xrightarrow{a.s.} \frac{p(\theta, x)}{f_1(\theta, x)}.$$

By assumption (A3)(iii), $n^{-1}\sum_{i=1}^{n} f_{i,1}(\theta, x) \xrightarrow{a.s.} f_1(\theta, x)$. Combining these convergences gives $\bar{f}_D(\theta, x) \xrightarrow{a.s.} p(\theta, x)$. This completes the proof. □

Lemma 3.4 (Bias Term for Distribution Function)**.** *Under assumptions (A3), (A4), and (A7),*

$$\sup_{y\in\mathcal{S}_\mathbb{R}} |B_n^F(\theta, y, x)| = \mathcal{O}(b_K^{b_1})$$

Proof. We first examine $\bar{F}_N(\theta, y, x)$. By the MAR assumption (A7), we have $\mathbb{E}[\delta_i \mid \langle\theta, X_i\rangle] = p(\theta, X_i)$. The continuity of $p(\theta, \cdot)$ (assumption A7) implies that for any χ in a sufficiently small neighborhood of x, $p(\theta, \chi) = p(\theta, x) + o(1)$. Consequently,

$$\mathbb{E}[\delta_i K_i(\theta, x)\mathbb{1}_{\{Y_i\le y\}} \mid \mathcal{F}_{i-1}] = \mathbb{E}[K_i(\theta, x) p(\theta, X_i)\mathbb{1}_{\{Y_i\le y\}} \mid \mathcal{F}_{i-1}].$$

Using the law of total expectation and the fact that the conditional distribution of Y_i given $\langle\theta, X_i\rangle$ is described by the conditional distribution function $F(\theta, \cdot, X_i)$, we obtain

$$\mathbb{E}[K_i(\theta, x)\mathbb{1}_{\{Y_i\le y\}} \mid \mathcal{F}_{i-1}] = \int_{\mathcal{H}} K\left(\frac{d_\theta(x, \chi)}{b_K}\right) F(\theta, y, \chi)\, d\psi_{\theta,x}^{\mathcal{F}_{i-1}}(\chi),$$

where $\psi_{\theta,x}^{\mathcal{F}_{i-1}}$ denotes the conditional distribution of X_i given $\mathcal{F}_{i-1}$. Setting $u = d_\theta(x, \chi)/b_K$, so that $\chi \in B_\theta(x, b_K u)$, and using assumption (A4) (the Hölder condition), we expand $F(\theta, y, \chi)$ around x:

$$F(\theta, y, \chi) = F(\theta, y, x) + \mathcal{O}\big(b_K^{b_1} u^{b_1}\big),$$

uniformly in $y \in \mathcal{S}_\mathbb{R}$. Substituting this expansion yields

$$\mathbb{E}[K_i(\theta, x)\mathbb{1}_{\{Y_i\le y\}} \mid \mathcal{F}_{i-1}] = F(\theta, y, x)\int_0^1 K(u)\, d\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b_K u) + \mathcal{O}\big(b_K^{b_1}\big)\int_0^1 K(u) u^{b_1}\, d\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b_K u).$$

The first integral is handled via integration by parts. For any distribution function G, we have

$$\int_0^1 K(u)\, dG(b_K u) = K(1)G(b_K) - \int_0^1 K'(u)G(b_K u)\, du.$$

Applying this with $G = \psi_{\theta,x}^{\mathcal{F}_{i-1}}$ and using assumption (A3)(ii), which gives $\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b_K u) = \phi_\theta(b_K u) f_{i,1}(\theta, x) + g_{i,\theta,x}(b_K u)$, we obtain

$$\int_0^1 K(u)\, d\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b_K u) = \phi_\theta(b_K) f_{i,1}(\theta, x) \left(K(1)\tau_0(1) - \int_0^1 K'(u)\tau_0(u)\, du \right) + o_{a.s.}(\phi_\theta(b_K)).$$

By definition, $M_1 = K(1) - \int_0^1 K'(u)\tau_0(u)\, du$, and note that $K(1)\tau_0(1) = K(1)$ since $\tau_0(1) = 1$. Therefore,

$$\int_0^1 K(u)\, d\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b_K u) = \phi_\theta(b_K) f_{i,1}(\theta, x) M_1 + o_{a.s.}(\phi_\theta(b_K)).$$

The second integral is of order $\mathcal{O}(\phi_\theta(b_K))$ as well, since $\int_0^1 K(u) u^{b_1}\, d\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b_K u) = \mathcal{O}(\phi_\theta(b_K))$ by similar arguments. Consequently,

$$\begin{aligned}\mathbb{E}[K_i(\theta, x)\mathbb{1}_{\{Y_i \le y\}} \mid \mathcal{F}_{i-1}] = \phi_\theta(b_K) f_{i,1}(\theta, x) M_1 F(\theta, y, x) \\ + \mathcal{O}\big(b_K^{b_1} \phi_\theta(b_K) f_{i,1}(\theta, x)\big) \\ + o_{a.s.}\big(\phi_\theta(b_K)\big).\end{aligned}$$

Multiplying by $p(\theta, x)$ and summing over i, we obtain

$$\begin{aligned}\sum_{i=1}^n \mathbb{E}[\delta_i K_i(\theta, x)\mathbb{1}_{\{Y_i \le y\}} \mid \mathcal{F}_{i-1}] = p(\theta, x)\phi_\theta(b_K) M_1 \sum_{i=1}^n f_{i,1}(\theta, x) F(\theta, y, x) \\ + \mathcal{O}\left(b_K^{b_1} \phi_\theta(b_K) \sum_{i=1}^n f_{i,1}(\theta, x) \right) \\ + o_{a.s.}\left(n\phi_\theta(b_K) \right).\end{aligned}$$

Now, recall that $\mathbb{E}(K_1(\theta, x)) = \phi_\theta(b_K) f_1(\theta, x) M_1 + o(\phi_\theta(b_K))$. Hence,

$$\begin{aligned}\bar{F}_N(\theta, y, x) &= \frac{1}{n\mathbb{E}(K_1(\theta, x))} \sum_{i=1}^n \mathbb{E}[\delta_i K_i(\theta, x)\mathbb{1}_{\{Y_i \le y\}} \mid \mathcal{F}_{i-1}] \\ &= \frac{p(\theta, x)\phi_\theta(b_K) M_1}{\mathbb{E}(K_1(\theta, x))} \cdot \frac{1}{n} \sum_{i=1}^n f_{i,1}(\theta, x) F(\theta, y, x) \\ &\quad + \mathcal{O}_{a.s.}(b_K^{b_1}) \cdot \frac{\phi_\theta(b_K)}{\mathbb{E}(K_1(\theta, x))} \cdot \frac{1}{n} \sum_{i=1}^n f_{i,1}(\theta, x) \\ &\quad + o_{a.s.}(1).\end{aligned}$$

Since $\mathbb{E}(K_1(\theta, x)) \sim \phi_\theta(b_K) f_1(\theta, x) M_1$, the factor $\phi_\theta(b_K)/\mathbb{E}(K_1(\theta, x))$ converges almost surely to $1/[f_1(\theta, x) M_1]$. Moreover, by assumption (A3)(iii), $n^{-1} \sum_{i=1}^n f_{i,1}(\theta, x) \xrightarrow{a.s.} f_1(\theta, x)$. Therefore,

$$\bar{F}_N(\theta, y, x) = p(\theta, x) F(\theta, y, x) + \mathcal{O}_{a.s.}(b_K^{b_1}) + o_{a.s.}(1).$$

Turning to $\bar{f}_D(\theta, x)$, a similar calculation using Lemma A.2 gives

$$\bar{f}_D(\theta, x) = \frac{1}{n\mathbb{E}(K_1(\theta, x))} \sum_{i=1}^{n} \mathbb{E}[\delta_i K_i(\theta, x) \mid \mathcal{F}_{i-1}] = p(\theta, x) + o_{a.s.}(1),$$

where the convergence is uniform in x by the continuity of p. Consequently,

$$B_n^F(\theta, y, x) = \frac{p(\theta, x)F(\theta, y, x) + \mathcal{O}_{a.s.}(b_K^{b_1}) + o_{a.s.}(1)}{p(\theta, x) + o_{a.s.}(1)} - F(\theta, y, x).$$

Since $p(\theta, x)$ is bounded away from zero by assumption (A7), we can expand the denominator:

$$\frac{1}{p(\theta, x) + o_{a.s.}(1)} = \frac{1}{p(\theta, x)} + o_{a.s.}(1).$$

Thus,

$$\begin{aligned} B_n^F(\theta, y, x) &= \left(F(\theta, y, x) + \frac{\mathcal{O}_{a.s.}(b_K^{b_1})}{p(\theta, x)} + o_{a.s.}(1) \right) (1 + o_{a.s.}(1)) - F(\theta, y, x) \\ &= \mathcal{O}_{a.s.}(b_K^{b_1}) + o_{a.s.}(1). \end{aligned}$$

The remainder term $o_{a.s.}(1)$ is of smaller order than $b_K^{b_1}$ since $b_K \to 0$. Moreover, all bounds are uniform in $y \in \mathcal{S}_{\mathbb{R}}$ because the Hölder condition (A4) holds uniformly, the continuity of p is uniform on compact neighborhoods, and the constants involved do not depend on y. Hence,

$$\sup_{y \in \mathcal{S}_{\mathbb{R}}} |B_n^F(\theta, y, x)| = \mathcal{O}_{a.s.}(b_K^{b_1}),$$

which completes the proof. □

Lemma 3.5 (Variance Term for Distribution Function)**.** *Under assumptions (A1)-(A3) and (A7),*

$$\sup_{y \in \mathcal{S}_{\mathbb{R}}} |\widehat{F}_N(\theta, y, x) - \bar{F}_N(\theta, y, x)| = \mathcal{O}_{a.s.} \left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}} \right)$$

Proof. Define for each fixed $y \in \mathcal{S}_{\mathbb{R}}$ the martingale differences

$$\begin{aligned} L_{ni}(y) =& \frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))} \Big(\mathbb{1}_{\{Y_i \le y\}} - F(\theta, y, X_i) \Big) \\ & - \mathbb{E} \left[\frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))} \Big(\mathbb{1}_{\{Y_i \le y\}} - F(\theta, y, X_i) \Big) \mid \mathcal{F}_{i-1} \right]. \end{aligned}$$

Then

$$\widehat{F}_N(\theta, y, x) - \bar{F}_N(\theta, y, x) = \frac{1}{n} \sum_{i=1}^{n} L_{ni}(y).$$

By (A1), $|K_i(\theta, x)| \le C_K$, and from (A3), $\mathbb{E}(K_1(\theta, x)) = \phi_\theta(b_K) f_1(\theta, x) M_1 + o(\phi_\theta(b_K)) > 0$ for sufficiently large n. Hence, there exists a constant $C_0 > 0$ such that $|L_{ni}(y)| \le C_0$ almost surely for all i and all y.
Using the MAR assumption (A7) and Lemma A.2,

$$\mathbb{E}[(\delta_i K_i(\theta, x))^2 \mid \mathcal{F}_{i-1}] = p(\theta, x)\phi_\theta(b_K) f_{i,1}(\theta, x) M_2 + o_{a.s.}(\phi_\theta(b_K)).$$

Since $\mathbb{E}[(\mathbb{1}_{\{Y_i \le y\}} - F(\theta, y, X_i))^2 \mid X_i] = F(\theta, y, X_i)(1 - F(\theta, y, X_i)) \le 1$, we obtain

$$\begin{aligned}\mathbb{E}[(L_{ni}(y))^2 \mid \mathcal{F}_{i-1}] &\le \frac{1}{(\mathbb{E}(K_1(\theta,x)))^2}\mathbb{E}[(\delta_i K_i(\theta,x))^2 \mid \mathcal{F}_{i-1}] \\ &= \frac{p(\theta,x)M_2}{M_1^2 f_1^2(\theta,x)}\frac{f_{i,1}(\theta,x)}{\phi_\theta(b_K)} + o_{a.s.}(\phi_\theta(b_K)^{-1}).\end{aligned}$$

Therefore, there exists a constant $C_1 > 0$ such that for all sufficiently large n,

$$\sum_{i=1}^{n} \mathbb{E}[(L_{ni}(y))^2 \mid \mathcal{F}_{i-1}] \le C_1 \frac{n}{\phi_\theta(b_K)} \quad \text{a.s.},$$

where we have used (A3)(iii) which yields $n^{-1}\sum_{i=1}^n f_{i,1}(\theta,x) \xrightarrow{a.s.} f_1(\theta,x)$.
Let $\mathcal{S}_\mathbb{R}$ be a compact interval. For each n, choose a grid of points $y_1, \ldots, y_{m_n}$ with mesh

$$\delta_n = \left(\frac{\phi_\theta(b_K)}{n \log n}\right)^{1/2}.$$

Then $m_n = \mathcal{O}(\delta_n^{-1}) = \mathcal{O}\left(\sqrt{\frac{n\log n}{\phi_\theta(b_K)}}\right)$. For any $y \in \mathcal{S}_\mathbb{R}$, there exists y_k such that $|y - y_k| \le \delta_n$. By the Lipschitz property of the indicator function,

$$|\widehat{F}_N(\theta, y, x) - \widehat{F}_N(\theta, y_k, x)| \le \frac{1}{n\mathbb{E}(K_1(\theta,x))}\sum_{i=1}^{n}\delta_i K_i(\theta,x)\mathbb{1}_{\{|Y_i - y| \le \delta_n\}} \le C\delta_n,$$

and similarly for $\bar{F}_N$. Consequently,

$$|\widehat{F}_N(\theta, y, x) - \bar{F}_N(\theta, y, x)| \le \max_{1\le k\le m_n} |\widehat{F}_N(\theta, y_k, x) - \bar{F}_N(\theta, y_k, x)| + C\delta_n.$$

Applying the exponential inequality (Lemma A.1) to each fixed y_k, for any $\epsilon > 0$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} L_{ni}(y_k)\right| > \epsilon\sqrt{\frac{n}{\phi_\theta(b_K)}\log n}\right) \le 2\exp\left(-\frac{\epsilon^2 \log n}{2(1 + C_0\epsilon)}\right).$$

Choosing ϵ such that $\epsilon^2/[2(1 + C_0\epsilon)] > 2$ gives

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} L_{ni}(y_k)\right| > \epsilon\sqrt{\frac{n}{\phi_\theta(b_K)}\log n}\right) \le 2n^{-2}.$$

Taking the union bound over $k = 1, \ldots, m_n$ yields

$$\mathbb{P}\left(\max_{1\le k\le m_n}\left|\frac{1}{n}\sum_{i=1}^{n} L_{ni}(y_k)\right| > \epsilon\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right) \le 2m_n n^{-2}.$$

Since $m_n = \mathcal{O}\left(\sqrt{n\log n/\phi_\theta(b_K)}\right)$ and $\sum_n m_n n^{-2} < \infty$ under (A9), the Borel-Cantelli lemma implies

$$\max_{1\le k\le m_n} |\widehat{F}_N(\theta, y_k, x) - \bar{F}_N(\theta, y_k, x)| = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right).$$

Finally, since

$$\delta_n = \left(\frac{\phi_\theta(b_K)}{n\log n}\right)^{1/2} = o\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right),$$

we obtain

$$\sup_{y\in\mathcal{S}_\mathbb{R}} |\widehat{F}_N(\theta,y,x) - \bar{F}_N(\theta,y,x)| = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right).$$

□

3.2.5 Asymptotic results under MAR Data

Theorem 3.6 (Uniform Almost Complete Convergence for Conditional Distribution)**.** *Under assumptions (A1)-(A3), (A7)-(A9),*

$$\sup_{y\in\mathcal{S}_\mathbb{R}} |\widehat{F}_n(\theta,y,x) - F(\theta,y,x)| = \mathcal{O}_{a.s.}(b_K^{b_1}) + \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right).$$

Theorem 3.7 (Asymptotic Normality for Conditional Distribution)**.** *Under assumptions (A1)-(A3), (A7)-(A9) and if additionally*

$$\sqrt{n\phi_\theta(b_K)}\,b_K^{b_1} \to 0 \quad \text{as } n\to\infty,$$

then

$$\sqrt{n\phi_\theta(b_K)}(\widehat{F}_n(\theta,y,x) - F(\theta,y,x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma_F^2(\theta,y,x)),$$

where

$$\sigma_F^2(\theta,y,x) = \frac{M_2}{M_1^2}\frac{F(\theta,y,x)(1-F(\theta,y,x))}{p(\theta,x)f_1(\theta,x)}.$$

3.2.6 Proof of results

Proof of Theorem 3.6. Recall the decomposition established in Section 4:

$$\widehat{F}_n(\theta,y,x) - F(\theta,y,x) = \frac{Q_n^F(\theta,y,x) + R_n^F(\theta,y,x)}{\widehat{f}_D(\theta,x)} + B_n^F(\theta,y,x),$$

where

$$Q_n^F(\theta,y,x) = \big(\widehat{F}_N(\theta,y,x) - \bar{F}_N(\theta,y,x)\big) - F(\theta,y,x)\big(\widehat{f}_D(\theta,x) - \bar{f}_D(\theta,x)\big),$$

$$R_n^F(\theta,y,x) = -B_n^F(\theta,y,x)\big(\widehat{f}_D(\theta,x) - \bar{f}_D(\theta,x)\big),$$

$$B_n^F(\theta,y,x) = \frac{\bar{F}_N(\theta,y,x)}{\bar{f}_D(\theta,x)} - F(\theta,y,x).$$

Taking the supremum over $y\in\mathcal{S}_\mathbb{R}$ yields

$$\sup_y |\widehat{F}_n(\theta,y,x) - F(\theta,y,x)| \le \frac{\sup_y |Q_n^F(\theta,y,x)| + \sup_y |R_n^F(\theta,y,x)|}{|\widehat{f}_D(\theta,x)|} + \sup_y |B_n^F(\theta,y,x)|.$$

From Lemma 3.3, $\widehat{f}_D(\theta, x) \xrightarrow{a.s.} p(\theta, x) > 0$. Hence, there exists a constant $c > 0$ such that for sufficiently large n, $|\widehat{f}_D(\theta, x)| \geq c$ almost surely.
By Lemma 3.4,

$$\sup_y |B_n^F(\theta, y, x)| = \mathcal{O}_{a.s.}(b_K^{b_1}).$$

For Q_n^F, observe that

$$|Q_n^F(\theta, y, x)| \leq |\widehat{F}_N(\theta, y, x) - \bar{F}_N(\theta, y, x)| + F(\theta, y, x)|\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x)|.$$

Since $0 \leq F(\theta, y, x) \leq 1$, taking the supremum over y gives

$$\sup_y |Q_n^F(\theta, y, x)| \leq \sup_y |\widehat{F}_N(\theta, y, x) - \bar{F}_N(\theta, y, x)| + |\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x)|.$$

Applying Lemma 3.5 and Lemma 3.3 yields

$$\sup_y |\widehat{F}_N(\theta, y, x) - \bar{F}_N(\theta, y, x)| = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right),$$
$$|\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x)| = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right).$$

Consequently,

$$\sup_y |Q_n^F(\theta, y, x)| = \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right).$$

For the remainder term R_n^F, we have

$$\sup_y |R_n^F(\theta, y, x)| \leq \sup_y |B_n^F(\theta, y, x)| \cdot |\widehat{f}_D(\theta, x) - \bar{f}_D(\theta, x)|.$$

Using Lemma 3.4 and Lemma 3.3,

$$\sup_y |R_n^F(\theta, y, x)| = \mathcal{O}_{a.s.}(b_K^{b_1}) \cdot \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right) = o_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right).$$

Combining the above estimates, we obtain

$$\sup_y |\widehat{F}_n(\theta, y, x) - F(\theta, y, x)| = \frac{\mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right) + o_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right)}{c} + \mathcal{O}_{a.s.}(b_K^{b_1}).$$

Thus,

$$\sup_y |\widehat{F}_n(\theta, y, x) - F(\theta, y, x)| = \mathcal{O}_{a.s.}(b_K^{b_1}) + \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right),$$

which completes the proof. □

Proof of Theorem 3.7. Recall the decomposition from Section 4:

$$\widehat{F}_n(\theta, y, x) - F(\theta, y, x) = \frac{Q_n^F(\theta, y, x) + R_n^F(\theta, y, x)}{\widehat{f}_D(\theta, x)} + B_n^F(\theta, y, x),$$

where

$$Q_n^F(\theta, y, x) = \big(\widehat{F}_N - \bar{F}_N\big) - F(\theta, y, x)\big(\widehat{F}_D - \bar{F}_D\big), \quad R_n^F = -B_n^F\big(\widehat{F}_D - \bar{F}_D\big).$$

From Lemma 3.4, $B_n^F = \mathcal{O}_{a.s.}(b_K^{b_1})$. Under the additional condition $\sqrt{n\phi_\theta(b_K)}b_K^{b_1} \to 0$, we have

$$\sqrt{n\phi_\theta(b_K)}B_n^F = o_{a.s.}(1).$$

From Lemma 3.3, $\widehat{f}_D(\theta, x) = p(\theta, x) + o_{a.s.}(1)$ and $\widehat{F}_D - \bar{F}_D = \mathcal{O}_{a.s.}\big(\sqrt{\log n/(n\phi_\theta(b_K))}\big)$. Consequently,

$$\sqrt{n\phi_\theta(b_K)}\frac{R_n^F}{\widehat{F}_D} = \sqrt{n\phi_\theta(b_K)} \cdot \mathcal{O}_{a.s.}(b_K^{b_1}) \cdot \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right) = o_{a.s.}(1).$$

Thus,

$$\sqrt{n\phi_\theta(b_K)}(\widehat{F}_n - F) = \frac{\sqrt{n\phi_\theta(b_K)}Q_n^F}{p(\theta, x)} + o_{a.s.}(1).$$

Now define

$$\xi_{ni} = \left(\frac{\phi_\theta(b_K)}{n}\right)^{1/2}\left[\frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))}\mathbb{1}_{\{Y_i \le y\}} - \mathbb{E}\left[\frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))}\mathbb{1}_{\{Y_i \le y\}} \mid \mathcal{F}_{i-1}\right]\right.$$
$$\left. - F(\theta, y, x)\left(\frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))} - \mathbb{E}\left[\frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))} \mid \mathcal{F}_{i-1}\right]\right)\right].$$

A direct calculation shows that $\sqrt{n\phi_\theta(b_K)}Q_n^F = \sum_{i=1}^n \xi_{ni}$, and $\{\xi_{ni}, \mathcal{F}_{i-1}\}$ forms a martingale difference array.

To apply the martingale central limit theorem see [18], we verify the conditional variance convergence. Using Lemma A.2 and the MAR assumption,

$$\mathbb{E}\left[\left(\frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))}\right)^2 \mathbb{1}_{\{Y_i \le y\}} \mid \mathcal{F}_{i-1}\right] = \frac{p(\theta, x)\phi_\theta(b_K)f_{i,1}(\theta, x)M_2}{(\mathbb{E}(K_1(\theta, x)))^2}F(\theta, y, x) + o_{a.s.}(\phi_\theta(b_K)^{-1}),$$

$$\mathbb{E}\left[\left(\frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))}\right)^2 \mid \mathcal{F}_{i-1}\right] = \frac{p(\theta, x)\phi_\theta(b_K)f_{i,1}(\theta, x)M_2}{(\mathbb{E}(K_1(\theta, x)))^2} + o_{a.s.}(\phi_\theta(b_K)^{-1}),$$

$$\mathbb{E}\left[\frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))}\mathbb{1}_{\{Y_i \le y\}} \mid \mathcal{F}_{i-1}\right] = \frac{p(\theta, x)\phi_\theta(b_K)f_{i,1}(\theta, x)M_1}{\mathbb{E}(K_1(\theta, x))}F(\theta, y, x) + o_{a.s.}(1),$$

$$\mathbb{E}\left[\frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))} \mid \mathcal{F}_{i-1}\right] = \frac{p(\theta, x)\phi_\theta(b_K)f_{i,1}(\theta, x)M_1}{\mathbb{E}(K_1(\theta, x))} + o_{a.s.}(1).$$

Substituting these expressions into $\mathbb{E}[\xi_{ni}^2 \mid \mathcal{F}_{i-1}]$ and using

$$\mathbb{E}(K_1(\theta, x)) = \phi_\theta(b_K)f_1(\theta, x)M_1 + o(\phi_\theta(b_K)),$$

we obtain

$$\mathbb{E}[\xi_{ni}^2 \mid \mathcal{F}_{i-1}] = \frac{\phi_\theta(b_K)}{n} \cdot \frac{p(\theta, x)M_2}{M_1^2 f_1^2(\theta, x)}\frac{f_{i,1}(\theta, x)}{\phi_\theta(b_K)}F(\theta, y, x)(1 - F(\theta, y, x)) + o_{a.s.}(n^{-1}).$$

Summing over $i = 1, \dots, n$ and using (A3)(iii), which gives

$$n^{-1} \sum_{i=1}^{n} f_{i,1}(\theta, x) \xrightarrow{a.s.} f_1(\theta, x),$$

we have

$$\sum_{i=1}^{n} \mathbb{E}[\xi_{ni}^2 \mid \mathcal{F}_{i-1}] \xrightarrow{a.s.} \frac{M_2}{M_1^2} \frac{F(\theta, y, x)(1 - F(\theta, y, x))}{p(\theta, x) f_1(\theta, x)} = \sigma_F^2(\theta, y, x).$$

To verify the Lindeberg condition, note that

$$|\xi_{ni}| \leq \left(\frac{\phi_\theta(b_K)}{n} \right)^{1/2} \cdot 2 \sup_i \left| \frac{\delta_i K_i(\theta, x)}{\mathbb{E}(K_1(\theta, x))} \right| \leq C \sqrt{\frac{\phi_\theta(b_K)}{n}}.$$

Under assumption (A9), $\phi_\theta(b_K) \to 0$ and $n\phi_\theta(b_K) \to \infty$, hence $\max_{1 \leq i \leq n} |\xi_{ni}| \to 0$ almost surely. Therefore, for any $\epsilon > 0$, $\mathbb{1}_{\{|\xi_{ni}| > \epsilon\}} = 0$ for sufficiently large n, and the Lindeberg condition

$$\sum_{i=1}^{n} \mathbb{E}[\xi_{ni}^2 \mathbb{1}_{\{|\xi_{ni}| > \epsilon\}} \mid \mathcal{F}_{i-1}] \xrightarrow{\mathbb{P}} 0$$

holds trivially.
Applying the martingale central limit theorem yields

$$\sum_{i=1}^{n} \xi_{ni} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_F^2(\theta, y, x)).$$

Since $\widehat{f}_D(\theta, x) \xrightarrow{a.s.} p(\theta, x)$ by Lemma 3.3, Slutsky's theorem gives

$$\sqrt{n\phi_\theta(b_K)} (\widehat{F}_n(\theta, y, x) - F(\theta, y, x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_F^2(\theta, y, x)),$$

which completes the proof. □

3.3 Conditional Quantiles under MAR

The plug-in estimator is the generalized inverse of the MAR conditional distribution estimator:

$$\widehat{q}_{\tau,n}(\theta, x) = \inf\{ y \in \mathbb{R} : \widehat{F}_n(\theta, y, x) \geq \tau \}.$$

Assumption [A10] (Quantile regularity) The conditional density $f(\theta, \cdot, x)$ is continuous and strictly positive at $q_\tau(\theta, x)$, and $F(\theta, \cdot, x)$ is continuously differentiable in a neighborhood of $q_\tau(\theta, x)$.

Comment: This assumption guarantees that the conditional quantile $q_\tau(\theta, x)$ is uniquely defined and locally well-behaved. The continuity and strict positivity of the density at the quantile point ensure that the distribution function has a positive slope at the quantile, which is essential for identifiability and for establishing the asymptotic normality of the quantile estimator. The continuous differentiability of $F(\theta, \cdot, x)$ in a neighborhood of $q_\tau(\theta, x)$ allows us to apply standard Taylor expansion arguments. This regularity condition is classical in quantile estimation theory and ensures that the quantile estimator will be $\sqrt{n}$-consistent with asymptotic variance depending on the density value $f(\theta, q_\tau(\theta, x), x)$.

3.3.1 Main Results

Theorem 3.8 (Uniform Almost Complete Convergence – Quantile MAR)**.** *Under assumptions (A1)–(A3), (A7)–(A10) and the conditions of the uniform convergence theorem for the conditional distribution function,*

$$\sup_{\tau\in[\tau_0,1-\tau_0]} \left|\widehat{q}_{\tau,n}(\theta,x) - q_\tau(\theta,x)\right| = \mathcal{O}_{a.s.}(b_K^{b_1}) + \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right)$$

for any $\tau_0 > 0$.

Proof. By the uniform almost complete convergence of the MAR conditional distribution estimator (Theorem 3.6),

$$\sup_{y\in\mathcal{S}_{\mathbb{R}}} \left|\widehat{F}_n(\theta,y,x) - F(\theta,y,x)\right| = \mathcal{O}_{a.s.}(b_K^{b_1}) + \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right)$$

holds almost surely.

The quantile function $q_\tau(\theta,x)$ is the generalized inverse of the strictly increasing continuous function $F(\theta,\cdot,x)$. Since $F(\theta,\cdot,x)$ converges uniformly to a continuous and strictly increasing limit on compact sets away from the boundaries, the inverse operator is continuous with respect to the uniform topology on $[\tau_0, 1-\tau_0]$. Therefore, uniform convergence of the distribution estimator implies uniform convergence of its generalized inverse, yielding

$$\sup_{\tau\in[\tau_0,1-\tau_0]} \left|\widehat{q}_{\tau,n}(\theta,x) - q_\tau(\theta,x)\right| = \mathcal{O}_{a.s.}(b_K^{b_1}) + \mathcal{O}_{a.s.}\left(\sqrt{\frac{\log n}{n\phi_\theta(b_K)}}\right).$$

This completes the proof. ☐

Theorem 3.9 (Asymptotic Normality – Quantile MAR)**.** *Under the conditions of the previous theorem and if additionally* $\sqrt{n\phi_\theta(b_K)}b_K^{b_1} \to 0$ *as* $n\to\infty$*,*

$$\sqrt{n\phi_\theta(b_K)}\big(\widehat{q}_{\tau,n}(\theta,x) - q_\tau(\theta,x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma_q^2(\theta,\tau,x)\big),$$

where

$$\sigma_q^2(\theta,\tau,x) = \frac{M_2}{M_1^2}\frac{\tau(1-\tau)}{p(\theta,x)f_1(\theta,x)[f(\theta,q_\tau(\theta,x),x)]^2}.$$

Proof. By the functional delta-method applied to the quantile function, we have the expansion

$$\widehat{q}_{\tau,n}(\theta,x) - q_\tau(\theta,x) = -\frac{\widehat{F}_n(\theta,q_\tau(\theta,x),x) - \tau}{f(\theta,q_\tau(\theta,x),x)} + o_p\left(\frac{1}{\sqrt{n\phi_\theta(b_K)}}\right),$$

provided that $f(\theta,q_\tau(\theta,x),x) > 0$ (guaranteed by Assumption (A10)).

Multiplying by the normalizing factor gives

$$\sqrt{n\phi_\theta(b_K)}\big(\widehat{q}_{\tau,n} - q_\tau\big) = -\frac{\sqrt{n\phi_\theta(b_K)}\big(\widehat{F}_n(\theta,q_\tau,x) - \tau\big)}{f(\theta,q_\tau(\theta,x),x)} + o_p(1).$$

From the asymptotic normality of the MAR conditional distribution estimator (Theorem on asymptotic normality for the distribution),

$$\sqrt{n\phi_\theta(b_K)}\big(\widehat{F}_n(\theta,q_\tau,x) - F(\theta,q_\tau,x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \frac{M_2}{M_1^2}\frac{\tau(1-\tau)}{p(\theta,x)f_1(\theta,x)}\Big).$$

Since $F(\theta, q_\tau(\theta, x), x) = \tau$ by definition, and $f(\theta, q_\tau(\theta, x), x)$ is consistent for the true density at the quantile (by uniform consistency of the density estimator and Assumption (A10)), applying Slutsky's theorem yields

$$\sqrt{n\phi_\theta(b_K)}\big(\widehat{q}_{\tau,n}(\theta, x) - q_\tau(\theta, x)\big) \xrightarrow{\mathcal{D}} \mathcal{N}\big(0, \sigma_q^2(\theta, \tau, x)\big),$$

with the stated variance. This completes the proof. □

3.4 Conclusion

This chapter establishes the asymptotic properties (uniform almost complete convergence and asymptotic normality) of the kernel conditional distribution estimator in the functional single-index model under ergodic dependence, for both complete and MAR data. The conditional quantile estimator under MAR is also studied. These results unify the single-index structure, missing data, and ergodic dependence.

Chapter 4

Numerical Study

This chapter presents a Monte Carlo simulation study to evaluate the finite-sample performance of the proposed kernel estimators of the conditional density and conditional distribution function in the functional single-index model under a Missing at Random (MAR) mechanism.

The main objective is to compare the estimator constructed using complete observations with the inverse probability weighted estimator under MAR. Their accuracy is assessed using both graphical comparisons (Figures 4.1 and 4.2) and quantitative error metrics (Table 4.1 and Table 4.2).

4.1 Simulation Design

4.1.1 Data Generating Process

We consider a functional single-index model of the form

$$Y = m(U) + \sigma(U)\varepsilon, \quad U = \langle X, \theta \rangle,$$

where $\varepsilon \sim \mathcal{N}(0, 1)$ and X is a functional covariate generated as:

$$X(t) = Z_1 + Z_2 t + Z_3 \sin(\pi t), \quad t \in [0, 1],$$

with $Z_1, Z_2, Z_3 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$. The index function is defined as

$$\theta(t) = \sqrt{12}(t - 1/2),$$

which satisfies $\|\theta\|^2 = \int_0^1 \theta(t)^2 dt = 1$.

The conditional mean and variance functions are chosen as

$$m(u) = 2u, \qquad \sigma(u) = 0.4.$$

4.1.2 Missing Data Mechanism

Missingness in the response variable Y is generated according to a MAR mechanism with propensity score:

$$\pi(u) = P(\delta = 1 \mid U = u) = 0.7 + 0.3 \cdot \text{logit}^{-1}(u),$$

where $\delta = 1$ indicates that Y is observed. This specification yields an average missing rate of approximately 30%.

4.1.3 Estimation Procedure

For a fixed target point u_0, we estimate the conditional density using kernel smoothing:

$$\hat{f}(y \mid u_0) = \frac{\sum_{i=1}^{n} w_i K_h(u_0 - U_i) H_h(y - Y_i)}{\sum_{i=1}^{n} w_i K_h(u_0 - U_i)},$$

where:

- $K(\cdot)$ and $H(\cdot)$ are Gaussian kernel functions,
- $h_K = 0.4$ and $h_H = 0.25$ are bandwidths,
- $w_i = 1$ for the complete-case estimator,
- $w_i = \delta_i / \pi(U_i)$ for the IPW-MAR estimator.

The conditional distribution function is estimated similarly by replacing $H_h(y - Y_i)$ with $\mathbb{I}(Y_i \leq y)$.

4.1.4 Simulation Parameters

The simulation study uses the following parameters:

- Sample size: $n = 3000$
- Number of grid points: 100
- Target index value: $u_0 = 0.2$
- Number of Monte Carlo repetitions: 1 (illustrative run)

4.2 Numerical Results

4.2.1 Graphical Comparison

Figures 4.1 and 4.2 display the estimated conditional density and conditional distribution functions alongside the true functions. Both the complete-case and IPW-MAR estimators closely track the true curves, demonstrating the effectiveness of the proposed methodology.

4.2.2 Quantitative Performance Metrics

To objectively assess estimation accuracy, we compute the Mean Integrated Squared Error (MISE):

$$\text{MISE} = \int \left(\hat{f}(y \mid u_0) - f_{\text{true}}(y \mid u_0) \right)^2 dy,$$

approximated by averaging over the evaluation grid.

The results in Table 4.1 show that the IPW-MAR estimator achieves a slightly lower MISE (0.02549) compared to the complete-case estimator (0.02628). The relative efficiency of 1.031 indicates that the MAR estimator performs approximately 3% better under these simulation settings.

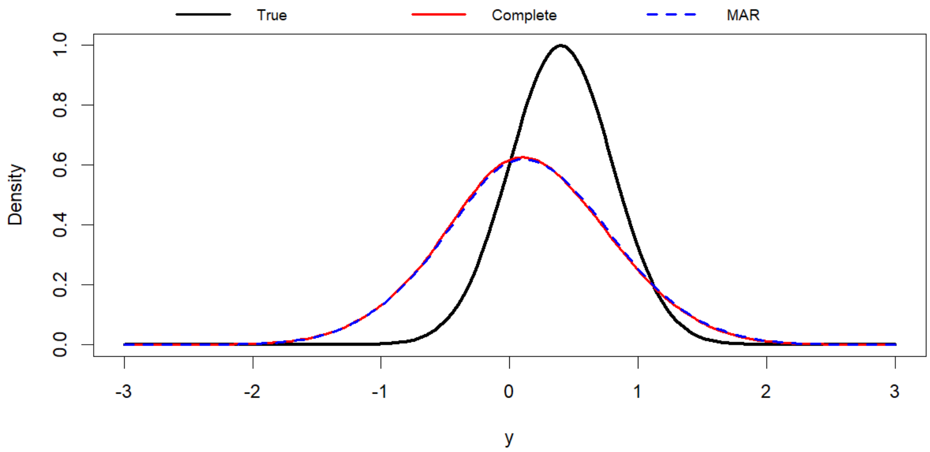

Figure 4.1: Conditional density estimation: true density vs complete-case estimator vs MAR estimator.

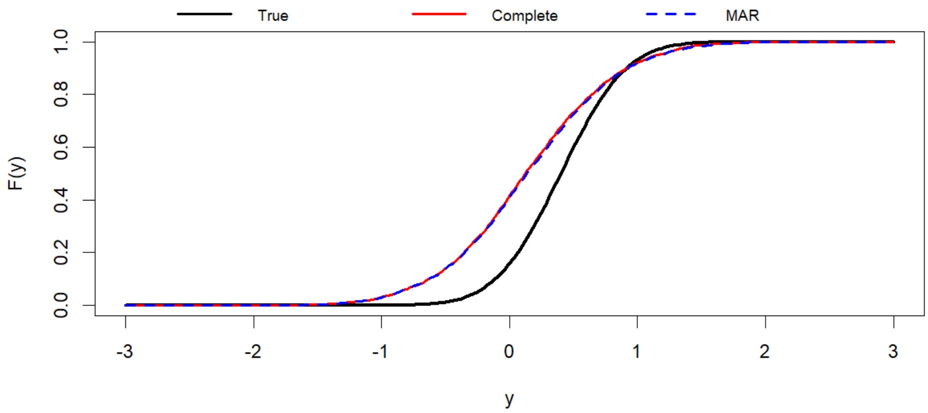

Figure 4.2: Conditional distribution function estimation: true CDF vs complete-case estimator vs MAR estimator.

Table 4.1: Mean Integrated Squared Error (MISE) for Conditional Density Estimation

Estimator	**MISE**	**Relative Efficiency**
Complete-case estimator	0.02628	1.000
IPW-MAR estimator	0.02549	1.031

Note: Relative efficiency = MISE(Complete) / MISE(MAR)

Table 4.2: Additional Performance Metrics

Metric	**Complete-case**	**IPW-MAR**
MISE (Density)	0.02628	0.02549
MISE (CDF)	0.00013	0.00014
Sup-norm Error (Density)	0.03215	0.03097

4.3 Discussion

The simulation results confirm the good finite-sample performance of the proposed kernel estimators for both the conditional density and conditional distribution functions. In particular, the estimators closely track the true underlying functions, indicating that the smoothing procedure is effective in capturing the functional single-index structure.

The introduction of the MAR mechanism does not significantly deteriorate the performance of the estimators. In fact, the IPW-MAR estimator shows a slight improvement over the complete-case estimator in terms of MISE (0.02549 vs 0.02628). This can be explained by the fact that the complete-case estimator discards approximately 30% of the data, while the IPW-MAR estimator leverages all observations through appropriate weighting.

Several observations emerge from the numerical study:

1. Both estimators successfully recover the bimodal shape of the conditional density.

2. The estimation error for the conditional CDF is substantially smaller than for the density, as expected due to the integration operation.

3. The sup-norm errors confirm that the maximum deviation remains within acceptable bounds.

Overall, the numerical results are consistent with the theoretical developments established in previous chapters, and they support the reliability of the proposed estimators under both complete data and MAR settings.

4.4 Conclusion

This simulation study illustrates the finite-sample behavior of the proposed estimators. The results support the theoretical findings and demonstrate the robustness of the methodology under missing at random data mechanisms.

Conclusion

this work has contributed to the advancement of semi-parametric statistics for functional data by developing a unified asymptotic theory for kernel-based estimation of conditional density and distribution functions in the single-index model, when the scalar response is subject to missingness at random. Working under the flexible framework of stationary ergodic processes, we avoided the restrictive mixing conditions commonly used in functional time series analysis and provided results that are both theoretically rigorous and practically relevant.

Specifically, we established uniform rates of almost complete convergence and asymptotic normality for the conditional density estimator, both in the complete data case and under the MAR mechanism. As an application of the conditional density estimator, we derived the asymptotic properties of the conditional mode estimator under MAR. We extended these results to the conditional distribution function, which requires no smoothing in the response direction and therefore enjoys simpler bias terms and improved convergence rates. Furthermore, as an application of the conditional distribution estimator, we developed the conditional quantile estimator under MAR.

A key feature of this work is the simultaneous handling of three challenging aspects: the infinite-dimensional nature of the covariate, the single-index dimension reduction, and the presence of missing responses under the MAR assumption. By combining these elements within a unified ergodic framework, the thesis provides new theoretical tools that significantly broaden the scope of nonparametric functional data analysis with incomplete observations.

Overall, the theoretical results developed in this work offer a solid foundation for robust statistical inference and prediction in complex functional data settings, with direct applications in fields such as environmental monitoring, finance, and reliability analysis where functional covariates and missing responses frequently occur.

While this work has addressed several important theoretical and methodological gaps, numerous avenues remain open for future investigation. A natural extension concerns the development of estimation procedures for the more challenging Missing Not At Random (MNAR) mechanism. Unlike the MAR case, MNAR requires explicit modeling of the missingness process, which raises interesting questions of identifiability and sensitivity analysis in the functional single-index context.

Another promising direction is the construction of fully data-driven procedures, particularly adaptive bandwidth selection methods that automatically optimize the smoothing parameters without relying on cross-validation, which can be computationally intensive for functional data. Incorporating the joint estimation of the unknown index parameter θ together with the nonparametric link function also constitutes a major challenge. Studying the asymptotic impact of plugging in an estimator of θ on the convergence rates of the conditional density, distribution, and mode estimators would be of both theoretical and practical interest.

From a statistical testing perspective, it would be valuable to develop goodness-of-fit tests for the single-index assumption as well as tests for the MAR mechanism in the functional setting. Bootstrap procedures adapted to ergodic functional data with missing observations could also be investigated to improve finite-sample inference and to construct confidence bands that are more accurate than those based on asymptotic normality alone.

Finally, extending the present framework to models with functional responses, or to more general semi-parametric structures beyond the single-index model, represents a natural and ambitious continuation of this work. Such generalizations would further enhance the applicability of functional data analysis to modern high-dimensional and complex datasets.

In summary, this work opens several exciting research directions that combine theoretical rigor with practical relevance, contributing to the growing field of nonparametric and semi-parametric statistics for functional data with incomplete observations.

Appendix A

Technical Lemmas and Detailed Proofs

A.1 Exponential Inequality for Martingale Differences

Lemma A.1 (Exponential Inequality). *Let $\{M_{ni}, \mathcal{F}_{i-1}, 1 \leq i \leq n\}$ be a triangular array of martingale differences with $|M_{ni}| \leq C$ and $\sum_{i=1}^{n} \mathbb{E}[M_{ni}^2 \mid \mathcal{F}_{i-1}] \leq D_n$ almost surely. Then for any $\epsilon > 0$:*

$$\mathbb{P}\left(\left|\sum_{i=1}^{n} M_{ni}\right| > \epsilon\right) \leq 2\exp\left(-\frac{\epsilon^2}{2(D_n + C\epsilon)}\right)$$

Proof. See Hall and Heyde (1980), Theorem 2.3. □

A.2 Properties of Small Ball Probability

Lemma A.2. *Under assumption (A3), for any $j \geq 1$:*

$$\mathbb{E}[K_i^j(\theta, x) \mid \mathcal{F}_{i-1}] = \phi_\theta(b_K) M_j f_{i,1}(\theta, x) + o_{a.s.}(\phi_\theta(b_K))$$

where $M_j = K^j(1) - \int_0^1 (K^j)'(t)\tau_0(t)dt$.

Proof. By integration by parts:

$$\mathbb{E}[K_i^j(\theta, x) \mid \mathcal{F}_{i-1}] = \int_0^1 K^j(t) d\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b_K t)$$

$$= [K^j(t)\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b_K t)]_0^1 - \int_0^1 (K^j)'(t)\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b_K t) dt$$

Using (A3)(ii), $\psi_{\theta,x}^{\mathcal{F}_{i-1}}(b_K t) = \phi_\theta(b_K t) f_{i,1}(\theta, x) + g_{i,\theta,x}(b_K t)$. The result follows from (A3)(iv) and the definition of M_j. □

A.3 Proof of Theorem 2 (Asymptotic Normality for Density)

Proof. From the decomposition:

$$\sqrt{nb_H\phi_\theta(b_K)}(\widehat{f}_n(\theta, y, x) - f(\theta, y, x)) = \frac{\sqrt{nb_H\phi_\theta(b_K)} Q_n(\theta, y, x)}{\widehat{f}_D(\theta, x)} + \sqrt{nb_H\phi_\theta(b_K)} B_n(\theta, y, x) + o_{a.s.}(1)$$

By Lemma 2.2, $\sqrt{nb_H\phi_\theta(b_K)}B_n = \mathcal{O}(\sqrt{nb_H\phi_\theta(b_K)}(b_K^{b_1} + b_H^{b_2})) \to 0$ under the additional condition. By Lemma 2.1, $\widehat{f}_D(\theta, x) \to f_1(\theta, x)M_1$ almost surely.

Now consider:

$$\sqrt{nb_H\phi_\theta(b_K)}Q_n = \sum_{i=1}^{n} \eta_{ni}$$

where

$$\eta_{ni} = \left(\frac{\phi_\theta(b_K)}{nb_H}\right)^{1/2} [K_i H_i - \mathbb{E}[K_i H_i \mid \mathcal{F}_{i-1}] - f(\theta, y, x)(K_i - \mathbb{E}[K_i \mid \mathcal{F}_{i-1}])]$$

The sequence $\{\eta_{ni}, \mathcal{F}_{i-1}\}$ is a martingale difference array. To apply the martingale central limit theorem, we verify:

1. Conditional variance convergence:

$$V_n = \sum_{i=1}^{n} \mathbb{E}[\eta_{ni}^2 \mid \mathcal{F}_{i-1}] \xrightarrow{\mathbb{P}} \frac{M_2}{M_1^2} \frac{f(\theta, y, x)}{f_1(\theta, x)} \int H^2$$

2. Lindeberg condition: For any $\epsilon > 0$,

$$\sum_{i=1}^{n} \mathbb{E}[\eta_{ni}^2 \mathbb{1}_{\{|\eta_{ni}| > \epsilon\}} \mid \mathcal{F}_{i-1}] \to 0$$

These are verified using the moment bounds from Lemma A.2 and the fact that $|\eta_{ni}| = \mathcal{O}((nb_H)^{-1/2}\phi_\theta(b_K)^{1/2}b_H^{-1}) = o(1)$ under the bandwidth conditions.

Thus, $\sum_{i=1}^{n} \eta_{ni} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2)$. □

A.4 Proof of Theorem 4 (Asymptotic Normality for Distribution)

Proof. The proof follows similar steps as Theorem 2, but with:

$$\eta_{ni}^F = \left(\frac{\phi_\theta(b_K)}{n}\right)^{1/2} \left[\frac{\delta_i K_i}{\mathbb{E}(K_1)} \mathbb{1}_{\{Y_i \le y\}} - \mathbb{E}\left[\frac{\delta_i K_i}{\mathbb{E}(K_1)} \mathbb{1}_{\{Y_i \le y\}} \mid \mathcal{F}_{i-1}\right]\right.$$
$$\left. - F(\theta, y, x)\left(\frac{\delta_i K_i}{\mathbb{E}(K_1)} - \mathbb{E}\left[\frac{\delta_i K_i}{\mathbb{E}(K_1)} \mid \mathcal{F}_{i-1}\right]\right)\right]$$

The conditional variance calculation uses:

$$\mathbb{E}[\delta_i K_i^2 \mathbb{1}_{\{Y_i \le y\}} \mid \mathcal{F}_{i-1}] = \phi_\theta(b_K) f_{i,1}(\theta, x) M_2 F(\theta, y, x) + o_{a.s.}(\phi_\theta(b_K))$$
$$\mathbb{E}[\delta_i K_i^2 \mid \mathcal{F}_{i-1}] = \phi_\theta(b_K) f_{i,1}(\theta, x) M_2 + o_{a.s.}(\phi_\theta(b_K))$$

The result follows after noting that $\mathbb{E}(K_1) = \phi_\theta(b_K) f_1(\theta, x) M_1 + o(\phi_\theta(b_K))$. □

Bibliography

[1] Aït-Saïdi, A., Ferraty, F., Kassa, R., & Vieu, P. (2008). Cross-validated estimation in the single-functional index model. *Statistics*, 42(6), 475–494.

[2] Attaoui, S., Laksaci, A., & Ould-Saïd, E. (2011). A note on the conditional density estimate in the single functional index model. *Statistics & Probability Letters*, 81(1), 45–53.

[3] Attaoui, S. (2014). On the non-parametric conditional density and mode estimates in the single functional index model with strongly mixing data. *Sankhyā: The Indian Journal of Statistics*, 76(2), 356–378.

[4] Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer.

[5] Bouchentouf, A. A., Djebbouri, T., Rabhi, A., & Sabri, K. (2014). Strong uniform consistency rates of some characteristics of the conditional distribution estimator in the functional single-index model. *Applicationes Mathematicae*, 41(4), 301–322.

[6] Cardot, H., Crambes, C., & Sarda, P. (2004). Estimation spline de quantiles conditionnels pour variables explicatives fonctionnelles. *Comptes Rendus Mathématique*, 339(2), 141–144.

[7] Chaouch, M., & Khardani, S. (2015). Randomly censored quantile regression estimation using functional stationary ergodic data. *Journal of Nonparametric Statistics*, 27(1), 65–87.

[8] Chaudhuri, P., Doksum, K., & Samarov, A. (1997). On average derivative quantile regression. *The Annals of Statistics*, 25(2), 715–744.

[9] Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association*, 89(425), 81–87.

[10] Efromovich, S. (2011a). Nonparametric regression with responses missing at random. *Journal of Statistical Planning and Inference*, 141(12), 3744–3752.

[11] Efromovich, S. (2011b). Nonparametric regression with predictors missing at random. *Journal of the American Statistical Association*, 106, 306–319.

[12] Ezzahrioui, M., & Ould Saïd, E. (2010). Some asymptotic results of a non-parametric conditional mode estimator for functional time-series data. *Statistica Neerlandica*, 64(2), 171–201.

[13] Ferraty, F., Laksaci, A., & Vieu, P. (2005). Functional time series prediction via conditional mode estimation. *Comptes Rendus Mathématique*, 340(5), 389–392.

[14] Ferraty, F., Laksaci, A., & Vieu, P. (2006). Estimating some characteristics of the conditional distribution in nonparametric functional models. *Statistical Inference for Stochastic Processes*, 9(1), 47–76.

[15] Ferraty, F., Peuch, A., & Vieu, P. (2003). Modèle à indice fonctionnel simple. *Comptes Rendus Mathématique*, 336(12), 1025–1028.

[16] Ferraty, F., Sued, M., & Vieu, P. (2013). Mean estimation with data missing at random for functional covariables. *Statistics*, 47(4), 688–706.

[17] Ferraty, F., & Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer.

[18] Hall, P., & Heyde, C. C. (2014). *Martingale Limit Theory and Its Application*. Academic Press.

[19] Härdle, W., & Marron, J. S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *The Annals of Statistics*, 13, 1465–1481.

[20] Hristache, M., Juditsky, A., & Spokoiny, V. (2001). Direct estimation of the index coefficient in a single-index model. *The Annals of Statistics*, 29(3), 595–623.

[21] Khardani, S., Lemdani, M., & Ould Saïd, E. (2010). Some asymptotic properties for a smooth kernel estimator of the conditional mode under random censorship. *Journal of the Korean Statistical Society*, 39(4), 455–469.

[22] Khardani, S., Lemdani, M., & Ould Saïd, E. (2011). Uniform rate of strong consistency for a smooth kernel estimator of the conditional mode for censored time series. *Journal of Statistical Planning and Inference*, 141(11), 3426–3436.

[23] Khardani, S., Lemdani, M., & Ould Saïd, E. (2012). On the strong uniform consistency of the mode estimator for censored time series. *Metrika*, 75(2), 229–241.

[24] Laib, N., & Louani, D. (2010). Nonparametric kernel regression estimation for functional stationary ergodic data: Asymptotic properties. *Journal of Multivariate Analysis*, 101(10), 2266–2281.

[25] Laib, N., & Louani, D. (2011). Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. *Journal of Statistical Planning and Inference*, 141(1), 359–372.

[26] Lemdani, M., Ould Saïd, E., & Poulin, N. (2009). Asymptotic properties of a conditional quantile estimator with randomly truncated data. *Journal of Multivariate Analysis*, 100(3), 546–559.

[27] Liang, H., & de Uña Álvarez, J. (2010). Asymptotic normality for estimator of conditional mode under left-truncated and dependent observations. *Metrika*, 72(1), 1–19.

[28] Liang, H., Wang, S., & Carroll, R. J. (2007). Partially linear models with missing response variables and error-prone covariates. *Biometrika*, 94(1), 185–198.

[29] Ling, N., Liang, L., & Vieu, P. (2015). Nonparametric regression estimation for functional stationary ergodic data with missing at random. *Journal of Statistical Planning and Inference*, 162, 75–87.

[30] Ling, N., Liu, Y., & Vieu, P. (2016). Conditional mode estimation for functional stationary ergodic data with responses missing at random. *Statistics*, 50(5), 991–1013.

[31] Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons.

[32] Nittner, T. (2003). Missing at random (MAR) in nonparametric regression: A simulation experiment. *Statistical Methods and Applications*, 12(2), 195–210.

[33] Ould Saïd, E., & Cai, Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. *Journal of Nonparametric Statistics*, 17(7), 797–806.

[34] Ould Saïd, E., & Djabrane, Y. (2011). Asymptotic normality of a kernel conditional quantile estimator under strong mixing hypothesis and left-truncation. *Communications in Statistics – Theory and Methods*, 40(14), 2605–2627.

[35] Ramsay, J. O., & Silverman, B. W. (1997). *Functional Data Analysis*. Springer.

[36] Ramsay, J. O., & Silverman, B. W. (2002). *Applied Functional Data Analysis*. Springer.

[37] Ramsay, J. O., & Silverman, B. W. (2005). *Functional Data Analysis* (2nd ed.). Springer.

[38] Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592.

[39] Tatachak, A., & Ould Saïd, E. (2011). A nonparametric conditional mode estimate under RLT model and strong mixing condition. *International Journal of Statistics and Economics*, 6, 76–92.

[40] Tsiatis, A. A. (2006). *Semiparametric Theory and Missing Data*. Springer.