Functional Uniform-in-Bandwidth Moderate Deviation Principle for the Local Empirical Processes Involving Functional Data

Nour-Eddine Berrahou[1*], Salim Bouzebda[2**], and Lahcen Douge[1***]

[1]*Laboratoire de Modélisation des Systèmes Complexes, University Cadi Ayyad Marrakesh, Marrakesh, Morocco*

[2]*Université de technologie de Compiègne, LMAC (Laboratory of Applied Mathematics of Compiègne), France*

Abstract—Our research employs general empirical process methods to investigate and establish moderate deviation principles for kernel-type function estimators that rely on an infinite-dimensional covariate, subject to mild regularity conditions. In doing so, we introduce a valuable moderate deviation principle for a function-indexed process, utilizing intricate exponential contiguity arguments. The primary objective of this paper is to contribute to the existing literature on functional data analysis by establishing functional moderate deviation principles for both Nadaraya–Watson and conditional distribution processes. These principles serve as fundamental tools for analyzing and understanding the behavior of these processes in the context of functional data analysis. By extending the scope of moderate deviation principles to the realm of functional data analysis, we enhance our understanding of the statistical properties and limitations of kernel-type function estimators when dealing with infinite-dimensional covariates. Our findings provide valuable insights and contribute to the advancement of statistical methodology in functional data analysis.

1. INTRODUCTION

The regression analysis has proved to be a flexible tool and provided a powerful statistical modeling framework in various applied and theoretical contexts where one intends to model the predictive relationship between related responses and predictors. It is worth noticing that the parametric regression models provide useful tools for analyzing practical data when the models are correctly specified but may suffer from large modelling biases if the structures of the models are misspecified, which is the case in many practical problems. As an alternative, nonparametric smoothing methods ease the concerns of modeling biases. Kernel nonparametric function estimation methods are popular, presenting only one of many approaches to constructing good function estimators, including nearest-neighbor, spline and wavelet methods. These methods have been applied to a wide variety of data. In the present paper, we shall focus on constructing consistent kernel-type estimators. For good sources of references to the research literature in this area along with statistical applications, consult [20, 26, 27, 30, 31, 35, 39, 51, 52, 57, 76, 78, 106, 108, 124, 126, 128, 130, 136, 138] and the references therein.

[*]E-mail: `n.berrahou@uca.ma`

[**]E-mail (Corresponding author): `salim.bouzebda@utc.fr`

[***]E-mail: `lahcen.douge@uca.ac.ma`

Recently, increasing interest has been given to regression models in which the response variable is real-valued, and the explanatory variable takes the form of smooth functions that vary randomly between repeated observations or measurements. Statistical problems related to the study of functional random variables, that is to say, variables with values in an infinite-dimensional space, have known, since the last decades, a growing interest in the statistics literature. The development of this research theme is indeed motivated by the abundance of data measured on an increasingly fine temporal/spatial grid as is the case, for instance, in meteorology, medicine, satellite imagery, and many other research areas. Thus, the statistical modeling of these data, seen as random functions, led to several challenging theoretical and numerical research questions. For an overview of theoretical as well as practical aspects of functional data analysis, the reader can refer to the monographs of [14] for linear models for random variables taking values in a Hilbert space, [120] for scalar-on-function and function-on-function linear models, functional principle component analysis, and parametric discriminant analysis. Ferraty and Vieu [62] however focused more on nonparametric methods and mainly kernel-type estimation for scalar-on-function nonlinear regression models. They extended such tools to classification and discrimination analysis. Horváth and Kokoszka [82] discussed the generalization of several interesting concepts in statistics, such as goodness-of-fit tests, portmanteau tests, and change detection, to the functional data framework. For the latest contributions in FDA and its related topics, one can refer to [1−4, 18, 25, 37, 54, 93, 103, 139]. In various scenarios, there is a keen interest in gauging the rate at which specific probabilities converge. Frequently, these probabilities exhibit rapid exponential convergence. Numerous researchers have explored large deviations and unearthed various applications, primarily within the realm of mathematical physics. To be more precise, the utilization of large deviation results in the fields of probability and statistics encompasses a wide range of applications. The primary application of statistics is in the evaluation of efficiency through the comparison of test procedures. This allows for the determination of the procedure that is most efficient based on the minimum amount of data required to achieve predefined performance levels, which are typically expressed in terms of the risk of the first kind, test power, and alternative hypotheses. For further information, refer to [112]. Additional applications arise in the assessment of estimate techniques, wherein the inaccuracy rates associated with each method are taken into account and subsequently compared. It is important to observe that large deviation results can serve as effective tools for establishing the consistency of estimators and their convergence rates. Valuable resources on the subject of large deviations can be found in [9, 49, 50, 132]. Extending beyond the conventional results of weak and strong convergence in regression analysis, the problem of functional moderate deviations introduces new challenges that these existing frameworks cannot readily address. There exists an extensive large and moderate deviation literature involving many areas of probability and statistics. We refer to the book of [49] and the references therein for an account of large deviation results and applications. In the nonparametric function estimation setting, several results have been obtained in these last years. We refer to [114], where the studies involve the Nadaraya−Watson and histogram estimates of the regression function, respectively both in the real vector case. Louani and Ould Maouloud [95] established a large deviation principle (LDP, in short) for a vector process, allowing to derive LDP's for the kernel density and regression function estimators by the contraction principle. Further results for the multivariate regression estimates are due to [105], where large together with moderate deviation principles are stated for the Nadaraya−Watson estimator as well as for the semi-recursive kernel estimator. Large deviation results for the kernel regression function estimate on a functional covariate are obtained by [95]. For more references we refer to [10−12, 42, 60, 66, 75, 84−86, 88, 104, 117, 119, 123, 125, 133].

The selection of the kernel function in our setup is mostly unconstrained, except for meeting certain mild requirements that will be provided subsequently. However, the choice of bandwidth presents a greater challenge. It is important to acknowledge that the selection of the bandwidth plays a critical role in achieving a good rate of consistency. Specifically, it significantly impacts the magnitude of bias in the estimate. Broadly speaking, our focus lies in determining bandwidth that yields an estimator exhibiting a favorable trade-off between the bias and variance of the estimators under consideration. It is more suitable to evaluate the variability of bandwidth based on the applied criteria, available data, and location, which cannot be obtained by conventional methods. For further elaboration and analysis on the subject, readers are encouraged to consult [96]. The main aim of the present paper is to establish

large deviation and moderate deviation[1] results in the functional data uniformly in the bandwidth. The uniform in bandwidth problem has attracted great attention, we refer to [15, 17, 21, 22, 25, 27, 29, 33, 34, 43, 47, 53, 59, 96, 98, 99]. To effectively tackle the challenges posed by functional moderate deviations, novel methodologies, and theoretical frameworks must be developed. Advanced statistical techniques, such as empirical process theory offer promising avenues for addressing these challenges. By extending our focus beyond the conventional regression models and embracing the complexities of functional moderate deviations, we can enhance our understanding of the behavior and limitations of kernel estimators. This, in turn, provides valuable insights into the underlying processes and patterns within functional data.

The layout of the article is as follows. Section 2 gives the notation and the definitions we need. Section 3 shows the moderate deviation principle, which is equivalent to the moderate deviation principle of the finite-dimensional distributions, given in Theorem 3.1, plus an exponential asymptotic equicontinuity condition concerning a pseudometric, given in Theorem 3.2. Section 4 provides applications of our main results including the kernel regression function estimate in Subsection 4.1, the kernel conditional distribution function in Subsection 4.2, The kernel quantile regression in Subsection 4.3, the kernel conditional density function in Subsection 4.4 and finally the kernel conditional copula function in Subsection 4.5. We discuss a bandwidth choice for practical use in Section 5. A summary of the findings highlighting remaining open issues may be found in Section 6. All proofs are deferred to Section 7. Due to the lengthiness of the proofs, we limit ourselves to the most important arguments. Finally, a few relevant technical results are given in Appendix A.

2. THE GENERAL PROCESS

We consider a sequence $\{(X_i, Y_i) : i \geq 1\}$ of i.i.d. pairs of random copies of the random element [rv] (X, Y), where X takes its values in some abstract space $\mathcal{E}$ and Y is a $\mathbb{R}^q$-valued random variable, $q \geq 1$, with density $g(\cdot)$, concerning the Lebesgue measure on $\mathbb{R}^q$. Suppose that $\mathcal{E}$ is endowed with a semi-metric $d(\cdot, \cdot)$[2] defining a topology to measure the proximity between two elements of $\mathcal{E}$ and which is disconnected from the definition of X to avoid measurability problems. This covers the case of semi-normed spaces of possibly infinite dimension (e.g., Hilbert or Banach spaces). We will consider especially the conditional expectation of $l(Y)$ given $X = x$,

$$r^l(x) := \mathbb{E}(l(Y_1)|X_1 = x), \quad x \in \mathcal{E},$$

whenever this regression function is meaningful. Here and elsewhere, $l(\cdot)$ denotes a specified measurable function from $\mathbb{R}^q$ to $\mathbb{R}$, which is assumed to be bounded on each compact sub-interval of $\mathbb{R}^q$. The general Nadaraya–Watson [107, 137] type estimator of $r^l(\cdot)$ has been introduced by [61], for $l(y) = y$. It is defined, for any fixed $x \in \mathcal{E}$, by

$$\widehat{r}_n^l(x, h) = \frac{\sum_{i=1}^n l(Y_i) K(h^{-1} d(x, X_i))}{\sum_{i=1}^n K(h^{-1} d(x, X_i))} := \frac{\widehat{r}_{n,2}^l(x, h)}{\widehat{r}_{n,1}^l(x, h)}, \tag{2.1}$$

where $K(\cdot)$ is a real-valued kernel function, h is the bandwidth parameter and, for $k = 1, 2$ and

$$\widehat{r}_{n,k}^l(x, h) = \frac{1}{n \mathbb{E}[\Delta_1(x, h)]} \sum_{i=1}^n l^{k-1}(Y_i) \Delta_i(x, h),$$

---

[1] Let us first recall the concept of large and moderate deviations. A sequence $\{Z_n, n \geq 1\}$ of $\mathbb{R}$-valued random variables is said to satisfy a large deviation principle (LDP) with speed v_n and rate function $I(\cdot)$ if for any closed set $F \subset \mathbb{R}$,

$$\limsup_{n \to \infty} v_n^{-1} \log \left(\mathbb{P}(Z_n \in F) \right) \leq -\inf_{x \in F} I(x)$$

and any open set $G \subset \mathbb{R}$,

$$\liminf_{n \to \infty} v_n^{-1} \log \left(\mathbb{P}(Z_n \in G) \right) \geq -\inf_{x \in G} I(x).$$

Let a_n be a nonrandom sequence that goes to infinity, if there exists function $c(n)$, and $(a_n (Z_n - c(n)))$ satisfies an LDP, then Z_n is said to satisfy a moderate deviation principles (MDP). Roughly speaking, the MDP for Z_n is the LDP for $(a_n (Z_n - c(n)))$.

[2] A semi-metric (sometimes called pseudo-metric) $d(\cdot, \cdot)$ is a metric which allows $d(x_1, x_2) = 0$ for some $x_1 \neq x_2$.

where $\Delta_i(x, h) = K(h^{-1}d(x, X_i))$. Notice that taking $l_A(y) = \mathbb{1}_{\{y \in A\}}$ in the statement (2.1), where A is a subset of $\mathbb{R}$, we obtain the well-known kernel estimator $\hat{\mu}(A|x)$ of the conditional empirical measure

$$\mu(A|x) := \mathbb{P}(Y_1 \in A | X = x).$$

Properties of $\hat{\mu}(A|x)$, whenever $A = (\infty, t]$ for $t \in \mathbb{R}$ and $x \in \mathbb{R}$, have been investigated by several authors among whom we cite [16, 17, 121, 122, 129]. In the functional data case, see [18, 65, 103].

The purpose of this paper is to establish some general moderate deviation results which allow us to derive, under mild regularity conditions, as a by-product the uniform functional moderate deviation principle for the kernel l-indexed regression function estimator, whenever $l(\cdot)$ belongs to an appropriate class $\mathcal{L}$. Towards this end, consider two real continuous functions, $c_l(\cdot)$ from $\mathcal{E}$ to $\mathbb{R}$, $d_l(\cdot)$ from $\mathcal{E}$ to $\mathbb{R}$ and define the following process. For any $x \in \mathcal{E}$ and $z = (l, \varrho) \in \mathcal{L} \times \mathcal{H}_0$, where $\mathcal{H}_0 = [\vartheta_1, \vartheta_2]$ and $0 < \vartheta_1 < \vartheta_2 < \infty$, set (assuming that this expression is meaningful)

$$W_n(x, z) = \frac{1}{n\mathbf{E}[\Delta_1(x, \varrho h)]} \sum_{i=1}^{n} \left\{ \left(c_l(x)l(Y_i) + d_l(x)\right)\Delta_i(x, \varrho h) - \mathbf{E}\left[\left(c_l(x)l(Y_i) + d_l(x)\right)\Delta_i(x, \varrho h)\right] \right\}$$

$$:= \frac{1}{n\mathbf{E}[\Delta_1(x, \varrho h)]} \sum_{i=1}^{n} \left\{ \mathcal{M}_{x,l}(Y_i)\Delta_i(x, \varrho h) - \mathbf{E}\left[\mathcal{M}_{x,l}(Y_i)\Delta_i(x, \varrho h)\right] \right\}. \qquad (2.2)$$

In what follows, we first establish a functional uniform moderate deviation principle for the process $\{W_n(x, z) : z \in \mathcal{L} \times \mathcal{H}_0\}$, with a fixed $x \in \mathcal{J}$, where $\mathcal{J}$ denotes a suitable subset of $\mathcal{E}$. Subsequently, through the utilization of exponential contiguity arguments, we deduce the corresponding moderate deviation principle for the regression estimate $\hat{r}_n^l(\cdot)$, encompassing the kernel distribution estimator. To provide more clarity, we present a corollary that delineates the behavior of the conditional distribution kernel estimator under scenarios of moderate deviations. This is complemented by the introduction of innovative applications of our main findings, as detailed in Section 4. Finally, we establish the moderate deviation principle for

$$\left\{ \sup_{(x,z) \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0} |W_n(x, z)| \right\}.$$

3. MAIN RESULTS

We will impose the following set of assumptions for our main results.

(A1) $K(\cdot)$ is a nonnegative bounded differentiable kernel over its support $[0, 1]$ and $K(1) > 0$. The derivative $K'(\cdot)$ of $K(\cdot)$ exists on the interval $[0, 1]$, and it is bounded and satisfies the condition $K'(t) \leq 0, \forall t \in (0, 1)$.

(A2) For each $x \in \mathcal{E}$ and a real number v, there exist a nonnegative functional $f_v(\cdot)$, a function $g_{x,v}(\cdot)$ and a nonnegative real function $\phi(\cdot)$ tending to zero, as its argument tends to 0, such that,

  (i) $F_x^v(u) := \mathbb{P}(d(x, X_1) \leq u | Y = v) = \phi(u)f_v(x) + g_{x,v}(u)$ with, uniformly in v, $g_{x,v}(u) = o(\phi(u))$ as $u \to 0$ and $g_{x,v}(u)/\phi(u)$ is almost surely bounded;

  (ii) there exists a nondecreasing bounded function $\tau_0(u)$ such that, uniformly in $u \in [0, 1]$,

$$\frac{\phi(uh)}{\phi(h)} = \tau_0(u) + o(1), \quad \text{as} \quad h \downarrow 0.$$

(A3) Let $h = h_n$ and w_n be sequences of positive numbers such that, as $n \to \infty$,

$$h_n \to 0, \quad w_n \to \infty, \quad \frac{w_n^2}{n\phi(\gamma h)} \to 0 \quad \text{for some} \quad \gamma \in [\vartheta_1, \vartheta_2].$$

(A4) For any real numbers a and b, and any $(x, l) \in \mathcal{E} \times \mathcal{L}$

$$\text{(i)} \int |l_1(v)l_2(v)|g(v)dv < \infty \quad \text{for any} \quad l_1, \, l_2 \in \mathcal{L};$$

$$\text{(ii)} \int e^{|a+bl(v)|} f_v(x)g(v)dv < \infty; \quad \text{(iii)} \int e^{|a+bl(v)|} g(v)dv < \infty,$$

Discussion of Assumptions

Condition (A1) is very usual in nonparametric estimation literature devoted to the functional data context. Notice that [115] symmetric kernel is not adequate in this context since the random process $d(x, X_i)$ is positive, therefore we consider $K(\cdot)$ with support $[0, 1]$. This is a natural generalization of the assumption usually made on the kernel in the multivariate case where $K(\cdot)$ is supposed to be a spherically symmetric density function. Because the fact that Lebesgue measure does not exist on infinite dimension space, assumptions (A2) involve the small ball techniques related to the fractal dimension used in this paper, for instance, see [100], who in turn was inspired by [68] in his nonparametric density estimation under functional observations. From [64], one can cite:

1. in the case when $\mathcal{E} = \mathbb{R}^d$, $\mathbb{P}(d(x, X_1) \leq u | Y = v) \approx C(d)u^d f_v(x)$, where $C(d)$ is the volume of the unit ball in $\mathbb{R}^d$;

2. $\mathbb{P}(d(x, X_1) \leq u | Y = v) \approx f_v(x)u^\gamma$ for some $\gamma > 0$, then $\tau_0(u) = u^\gamma$;

3. $\mathbb{P}(d(x, X_1) \leq u | Y = v) \approx f_v(x)u^\gamma \exp\{-c/u^\kappa\}$ for some $\gamma > 0$ and $\kappa > 0$, then $\tau_0(u) = \delta_1(u)$, where $\delta_1(\cdot)$ is a Dirac function.

Masry [100] explains that if $\mathcal{E} = \mathbb{R}$, then the condition coincides with the fundamental axioms of probability calculus, furthermore if $\mathcal{E}$ is an infinitely dimensional Hilbert space then $\phi(h)$ can decrease toward 0 by an exponential speed as $n \to \infty$. (A2)(ii) approximately shows that the small ball probability can be written approximately as the product of two independent functions, refer to [101] for the diffusion process, [13] for a Gaussian measure, [91] for a general Gaussian process and has employed these assumptions [100] for strongly mixing processes. For example, the function $\phi(\cdot)$ can be expressed as $\phi(\epsilon) = \epsilon^\delta \exp(-C/\epsilon^a)$ with $\delta \geq 0$ and $a \geq 0$, and it corresponds to the Ornstein–Uhlenbeck and general diffusion processes (for such processes, $a = 2$ and $\delta = 0$) and the fractal processes (for such processes, $\delta > 0$ and $a = 0$). This class of processes also satisfies condition (A2). For other examples, we refer to [24, 28, 64, 128]. Since $n\phi(h) \to \infty$, suppose $\phi(h) = n^{-c}$ for some $0 < c < 1$. Then Condition (A3) is satisfied provided $w_n = n^\gamma$, for $0 < \gamma < (1-c)/2$. Assumptions (A4) are set on to ensure the needed properties of finiteness and differentiability of the finite-dimensional moment generating function associated with the process $\{W_n(x, z) : z \in \mathcal{L} \times \mathcal{H}_0\}$. All these general assumptions are sufficiently weak relative to the different objects involved in the statement of our main results. They cover and exploit the principal axes of this contribution, which are the topological structure of the functional variables, and the probability measure in this functional space. From now on, consider the following covariance function defined, for any $z_1 := (l_1, \varrho_1) \in \mathcal{L} \times \mathcal{H}_0$, $z_2 := (l_2, \varrho_2) \in \mathcal{L} \times \mathcal{H}_0$ and $\gamma \in \mathcal{H}_0$,

$$R_\gamma(x, z_1, z_2) := \frac{\alpha_1(\varrho_1, \varrho_2)\tau(\varrho, \gamma)}{\alpha_0^2 \tau(\varrho_1, \gamma)\tau(\varrho_2, \gamma)} \frac{\int \mathcal{M}_{x,l_1}(v)\mathcal{M}_{x,l_2}(v)f_v(x)g(v)\,dv}{\left(\int f_v(x)g(v)\,dv\right)^2}, \tag{3.1}$$

where

$$\alpha_0 = K(1) - \int_0^1 K'(u)\tau_0(u)\,du \tag{3.2}$$

and

$$\alpha_1(\varrho_1, \varrho_2) = K\left(\frac{\varrho}{\varrho_1}\right) K\left(\frac{\varrho}{\varrho_2}\right) - \int_0^1 \left(K\left(\frac{\varrho}{\varrho_1}u\right) K\left(\frac{\varrho}{\varrho_2}u\right)\right)' \tau_0(u) du$$

with $\varrho = \min(\varrho_1, \varrho_2)$ and

$$\tau(a, b) = \tau_0\left(\frac{a}{b}\right) \mathbb{1}_{\{a \leq b\}} + \left[\tau_0\left(\frac{b}{a}\right)\right]^{-1} \mathbb{1}_{\{a > b\}},$$

where $\mathbb{1}_A$ denotes the indicator of A. Note that, $\tau(a, b) = (\tau(b, a))^{-1}$ which gives that $R_\gamma(x, z_1, z_2) = R_\gamma(x, z_2, z_1)$. Let $\{\Xi_\gamma(x, z) : z \in \mathcal{L} \times \mathcal{H}_0\}$ be a mean-zero Gaussian process such that, for any $(z_1, z_2) \in \left(\mathcal{L} \times \mathcal{H}_0\right)^2$,

$$\mathbf{E}[\Xi_\gamma(x, z_1)\Xi_\gamma(x, z_2)] = R_\gamma(x, z_1, z_2).$$

Let $\mathscr{Z}_{x,\gamma}$ be the closed linear subspace of the space L_2, generated by

$$\left\{\Xi_\gamma(x, z) : z \in \mathcal{L} \times \mathcal{H}_0\right\}.$$

Define the function $\varphi : \mathscr{Z}_{x,\gamma} \to l_\infty\left(\mathcal{L} \times \mathcal{H}_0\right)$ by

$$\varphi(\xi)(z) = \mathbf{E}[\Xi_\gamma(x, z)\xi].$$

Note that the reproducing kernel Hilbert space is associated with the covariance function $R_\gamma(x, z_1, z_2)$ is the Hilbert space $\{\varphi(\xi) : \xi \in \mathscr{Z}_{x,\gamma}\}$ equipped with the inner product

$$\langle \varphi(\xi_1), \varphi(\xi_2) \rangle = \mathbf{E}[\xi_1 \xi_2].$$

For any $(z_1, \ldots, z_m) \in \mathcal{L} \times \mathcal{H}_0$ and any $\lambda_1, \ldots, \lambda_m \in \mathbb{R}$, set

$$\Gamma_{x,z_1 \ldots, z_m}^\gamma(\lambda_1, \ldots, \lambda_m) = \inf\left\{2^{-1}\mathbf{E}[\xi^2] : \xi \in \mathscr{Z}_{x,\gamma}, \; \varphi(\xi)(z_j) = \lambda_j, 1 \leq j \leq m\right\}. \tag{3.3}$$

The following theorem gives a finite-dimensional moderate deviation principle for the process $\{W_n(x, z) : z \in \mathcal{L} \times \mathcal{H}_0\}$. Let, for any $x \in \mathcal{E}$,

$$f(x) = \int f_v(x)g(v)\, dv.$$

Theorem 3.1. *Assume that the assumptions* (A1)–(A4) *are fulfilled. If* $f(x) > 0$ *and* $\tau_0(\frac{\vartheta_1}{\vartheta_2}) > 0$. *Then the random sequence* $w_n(W_n(x, z_1), \ldots, W_n(x, z_m))$ *satisfies a LDP with the speed* $n\phi(\gamma h)/w_n^2$ $(\gamma \in \mathcal{H}_0)$ *and the good rate function* $\Gamma_{x,z_1 \ldots, z_m}^\gamma(\cdot)$ *defined in* (3.3), *where* $z_i = (l_i, \varrho_i)$, *for* $i = 1, \ldots, m$.

The proof of Theorem 3.1 is postponed until Section 7.

Remark 3.1. For any $z := (l, \varrho) \in \mathcal{L} \times \mathcal{H}_0$, it follows by Theorem 3.1 that the random sequence $w_n W_n(x, z)$ satisfies a LDP with the speed $n\phi(\gamma h)/w_n^2$ $(\gamma \in \mathcal{H}_0)$ and the good rate function $\Gamma_{x,z}^\gamma(\cdot)$ given by

$$\Gamma_{x,z}^\gamma(\lambda) = \sup_a \left\{\lambda a - \frac{1}{2}a^2 R_\gamma(x, z, z)\right\} = \frac{\lambda^2}{2R_\gamma(x, z, z)},$$

where

$$R_\gamma(x, z, z) := \frac{\left(K^2(1) - \int_0^1 \left(K^2(u)\right)' \tau_0(u)\, du\right)}{\alpha_0^2 \tau(\varrho, \gamma)} \frac{\int \mathcal{M}_{x,l}^2(v) f_v(x)g(v)\, dv}{\left(\int f_v(x)g(v)\, dv\right)^2}. \tag{3.4}$$

Remark 3.2. If we suppose that the derivative of the function $\tau_0(\cdot)$ exists and considering the fact that $\tau_0(0) = 0$ and $\tau_0(1) = 1$, then, integrating by parts, we obtain

$$R_\gamma(x, z, z) = \frac{1}{\tau(\varrho, \gamma)} \frac{\int\limits_0^1 K^2(u)\tau_0'(u)du \int \mathcal{M}_{x,l}^2(v)f_v(x)g(v)dv}{\left[\int\limits_0^1 K(u)\tau_0'(u)du\right]^2 \left(\int f_v(x)g(v)dv\right)^2},$$

which gives a simpler form of the rate function. Also, we observe, whenever $K(u) = \mathbb{1}_{[0,1]}(u)$, that

$$R_\gamma(x, z, z) = \frac{1}{\tau(\varrho, \gamma)} \frac{\int \mathcal{M}_{x,l}^2(v)f_v(x)g(v)dv}{\left(\int f_v(x)g(v)dv\right)^2}.$$

In the sequel, we investigate the functional moderate deviation principle of the process

$$\left\{W_n(x, z) : z \in \mathcal{L} \times \mathcal{H}_0\right\}$$

in the space $l_\infty(\mathcal{L} \times \mathcal{H}_0)$ equipped with the uniform topology.

Let $L(\cdot)$ denote the finite-valued measurable envelope function of the class $\mathcal{L}$ of measurable functions on $\mathbb{R}$, that is,

$$L(y) \geq \sup_{l \in \mathcal{L}} |l(y)|, \quad y \in \mathbb{R}.$$

Define

$$N(\epsilon, \mathcal{L}) = \sup_Q \mathcal{N}\left(\epsilon\sqrt{Q(L^2)}, \mathcal{L}, d_Q\right),$$

where the supremum is taken over all probability measures Q on $\mathbb{R}$, for

$$0 < Q(L^2) = \int L^2(y)dQ(y) < \infty,$$

and d_Q is the $L_2(Q)$-metric. As usual, $N(\epsilon, \mathcal{L}, d_Q)$ is the minimal number of balls $\{l : d_Q(l, l') < \epsilon\}$ of d_Q-radius ϵ needed to cover $\mathcal{L}$. To formulate our functional moderate deviation principle, we consider some additional conditions.

(B1)   (i) For some $C' > 0$ and $\nu_2 > 0$,

$$N(\epsilon, \mathcal{L}) \leq C'\epsilon^{-\nu_2}, \quad 0 < \epsilon < 1;$$

  (ii) $\mathcal{L}$ is a pointwise measurable class, that is, there exists a countable subclass $\mathcal{L}_0$ of $\mathcal{L}$ such that, for any function $l \in \mathcal{L}$, we can find a sequence of function $\{l_n\}$ in $\mathcal{L}_0$ for which

$$l_n(y) \to l(y), \quad y \in \mathbb{R}.$$

(B2) The class

$$\mathcal{K} = \left\{x \mapsto K\left(\frac{d(x, u)}{h}\right) : u \in \mathcal{J}, \ h > 0\right\}$$

satisfies the Condition (B1).

(B3) Uniformly in $v \in \mathbb{R}$, the function $f_v(.)$ is continuous and strictly positive on $\mathcal{J}$.

As in [48], let us denote by $\{\mathcal{M}(x) : x \geqslant 0\}$ a continuous, increasing and non-negative function fulfilling, for some $q > 2$, ultimately as $x \uparrow \infty$,

$$(i)\; x^{-q}\mathcal{M}(x) \uparrow, \quad (ii)\; x^{-1} \log \mathcal{M}(x) \downarrow,$$

where '$\uparrow$' (resp. '$\downarrow$') stands for non-decreasing (resp. non-increasing). For each $t \geqslant \mathcal{M}(0)$, we denote by $\mathcal{M}^{\mathrm{inv}}(t)$ the uniquely defined non-negative number such that $\mathcal{M}\left(\mathcal{M}^{\mathrm{inv}}(t)\right) = t$. The following choices of $\mathcal{M}(\cdot)$ are of particular interest:

(i) $\mathcal{M}(x) = x^p$ for some $p > 2$;

(ii) $\mathcal{M}(x) = \exp(sx)$ for some $s > 0$.

(B4) (i) For some $t > \max(s, 1)$,

$$\mathbf{E}\left[\exp\left(tL(Y)\right)\right] < \infty;$$

(ii)

$$\lim_{n\to\infty} \frac{w_n^2 \max\left(\mathcal{M}^{\mathrm{inv}}(n), \log(n)\right)}{n\phi(\gamma h)} = \infty;$$

(iii)

$$\lim_{n\to\infty} \frac{n\phi(\gamma h)\mathcal{M}^{\mathrm{inv}}(n)^{-2}}{\max\left(\log\left(\mathcal{M}^{\mathrm{inv}}(n)\right), \log\left(1/\phi(\gamma h)\right)\right)} = \infty \quad \text{and} \quad \limsup_{n\to\infty} \frac{w_n^2 \log\left(\mathcal{M}^{\mathrm{inv}}(n)\right)}{n\phi(\gamma h)} < \infty;$$

(iv)

$$\limsup_{n\to\infty} \frac{w_n^2 \log\left(1/\phi(\gamma h)\right)}{n\phi(\gamma h)} = 0.$$

(B′4) (i) There exists a constant $L_0 > 0$ such that $L(Y)\mathbb{1}_{\{X \in \mathcal{J}\}} \leq L_0$ a.s.;

(ii)

$$\limsup_{n\to\infty} \frac{w_n^2 \log(1/\phi(\gamma h))}{n\phi(\gamma h)} = 0 \quad \text{and} \quad \limsup_{n\to\infty} \frac{n\phi(\gamma h)}{\log(1/\phi(\gamma h))} = \infty.$$

Comments on Additional Hypotheses

For Assumption (B1)(i), see [116, Examples 26 and 38], [113, Lemma 22], [55, Subsection 4.7.], [131, Theorem 2.6.7], [89, Subsection 9.1] provide a number of sufficient conditions under which (B1)(i) holds, we may also refer to [46, 15, 16, 27, Subsection 3.2] for further discussions. For instance, it is satisfied, for general $d \geq 1$, whenever $l(\mathbf{x}) = \Psi(p(\mathbf{x}))$, with $p(\mathbf{x})$ is either a polynomial in d variables or the αth power of the absolute value of a real polynomial for some $\alpha > 0$ and $\Psi(\cdot)$ is a real-valued function of bounded variation, which covers commonly used kernels, such as Gaussian, Epanečnikov, Uniform, etc, we refer the reader to [59, p. 1381]. We also mention that Condition (B1)(i) is satisfied whenever that class of functions contains functions of bounded variation on $\mathbb{R}^q$ (in the sense of Hardy and Kauser ([80, 90, 134]), see, e.g., [44, 81, 111, 135]). Assumption (B1)(ii) is made to avoid measurability difficulties. Our definition of "pointwise measurability" is borrowed from example 2.3.4 in [131], [72, p. 262] calls a pointwise measurable function class satisfying the pointwise countable approximation property. This condition is discussed in [131, Example 2.3.4, p. 110] and [89, Subsection 8.2, p. 110] and it is satisfied whenever $l(\cdot)$ is right continuous. Assumption (B1)(i) ensures that $\mathcal{L}$ is VC type with characteristics C' and ν. Condition (B2) in [73] is formulated as follows:

(B′2) $K(x) > 0$, is a bounded and compactly supported measurable function that belongs to the linear span (the set of finite linear combinations) of functions $k(x) \geq 0$ satisfying the following property: the subgraph of $k(\cdot)$, $\{(s, u) : k(s) \geq u\}$, can be represented as a finite number of Boolean operations among sets of the form

$$\left\{ (s, u) : p(s, u) \geq \varphi(u) \right\},$$

where p is a polynomial on $\mathbb{R} \times \mathbb{R}$ and φ is an arbitrary real function.

Indeed, for a fixed polynomial p, the family of sets

$$\left\{ \{(s, u) : p((s - t)/h, u) \geq \varphi(u)\} : t \in \mathbb{R}, h > 0 \right\}$$

is contained in the family of positivity sets of a finite-dimensional space of functions, and then the entropy bound follows by Theorems 4.2.1 and 4.2.4 in [55]. Our results are demonstrated through an examination of the bounded scenario, where the Condition (B′4) is imposed, and the unbounded scenarios, which are explored under the Condition (B4). To establish the result of Theorem 3.2, we may relax and replace the Assumption (B4)(iv) by the following assumption:

$$\limsup_{n \to \infty} \frac{w_n^2 \log(1/\phi(\gamma h))}{n \phi(\gamma h)} < \infty.$$

A function $\pi : T \to T$ is said to be a finite partition function of a set T if, for each $t \in T$, $\pi(\pi(t)) = \pi(t)$ and the cardinality of $\{\pi(t) : t \in T\}$ is finite. Let $\pi(T) = \{t_1, \ldots, t_m\}$ and $A_j = \{t \in T : \pi(t) = t_j\}$ for $1 \leq j \leq m$; then $\{A_1, \ldots, A_m\}$ is a partition of the set T. Note that finite partition functions can be used to characterize the compactness of $l_\infty(T)$. A set B of $l_\infty(T)$ is compact if and only if it is closed and bounded and if for each $\tau > 0$ there exists a finite partition function $\pi : T \to T$ such that

$$\sup_{\psi \in B} \left| \psi(t) - \psi(\pi(t)) \right| \leq \tau,$$

for instance, see (Theorem IV.5.6 in [56] or [6], p. 573). We also have that if B is a compact set of $l_\infty(T)$, then B is a set of uniformly bounded and uniformly equicontinuous functions in the pseudo-metric space (T, d_B), where

$$d_B(t_1, t_2) = \sup_{\psi \in B} |\psi(t_1) - \psi(t_2)|.$$

From now on, for any real-valued function ψ defined on a set T, we use the notation

$$\|\psi\|_T := \sup_{t \in T} |\psi(t)|.$$

For future use, let us introduce two classes of continuous and bounded functions on $\mathcal{J}$ indexed by $\mathcal{L}$

$$\mathcal{C} := \left\{ c_l : l \in \mathcal{L} \right\} \quad \text{and} \quad \mathcal{D} := \left\{ d_l : l \in \mathcal{L} \right\}.$$

We shall always assume that the classes $\mathcal{C}$ and $\mathcal{D}$ are compact with respect to the uniform topology. Let us define

$$C_{\mathcal{L}} := \sup \left\{ \|c_l\|_{\mathcal{J}} : l \in \mathcal{L} \right\} \quad \text{and} \quad D_{\mathcal{L}} := \sup \left\{ \|d_l\|_{\mathcal{J}} : l \in \mathcal{L} \right\}.$$

For any $x \in \mathcal{J}$, the moderate deviation principle for the process $\{W_n(x, z) : z \in \mathcal{L} \times \mathcal{H}_0\}$ in the space $l_\infty(\mathcal{L} \times \mathcal{H}_0)$ is presented in the following theorem.

Theorem 3.2. *Assume that Assumptions* (A1)–(A4), (B1)–(B3), *and* (B4) *or* (B′4) *hold true, and* $\tau_0(\frac{\vartheta_1}{\vartheta_2}) > 0$. *Furthermore, consider the classes of continuous functions* $\mathcal{C}$ *and* $\mathcal{D}$ *given above. Then we have, for any* $\gamma \in \mathcal{H}_0$ *and any* $x \in \mathcal{J}$,

(i) *for any* $0 < c < \infty$, $\{\psi \in l_\infty(\mathcal{L} \times \mathcal{H}_0) : I_x^\gamma(\psi) \leq c\}$ *is a compact set of* $l_\infty(\mathcal{L} \times \mathcal{H}_0)$;

(ii) *for all open subsets O of $l_\infty(\mathcal{L} \times \mathcal{H}_0)$,*

$$\liminf_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\left\{w_n W_n(x,\cdot) \in O\right\}\right) \geq -\inf_{\psi\in O} I_x^\gamma(\psi)$$

and for all closed subsets F of $l_\infty(\mathcal{L} \times \mathcal{H}_0)$,

$$\limsup_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\left\{w_n W_n(x,\cdot) \in F\right\}\right) \leq -\inf_{\psi\in F} I_x^\gamma(\psi),$$

where

$$I_x^\gamma(\psi) = \inf\left\{2^{-1}\mathbf{E}[\xi^2] : \xi \in \mathscr{Z}_{x,\gamma},\ \varphi(\xi) = \psi\right\}. \tag{3.5}$$

The proof of Theorem 3.2 is postponed until Section 7.

Remark 3.3. Making use of similar arguments as in the paper [8, p. 5], it follows, whenever $\sup\limits_{z\in\mathcal{L}\times\mathcal{H}_0} R_\gamma(x,z,z) > 0$, that for any $\lambda \geq 0$,

$$\inf_{\{\psi\in l_\infty(\mathcal{L}\times\mathcal{H}_0):\sup_{z\in\mathcal{L}\times\mathcal{H}_0}|\psi(z)|\geq\lambda\}} I_x^\gamma(\psi) = \frac{\lambda^2}{2\sup\limits_{z\in\mathcal{L}\times\mathcal{H}_0} R_\gamma(x,z,z)}$$

and, whenever $\sup_{z\in\mathcal{L}\times\mathcal{H}_0} R_\gamma(x,z,z) = 0$,

$$I_x^\gamma(\psi) = \begin{cases} 0, & \text{if } \sup\limits_{z\in\mathcal{L}\times\mathcal{H}_0} |\psi(z)| = 0 \\ \infty, & \text{if } \sup\limits_{z\in\mathcal{L}\times\mathcal{H}_0} |\psi(z)| > 0. \end{cases}$$

Therefore, by Theorem 3.2, for any $\lambda \geq 0$, we have

$$\lim_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\mathbb{P}\left(w_n \sup_{z\in\mathcal{L}\times\mathcal{H}_0} |W_n(x,z)| \geq \lambda\right)$$

$$= \begin{cases} -\dfrac{\lambda^2}{2\sup\limits_{z\in\mathcal{L}\times\mathcal{H}_0} R_\gamma(x,z,z)} & \text{if } \sup\limits_{z\in\mathcal{L}\times\mathcal{H}_0} R_\gamma(x,z,z) > 0 \\ -\infty & \text{if } \sup\limits_{z\in\mathcal{L}\times\mathcal{H}_0} R_\gamma(x,z,z) = 0 \end{cases}$$

$$=: -I_x^\gamma(\lambda).$$

The uniform moderate deviations principle on $\mathcal{J} \times \mathcal{L} \times \mathcal{H}_0$ is presented in the following theorem. Towards this end, for any $\epsilon > 0$, consider the following number

$$\mathcal{N}(\epsilon, \mathcal{J}, d) = \min\left\{n : \text{there exist } x_1, \ldots, x_n \text{ in } \mathcal{J} \text{ such that for any } x \in \mathcal{J} \right.$$

$$\left. \text{there exists } 1 \leq k \leq n \text{ such that } d(x, x_k) < \epsilon\right\},$$

which is the minimal number of open ball with d-radius ϵ needed to cover the subset $\mathcal{J}$. We assume that $\mathcal{J}$ satisfies the following property.

(J) For any $\epsilon > 0$, there exists $C > 0$ and $\nu > 0$ such that $\mathcal{N}(\epsilon, \mathcal{J}, d) \leq C\epsilon^{-\nu}$.

Theorem 3.3. *If assumptions of Theorem* 3.2 *and assumption* (J) *hold, then for any* $\lambda > 0$,

$$\lim_{n \to \infty} \frac{w_n^2}{n\phi(\gamma h)} \log \left(\mathbb{P} \left\{ \sup_{(x,z) \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0} w_n |W_n(x,z)| > \lambda \right\} \right) = -I^{\gamma}(\lambda),$$

where

$$I^{\gamma}(\lambda) = \inf_{x \in \mathcal{J}} I_x^{\gamma}(\lambda).$$

The proof of Theorem 3.3 is postponed until Section 7.

Remark 3.4. As in the [97] bootstrap, see also [19, 36, 127] for recent references. Following [46], we introduce an auxiliary i.i.d. sequence $Z = Z_1, Z_2, \ldots$ of real-valued rv's, independent of $\{(X_i, Y_i) : i \geq 1\}$, and such that

(R1) $\mathbb{E}(Z) = 1$; $\mathbb{E}(Z^2) = 2$;

(R2) for some $\epsilon > 0$, $\mathbb{E}(e^{tZ}) < \infty$ for all $|t| \leq \epsilon$.

Setting $T_n = Z_1 + \cdots + Z_n$ we define the $\{\mathfrak{W}_{i,n} : 1 \leq i \leq n\}$, by setting, for $i = 1, \ldots, n$

$$\mathfrak{W}_{i,n} = \begin{cases} \dfrac{Z_i}{T_n} = \dfrac{Z_i}{\displaystyle\sum_{j=1}^{n} Z_j} & \text{when} \quad T_n > 0 \\[4mm] \dfrac{1}{n} & \text{when} \quad T_n \leq 0. \end{cases}$$

Introduce the resampled version of the process (2.2) given by

$$W_n^*(x,z) = \frac{1}{n\mathbf{E}[\Delta_1(x,\varrho h)]} \sum_{i=1}^{n} \left\{ \left(c_l(x)l(Y_i)Z_i + d_l(x) \right) \Delta_i(x, \varrho h) \right.$$

$$\left. - \mathbf{E}\left[\left(c_l(x)l(Y_i)Z_i + d_l(x) \right) \Delta_i(x, \varrho h) \right] \right\}. \tag{3.6}$$

Following [46], we observe that $W_n^*(x, l, h)$ reduces to a process of the form $W_n(x, l^*, h)$, for a suitable measurable $l^*(\cdot)$, and after some easy changes. Without loss of generality, we set $Z = Q(U)$ and $Z_i = Q(U_i)$ for $i = 1, \ldots, n$, where U and $U_1, \ldots, U_n$ are independent rv's, with a uniform distribution on $(0,1)$, and independent of $\{(X_i, Y_i) : 1 \leq i \leq n\}$. This allows us to define a measurable function $\psi^*(\cdot)$ on $\mathbb{R}^{q+1}$, and a rv $\mathbf{Y}^*$, by

$$\mathbf{Y}^* = \begin{bmatrix} U & Y_1 \end{bmatrix}' \in \mathbb{R}^{1+q} \quad \text{and} \quad l^*(\mathbf{Y}^*) = Q(U)l(Y) = Zl(Y).$$

Letting $(X_i, \mathbf{Y}_i^*), i = 1, 2, \ldots$, denote i.i.d. random copies of $(X, \mathbf{Y}^*)$, it is readily checked that $\{(X_i, \mathbf{Y}_i^*) : 1 \leq i \leq n\}$, and $l^*(\cdot)$ fulfill the general assumptions imposed in Theorem 3.2. Then the process

$$\left\{ w_n W_n^*(x,z) : z \in \mathcal{L} \times \mathcal{H}_0 \right\}$$

satisfies a LDP in the space $l_\infty(\mathcal{L} \times \mathcal{H}_0)$ with the speed $n\phi(\gamma h)/w_n^2$ and the good rate function $I_x^{\gamma}(\cdot)$. Then we have, for any $\gamma \in \mathcal{H}_0$ and any $x \in \mathcal{J}$,

(i) for any $0 < c < \infty$, $\{\psi \in l_\infty(\mathcal{L} \times \mathcal{H}_0) : I_x^{\gamma}(\psi) \leq c\}$ is a compact set of $l_\infty(\mathcal{L} \times \mathcal{H}_0)$;

(ii) for all open subsets O of $l_\infty(\mathcal{L} \times \mathcal{H}_0)$,

$$\liminf_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\Big\{w_n W_n^*(x,\cdot) \in O\Big\}\right) \geq -\inf_{\psi \in O} I^\gamma(\psi)$$

and for all closed subsets F of $l_\infty(\mathcal{L} \times \mathcal{H}_0)$,

$$\limsup_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\Big\{w_n W_n^*(x,\cdot) \in F\Big\}\right) \leq -\inf_{\psi \in F} I_x^\gamma(\psi),$$

where we recall

$$I_x^\gamma(\psi) = \inf\left\{2^{-1}\mathbf{E}[\xi^2] : \xi \in \mathcal{Z}_\gamma,\ \varphi(\xi) = \psi\right\}. \tag{3.7}$$

By Theorem 3.3 we have, for any $\lambda > 0$,

$$\lim_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\Big\{\sup_{(x,z)\in\mathcal{J}\times\mathcal{L}\times\mathcal{H}_0} w_n |W_n^*(x,z)| > \lambda\Big\}\right) = -I^\gamma(\lambda),$$

while it is possible that the last result also holds for the exchangeably weighted bootstrap, such a determination is beyond the scope of this paper and appears to be quite difficult.

Remark 3.1. According to [64], our methodology is heavily dependent on the distribution function $\phi(\cdot)$. This is evident in our conditions (via the function $\tau_0(\cdot)$) and in the convergence rates of our estimate (via the asymptotic behavior of the quantity $n\phi(h)$). More precisely, the behavior of $\phi(\cdot)$ around 0 turns out to be of paramount importance. Thus, the tiny ball probabilities of the underlying functional variable X are crucial. In probability theory, the calculation of the quantity $\mathbb{P}(\|X - x\| < s)$ for "small" s (i.e., for s tending toward zero) and for a fixed x is known as the "small ball problem." Unfortunately, there are solutions for very few random variables (or processes) X even when $x = 0$. In several functional spaces, taking $x \neq 0$ results in formidable obstacles that may be insurmountable. Typically, authors emphasize Gaussian random variables. We refer you to [91] for a summary of the key findings regarding the probability of small balls. If X is a Gaussian random element on the separable Banach space $\mathcal{E}$ and x belongs to the reproducing kernel Hilbert space associated with X, then the following well-known result holds:

$$\mathbb{P}(\|X - x\| \leq s) \sim C_x \mathbb{P}(\|X\| \leq s), \quad \text{as} \quad s \to 0.$$

As far as we know, the results that are available in the published literature are basically all of the forms

$$\mathbb{P}(\|X - x\| < s) \sim c_x s^{-\alpha} \exp\left(-C/s^\beta\right),$$

where α, β, c_χ, and C are positive constants and $\|\cdot\|$ may be a supremum norm, a L^p norm or a Besov norm. The interested reader can refer to [62–64, 128] for more discussion. Notice that the pioneer book by [62] extensively comments on the links between nonparametric functional statistics small-ball probability theory and topological structure on the functional space $\mathcal{E}$.

4. APPLICATIONS

While only the examples provided below will be discussed, they serve as archetypes for various functionals and can be explored similarly.

4.1. The Kernel Regression Function Estimate

Some further conditions are needed to establish the functional moderate deviations principle for the kernel regression function estimate $\widehat{r}_n^l(x)$.

(C1)   (i) For each $(x, x') \in \mathcal{J}^2, l \in \mathcal{L}$, and some constants $\beta > 0$ and $\varsigma > 0$

$$|r^l(x) - r^l(x')| \leq \varsigma d(x, x')^\beta;$$

(ii) $w_n h^\beta \to 0$ as $n \to \infty$.

Assumption (C1)(i) imposes some smoothness of the regression operator. Define the function $I^\gamma_{x,1}(\cdot)$ as the function $I^\gamma_x(\cdot)$ in the statement (3.5), with $c_l(x) = 1$ and $d_l(x) = -r^l(x)$ for any $x \in \mathcal{J}$. The large deviation principle for the process $\{w_n(\widehat{r}^l_n(x, \varrho h) - r^l(x)) : (l, \varrho) \in \mathcal{L} \times \mathcal{H}_0\}$ in the space $l_\infty(\mathcal{L} \times \mathcal{H}_0)$ is presented in following corollary.

Corollary 4.1. *Under the assumption of Theorem* 3.2*, assume that the Conditions* (C1) *hold. Then the process*

$$\left\{ w_n(\widehat{r}^l_n(x, \varrho h) - r^l(x)) : (l, \varrho) \in \mathcal{L} \times \mathcal{H}_0 \right\}$$

satisfies a LDP *in the space* $l_\infty(\mathcal{L} \times \mathcal{H}_0)$ *with the speed* $n\phi(\gamma h)/w^2_n$ *and the good rate function* $I^\gamma_{x,1}(\cdot)$.

The proof of Corollary 4.1 is postponed until Section 7.

Proposition 4.1. *Under the assumption of Theorem* 3.2*, assume that the Conditions* (C1) *hold. Then, for any $\delta > 0$, we have*

$$\lim_{n \to \infty} \frac{w^2_n}{n\phi(\gamma h)} \log \mathbb{P}\left(\exists x \in \mathcal{E}, \quad r^l(x) \notin \left[\widehat{r}^l_n(x, h) - \delta \frac{\hat{\sigma}_n}{w_n}, \widehat{r}^l_n(x, h) + \delta \frac{\hat{\sigma}_n}{w_n} \right] \right) = -\frac{\delta^2}{2}.$$

Moreover, the sequence of sets of functions

$$D_n = \left\{ g : \mathcal{E} \to \mathbb{R}, \left| g(x) - \widehat{r}^l_n(x, h) \right| \leq \delta \frac{\hat{\sigma}_n}{w_n}, \quad \forall x \in \mathcal{E} \right\},$$

is an asymptotic almost sure sequence of confidence regions of $r^l(x)$, here $\hat{\sigma}^2_n := \hat{\sigma}^2_n(x)$ is any consistent estimate of the $\widehat{r}^l_n(x, h)$'s variance.

Remark 4.2. Let $|| \cdot ||$ be a norm on $\mathcal{E} = \mathbb{R}^d$. Denote by $B(x, r)$ the set of all points $z \in \mathbb{R}^d$ satisfying $||x - z|| \leq r$. For each $n \geq 1$ and $k \in \{1, \dots, n\}$, the k-nearest neighbor bandwidth at x is denoted by $\hat{k}_{n,x}$ and defined as the smallest radius $r \geq 0$ such that the ball $B(x, r)$ contains at least k points from the collection $\{X_1, \dots, X_n\}$, i.e.,

$$\hat{k}_{n,x} = \inf \left\{ r \geq 0 : \sum_{i=1}^n \mathbb{1}_{B(x,r)}(X_i) \geq k \right\}.$$

The k-nearest neighbor estimate of the regression function, $x \mapsto \mathbb{E}[l(Y)|X = x]$, is defined as, for all $x \in \mathbb{R}^d$,

$$\sum_{i=1}^n l(Y_i) \mathbb{1}_{B(x,\hat{k}_{n,x})}(X_i) / \sum_{i=1}^n \mathbb{1}_{B(x,\hat{k}_{n,x})}(X_i).$$

This estimate is an adaptive bandwidth version of the Nadaraya−Watson estimate which here would be defined in the same way except that a non-random bandwidth (depending only on n, e.g., $n^{-1/5}$) is used in place of $\hat{k}_{n,x}$. It would be of interest to investigate the process defined in (2.2) in the k-nearest neighbor setting.

4.2. The Kernel Conditional Distribution Function

To present the functional moderate deviations principle for the conditional distribution function estimate $\widehat{F}_{nh}(t|x) := \widehat{\mu}((\infty, t]|x)$ for $(x, t) \in \mathcal{J} \times \mathbb{R}$ as a special case, the Conditions (C1) have to be formulated for the conditional distribution function $F(t|x) := \mu((\infty, t]|x)$ for $(x, t) \in \mathcal{J} \times \mathbb{R}$ as follows.

(C2) (i) For any $(x, x') \in \mathcal{J}^2$, any $(t, t') \in \mathbb{R}^{2q}$, some $\beta_1 > 0$ $\beta_2 > 0$ and a constant $\varsigma' > 0$

$$|F(t|x) - F(t'|x')| \leq \varsigma' \left(d(x, x')^{\beta_1} + ||t - t'||^{\beta_2} \right);$$

(ii) $w_n h^{\beta_1} \to 0$ as $n \to \infty$.

Assumption $(C2)(i)$ introduces a level of smoothness to the conditional distribution. Define the function $I_{2,x}^{\gamma}(\cdot)$ as the function $I_x^{\gamma}(\cdot)$ in statement (3.5), where $l(y) = \mathbb{1}_{(\infty,t]}(y)$, $c_l(x) = 1$, and $d_l(x) = -F(t|x)$ for any $(x,t) \in \mathcal{J} \times \mathbb{R}$. Let the conditional distribution function estimate be defined as follows

$$\widehat{F}_{n,h}(t|x) = \frac{\sum_{i=1}^{n} \mathbb{1}_{(\infty,t]}(Y_i) K(h^{-1} d(x,X_i))}{\sum_{i=1}^{n} K(h^{-1} d(x,X_i))}. \tag{4.1}$$

The large deviation principle for the process $\{w_n(\widehat{F}_{n,\varrho h}(t|x) - F(t|x)) : (t,\varrho) \times \mathbb{R} \times \mathcal{H}_0\}$ in $l_\infty(\mathbb{R} \times \mathcal{H}_0)$ is presented in the following corollary.

Corollary 4.2. *Assume that assumptions* (A1)–(A5), (B2)–(B3), *and* (C2) *hold true. Then the process*

$$\left\{ w_n(\widehat{F}_{n,\varrho h}(t|x) - F(t|x)) : (t,\varrho) \in \mathbb{R} \times \mathcal{H}_0 \right\}$$

satisfies a large deviations in $l_\infty(\mathbb{R} \times \mathcal{H}_0)$ *with the speed* $n\phi(\gamma h)/w_n^2$ *and the good rate function* $I_{2,x}^{\gamma}(\cdot)$.

The proof of Corollary 4.2 is postponed until Section 7.

4.3. The Kernel Quantile Regression

For a given $\alpha \in (0,1)$, the αth-order conditional quantile of the distribution of a real-valued Y given $X = x$ is defined as

$$q_\alpha(x) = \inf\left\{ y \in \mathbb{R} : F(y|x) \geq \alpha \right\}.$$

Notice that, whenever $F(\cdot|x)$ is strictly increasing and continuous in a neighborhood of $q_\alpha(x)$, the function $F(\cdot|x)$ has a unique quantile of order α at a point $q_\alpha(x)$, that is $F(q_\alpha(x)|x) = \alpha$. In such case

$$q_\alpha(x) = F^{-1}(\alpha|x) = \inf\left\{ y \in \mathbb{R} : F(y|x) \geq \alpha \right\},$$

which may be estimated uniquely by

$$\widehat{q}_{n,\alpha}(x,\varrho) = \widehat{F}_{n,\varrho h}^{-1}(\alpha|x).$$

Conditional quantiles have been widely studied in the literature when the predictor X is of finite dimension, see for instance, [45]. Let us first recall some conceptions of Hadamard differentiability [70, 89, 131]. Let $\mathcal{X}$ and $\mathcal{Y}$ be two metrizable topological linear spaces. A map Φ defined on a subset $\mathcal{D}_\Phi$ of $\mathcal{X}$ with values in $\mathcal{Y}$ is called Hadamard differentiable at x if there exists a continuous mapping $\Phi_x' : \mathcal{X} \mapsto \mathcal{Y}$ such that

$$\lim_{n \to \infty} \frac{\Phi(x + t_n \nu_n) - \Phi(x)}{t_n} = \Phi_x'(h)$$

holds for all sequences t_n converging to $0+$ and ν_n converging to ν in $\mathcal{X}$ such that $x + t_n \nu_n \in \mathcal{D}_\Phi$ for every n.

Corollary 4.3. *Let* $0 < p < q < 1$ *be fixed and let* $F(\cdot|x)$ *be a conditional distribution function with continuous and positive derivative* $f(\cdot|x)$ *on the interval* $\left[F^{-1}(p|x) - \varepsilon, F^{-1}(q|x) + \varepsilon \right]$ *for some* $\varepsilon > 0$. *Then, under conditions of Corollary* 4.2, *the process* $\{w_n(\widehat{q}_{n,\alpha}(x,\varrho) - q_\alpha(x))\}$ *satisfies the LDP in* $l_\infty([p,q] \times \mathcal{H}_0)$ *with speed* $n\phi(\gamma h)/w_n^2$ *and rate function* $I_{\gamma,x}^{EQ}$ *given by*

$$I_{\gamma,x}^{EQ}(\phi) = \inf\left\{ I_{2,x}^{\gamma}(\psi) : -\frac{\psi\left(F^{-1}(t|x)\right)}{f\left(F^{-1}(t|x)\right)} = \phi(x) \text{ for all } t \in [p,q] \right\}.$$

The proof of Corollary 4.3 is postponed until Section 7.

Remark 4.3. It will be interesting to extend our findings to the following settings.

1. (Expectile regression.) For $p \in (0,1)$, let $l(T - \boldsymbol{\theta}) = (p - \{T - \boldsymbol{\theta} \leq 0\})|T - \boldsymbol{\theta}|$, then the zero of $r^l(\cdot)$ with respect to $\boldsymbol{\theta}$ leads to quantities called expectiles by [110]. Expectiles, as defined by [110], may be introduced either as a generalization of the mean or as an alternative to quantiles. Indeed, classical regression provides us with a high sensitivity to extreme values, allowing for more reactive risk management. Quantile regression, on the other hand, provides the ability to acquire exhaustive information on the effect of the explanatory variable on the response variable by examining its conditional distribution, refer to [102, 103] for further details on expectiles in functional data setting.

2. (Conditional winsorized mean.) As in [83], if we consider $l(T - \boldsymbol{\theta}) = -k, T - \boldsymbol{\theta}, k$ if $T - \boldsymbol{\theta} < -k$, $|T - \boldsymbol{\theta}| \leq k$, or $T - \boldsymbol{\theta} > k$, then the zero of $r^l(\cdot)$ with respect to $\boldsymbol{\theta}$ will be the conditional winsorized mean. Notably, this parameter was not considered in the literature on nonparametric functional data analysis involving wavelet estimators. Our paper offers asymptotic results for the conditional winsorized mean when the covariates are functions.

4.4. The Kernel Conditional Density Function

By setting $l(\cdot) = \frac{1}{h_1} K_1(h_1^{-1}(\cdot - t))$, for $t \in \mathbb{R}$, h_1 is a bandwidth parameter and $K_1(\cdot)$ is the kernel function, into (2.1), we obtain the kernel estimator of the conditional density function $f(t|x)$ given by

$$\widehat{f}_{n,bh}(t|x) = \frac{\sum_{i=1}^n \frac{1}{h_1} K_1(h_1^{-1}(Y_i - t)) K(bh^{-1}d(x, X_i))}{\sum_{i=1}^n K(bh^{-1}d(x, X_i))}. \tag{4.2}$$

(C3)   (i) For any $(x, x') \in \mathcal{J}^2$, any $(t, t') \in [a, b]^2 \subset \mathbb{R}^2$, some $\beta_1 > 0$, $\beta_2 > 0$, and a constant $\varrho > 0$

$$|f(t|x) - f(t'|x')| \leq \varrho \left(d(x, x')^{\beta_1} + |t - t'|^{\beta_2} \right);$$

(ii) $w_n(h^{\beta_1} + h_1^{\beta_2}) \to 0$ as $n \to \infty$.

(B″2) The class

$$\tilde{\mathcal{K}} = \left\{ (h, x, y) \mapsto K\left(\frac{y - t}{h_1}\right) K\left(\frac{d(x, x')}{h}\right) : x' \in \mathcal{J}, t \in \mathbb{R}, h > 0 \right\}$$

satisfies the Condition (B1).

Assumption (C3)(i) imposes some smoothness of the conditional density. Define the function $I_{3,x}^\gamma(\cdot)$ as the function $I_x^\gamma(\cdot)$ in the statement (3.5), with $l(\cdot) = \frac{1}{h_1} K_1(h_1^{-1}(\cdot - t))$, $c_l(x) = 1$ and $d_l(x) = -f(t|x)$ for any $(x, t) \in \mathcal{J} \times [a, b]$.

Corollary 4.4. *Assume that assumptions* (A1)–(A5), (B″2)–(B4), *and* (C3) *hold. Then the process*

$$\left\{ w_n(\widehat{f}_{n,\varrho h}(t|x) - f(t|x)) : (t, \varrho) \in [a, b] \times \mathcal{H}_0 \right\}$$

satisfies large deviations in $l_\infty([a, b] \times \mathcal{H}_0)$ with the speed $nh_1\phi(\gamma h)/w_n^2$ and the good rate function $I_{3,x}^\gamma(\cdot)$.

The proof of Corollary 4.4 is similar to the proof of Corollary 4.2 and therefore omitted.

4.5. The Kernel Conditional Copula Function

Let us recall the setting of [69]. Assume that $(X_1, Y_{11}, Y_{21}), \ldots, (X_n, Y_{1n}, Y_{2n})$ is a sample of n independent and identically distributed triples of random variables. The random variables Y_{1i} and Y_{2i} are real and the X_i's are random elements. Suppose that the conditional distribution of $(Y_1, Y_2)^\top$ given $X = x$ exists and denote the corresponding conditional joint distribution function by

$$H_x(y_1, y_2) = \mathbb{P}(Y_1 \le y_1, Y_2 \le y_2 | X = x).$$

If the marginals of $H_x(\cdot, \cdot)$ denoted as

$$F_{1x}(y_1) = \mathbb{P}(Y_1 \le y_1 | X = x), \quad F_{2x}(y_2) = \mathbb{P}(Y_2 \le y_2 | X = x)$$

are continuous, then according to Sklar's theorem (see e.g., [109]) there exists a unique copula $C_x(\cdot, \cdot)$ which equals

$$C_x(u_1, u_2) = H_x\left(F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\right),$$

where

$$F_{1x}^{-1}(u) = \inf\{y : F_{1x}(y) \ge u\}$$

is the conditional quantile function of Y_1 given $X = x$ and $F_{2x}^{-1}(\cdot)$ is the conditional quantile function of Y_2 given $X = x$. The conditional copula $C_x(\cdot, \cdot)$ fully describes the conditional dependence structure of $(Y_1, Y_2)^\top$ given $X = x$. An estimator of the joint conditional distribution function $H_x(\cdot, \cdot)$ is

$$H_{xh}(y_1, y_2) = \frac{\sum_{i=1}^n \mathbb{1}\{Y_{1i} \le y_1, Y_{2i} \le y_2\} K(h^{-1}d(x, X_i))}{\sum_{i=1}^n K(h^{-1}d(x, X_i))}. \tag{4.3}$$

Then analogously as in [69] and [23] one can suggest the following empirical estimator of the copula $C_x(\cdot, \cdot)$

$$C_{x,h}(u_1, u_2) = H_{xh}\left(F_{1xh}^{-1}(u_1), F_{2xh}^{-1}(u_2)\right), \quad 0 \le u_1, \quad u_2 \le 1,$$

where $F_{1xh}(\cdot)$ and $F_{2xh}(\cdot)$ are the corresponding marginal distribution functions of $H_{xh}(\cdot, \cdot)$, i.e., $F_{1xh}(y_1) = H_{xh}(y_1, +\infty)$ and $F_{2xh}(y_2) = H_{xh}(+\infty, y_2)$.

Corollary 4.5. *Let $0 < p < q < 1$ be fixed. Suppose that $F_{1x}(\cdot)$ and $F_{2x}(\cdot)$ are continuously differentiable on the intervals $\left[F_{1x}^{-1}(p) - \varepsilon, F_{1x}^{-1}(q) + \varepsilon\right]$ and $\left[F_{2x}^{-1}(p) - \varepsilon, F_{2x}^{-1}(q) + \varepsilon\right]$ with strictly positive derivatives $f_1(\cdot|x)$ and $f_2(\cdot|x)$, respectively, for some $\varepsilon > 0$. Furthermore, assume that $\partial H_x/\partial y_1$ and $\partial H_x/\partial y_2$ exist and are continuous on the product intervals. Then, under conditions of Corollary 4.2, the process*

$$\left\{w_n(C_{x,\varrho h}(u_1, u_2) - C_x(u_1, u_2)) : ((u_1, u_2), \varrho) \in [p, q]^2 \times \mathcal{H}_0\right\}$$

satisfies the LDP in $l_\infty\left([p, q]^2 \times \mathcal{H}_0\right)$ with speed $n\phi(\gamma h)/w_n^2$ and rate function $I_{\gamma,x}^C(\cdot)$ defined by

$$I_{\gamma,x}^C(\phi) = \inf\left\{I_{1,x}^\gamma(\varpi) : \Phi_H'(\varpi) = \phi\right\},$$

where

$$\Phi_H'(\varpi)(u_1, u_2) = \varpi\left(F_{1x}^{-1}(u_1), F_{2x}^{-1}(u_2)\right)$$

$$- \frac{\partial H_x}{\partial y_1}\left(F_{1,x}^{-1}(u_1), F_{2,x}^{-1}(u_2)\right) \frac{\varpi\left(F_{1,x}^{-1}(u_1), \infty\right)}{f_1\left(F_{1,x}^{-1}(u_1)|x\right)}$$

$$- \frac{\partial H_x}{\partial y_1}\left(F_{1,x}^{-1}(u_1), F_{2,x}^{-1}(u_2)\right) \frac{\varpi\left(F_{2,x}^{-1}(u_2), \infty\right)}{f_2\left(F_{2,x}^{-1}(u_2)|x\right)}.$$

Corollary 4.5 is a direct consequence Corollary 4.2, Lemma 3.9.28 of [131] (or [32]) and Theorem 3.1 of [67] in a similar way as in Theorem 4.6 of the last mentioned reference.

Remark 4.4. We define the conditional hazard function on $\mathbb{R}$ by

$$S(\cdot|x) = \frac{f(\cdot|x)}{1 - F(\cdot|x)}.$$

The kernel estimator $S_{n;h_n}(y|x)$ is defined for all $y \in \mathbb{R}$ such that $F(y|x) < 1$, by

$$S_{n;h_n}(t|\mathbf{x}) = \frac{\widehat{f}_n(t|x)}{1 - \widehat{F}_n(t|x)}.$$

Our result can be applied to $S_{n;h_n}(t|\mathbf{x})$ by combining Corollaries 4.2 and 4.4.

Remark 4.5. The use of the single index model has been adopted to decrease the dimensionality of the explanatory variable, aiming to circumvent the "curse of dimensionality" while preserving the benefits of nonparametric smoothing in multivariate regression over the past few decades. By assuming $\mathcal{X}$ is a Hilbert space, in the single index setting, the process in (2.2) takes the form

$$W_n(x, l, \theta, \vartheta) = \frac{1}{n\mathbf{E}[\Delta_1(x, \theta, \vartheta h)]} \sum_{i=1}^{n} \left\{ \left(c_l(x)l(Y_i) + d_l(x) \right) \Delta_i(x, \theta, \vartheta h) \right.$$
$$\left. - \mathbf{E}\left[\left(c_l(x)l(Y_i) + d_l(x) \right) \Delta_i(x, \theta, \vartheta h) \right] \right\}, \tag{4.4}$$

where θ is a functional single index valued in a subset Θ of a separable Hilbert space $\mathcal{X}$ and $\Delta_i(x, \theta, h) = K(h^{-1}|\langle X - x, \theta\rangle|)$, one can refer to [37, 38]. Although it is conceivable that our findings may apply to the single index model, establishing such a conclusion is outside the purview of this paper and seems to pose significant challenges.

Remark 4.6. This study holds significance in the realm of functional data analysis. Firstly, the results presented in this paper are enriched by an additional uniformity constraint, specifically $a_n \leq h \leq b_n$. Secondly, the scope of applications is broadened to encompass novel areas in the field, including kernel quantile regression, kernel conditional density function, and kernel conditional copula function. These extensions represent pioneering contributions in functional data analysis. The findings of this study play a crucial role in establishing uniform consistency for various estimators, employing data-driven bandwidths, associated with the aforementioned applications. Further insights and relevant results can be explored in [41] and [94].

5. THE BANDWIDTH SELECTION CRITERION

Numerous methods have been developed to construct asymptotically optimal bandwidth selection rules for nonparametric kernel estimators, particularly for the Nadaraya−Watson regression estimator. Prominent works in this regard include [25, 30, 77, 79, 118]. The selection of this parameter is crucial, whether in the standard finite-dimensional case or within the infinite-dimensional framework, to ensure effective practical performance. Let's define the leave-out-(X_i, Y_i) estimator for the regression function

$$\widehat{r}_n^{l;j}(x) = \frac{\sum_{i=1, i\neq j}^{n} l(Y_i)K(h^{-1}d(x, X_i))}{\sum_{i=1, i\neq j}^{n} K(h^{-1}d(x, X_i))}. \tag{5.1}$$

To minimize the quadratic loss function, we introduce the following criterion, where $\mathbb{W}(\cdot)$ is a known non-negative weight function

$$CV(\varphi, h_n) := \frac{1}{n} \sum_{j=1}^{n} \left(l(Y_j) - \widehat{r}_n^{l;j}(X_i) \right)^2 \mathbb{W}(X_j). \tag{5.2}$$

Following the ideas developed by [118], a natural way for choosing the bandwidth is to minimize the precedent criterion, so let's choose $\widehat{h}_n \in [a_n, b_n]$ minimizing among $h \in [a_n, b_n]$:

$$CV(\varphi, h_n).$$

One can replace (5.2) by

$$\widehat{CV}\left(\varphi, h\right) := \frac{1}{n} \sum_{j=1}^{n} \left(l\left(Y_j\right) - \widehat{r}_n^{l;j}(X_i)\right)^2 \widehat{\mathcal{W}}\left(\mathbf{x}_j, \mathbf{x}\right). \tag{5.3}$$

In practice, one takes for $j = 1, \ldots, n$, the uniform global weights $\mathbb{W}\left(\mathbf{x}_j\right) = 1$, and the local weights

$$\widehat{\mathcal{W}}(\mathbf{x}_j, \mathbf{x}) = \begin{cases} 1 & \text{if} \quad d(x_j, \mathbf{x}) \leq h \\ 0 & \text{otherwise.} \end{cases}$$

By similar reasoning, one can select $\widehat{h}_n$ for the process defined in (2.2). Let us define

$$\widehat{W}_n(x, l) = W_n(x, l, \widehat{h}_n) = \frac{1}{n\mathbf{E}[\Delta_1(x, \widehat{h}_n)]} \sum_{i=1}^{n} \left\{ \left(c_l(x)l(Y_i) + d_l(x)\right) \Delta_i(x, \widehat{h}_n) \right.$$

$$\left. - \mathbf{E}\left[\left(c_l(x)l(Y_i) + d_l(x)\right)(x, \widehat{h}_n)\right] \right\}. \tag{5.4}$$

The following corollary is an immediate consequence of Theorem 3.2.

Corollary 5.1. *Assume that assumptions* (A1)–(A4), (B1)–(B3), *and* (B4) *or* (B′4) *hold true. Furthermore, consider the classes of continuous functions* $\mathcal{C}$ *and* $\mathcal{D}$ *given above. Then we have, for any* $x \in \mathcal{J}$,

(i) *for any* $0 < c < \infty$, $\{\psi \in l_\infty(\mathcal{L}) : I_x(\psi) \leq c\}$ *is a compact set of* $l_\infty(\mathcal{L})$;

(ii) *for all open subsets* O *of* $l_\infty(\mathcal{L})$,

$$\liminf_{n \to \infty} \frac{w_n^2}{n\phi(\widehat{h}_n)} \log\left(\mathbb{P}\left\{w_n\widehat{W}_n(x, \cdot) \in O\right\}\right) \geq -\inf_{\psi \in O} I_x(\psi)$$

and for all closed subsets F *of* $l_\infty(\mathcal{L})$,

$$\limsup_{n \to \infty} \frac{w_n^2}{n\phi(\widehat{h}_n)} \log\left(\mathbb{P}\left\{w_n\widehat{W}_n(x, \cdot) \in F\right\}\right) \leq -\inf_{\psi \in F} I_x(\psi),$$

where

$$I_x(\psi) = \inf\left\{2^{-1}\mathbf{E}[\xi^2] : \xi \in \tilde{\mathcal{Z}}_x, \; \varphi(\xi) = \psi\right\}, \tag{5.5}$$

and $\tilde{\mathcal{Z}}_x$ *is the closed linear subspace of the space* L_2 *generated by the mean-zero Gaussian process* $\left\{\Xi(x, l) : l \in \mathcal{L}\right\}$.

As in [87], let us minimize the following errors:

$$\text{Err}_1\left(\widehat{f}_{n,b}, f\right) = \iint \left\{\widehat{f}_{n,b}(y|x) - f(y|x)\right\}^2 W_1(x)W_2(y)d\mathbb{P}(x, y),$$

$$\text{Err}_2\left(\widehat{f}_{n,b}, f\right) = \frac{1}{n} \sum_{i=1}^{n} \left\{\widehat{f}_{n,b}(Y_i|X_i) - f(Y_i|X_i)\right\}^2 \frac{W_1\left(X_i\right)W_2\left(Y_i\right)}{f(Y_i|X_i)},$$

or

$$\text{Err}_3\left(\widehat{f}_{n,b}, f\right) = \iint \mathbf{E}\left\{\widehat{f}_{n,b}(y|x) - f(y|x)\right\}^2 W_1(x)W_2(y)d\mathbb{P}(x, y),$$

where $W_1(\cdot)$ and $W_2(\cdot)$ are some non-negative weight functions. These theoretical errors are not computable in practice and the following leave-one-out cross-validation criterion can be constructed to approximate them in some fully data-driven way:

$$\mathrm{CV}(b) = \frac{1}{n}\sum_{i=1}^{n} W_1\left(X_i\right) \int \left(\widehat{f}_{n,b}^{-i}(y|X_i)\right)^2 W_2(y)d\mathbb{P}(y) - \frac{2}{n}\sum_{i=1}^{n} \widehat{f}_{n,b}^{-i}(Y_i|X_i)W_1\left(X_i\right)W_2\left(Y_i\right),$$

where

$$\widehat{f}_{n,b}^{-i}(t|X_i) = \frac{\sum_{j=1,j\neq i}^{n} \frac{1}{h_1}K_1(h_1^{-1}(Y_j - t))K(h^{-1}d(x, X_j))}{\sum_{j=1,i\neq j}^{n} K(b^{-1}d(x, X_j))}. \tag{5.6}$$

Then the smoothing parameter b is selected by the following procedure:

$$\widehat{h} = \arg\min_{a_n \leq h \leq b_n} \mathrm{CV}(h).$$

While the aforementioned cross-validation procedures focus on approximating quadratic errors of estimation, alternative approaches for selecting smoothing parameters may prioritize optimizing the predictive power of the method. This can be achieved by minimizing one of the following prediction criteria

$$\widetilde{h}^{(1)} = \arg\min_{a_n \leq h \leq b_n} \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i^{(1)}\right)^2,$$

where the prediction is performed using either of the conditional median

$$\widehat{F}_{nh}^{-i}\left(\frac{1}{2}|x, h\right) = \frac{\sum_{j=1,j\neq i}^{n} \mathbb{1}_{(\infty,t]}(Y_j)K(h^{-1}d(x, X_j))}{\sum_{j=1,j\neq i}^{n} K(h^{-1}d(x, X_j))} \text{ and } \widehat{Y}_i^{(1)} = \left\{\widehat{F}_{nh}^{-i}\right\}^{-1}\left(\frac{1}{2}|X_i, h\right) \tag{5.7}$$

or

$$\widetilde{h}^{(2)} = \arg\min_{a_n \leq b \leq b_n} \sum_{i=1}^{n} \left(Y_i - \widehat{Y}_i^{(2)}\right)^2,$$

using the conditional mode, viz.

$$\widehat{Y}_i^{(2)} = \arg\max_{y} \widehat{f}_{n,b}^{-i}(y|X_i).$$

For more discussion, one can refer to [16, 17, 29].

Remark 5.2. It is essential to highlight that the primary challenge in employing an estimator like the one in (2.1) lies in properly selecting the smoothing parameter h. The consistency results with uniformity in bandwidth, as presented in Corollary 4.1, indicate that the choice of h_1 and h_2 within certain intervals guarantees the moderate deviations principle for $\widehat{r}_n^l(x, h)$. In other words, the fluctuations in bandwidth within a small interval do not impact the moderate deviations principle of the nonparametric estimator $\widehat{r}_n^l(x, h)$ for $r^l(x)$.

Remark 5.3. It is straightforward to modify the proofs of our results to show that it remains true when the entropy condition is substituted by the bracketing condition.[3] For some $C_0 > 0$ and $v_0 > 0$,

$$\mathcal{N}_{[\,]}\left(\mathcal{F}, L_2(\mathbb{P}), \epsilon\right) \leq C_0\epsilon^{-v_0}, \quad 0 < \epsilon < 1.$$

Remark 5.4. Observe that the standardizing factor is $(n\phi(h_n))^{1/2}$ with $\phi(h_n) \to 0$, indicating a lower rate of convergence. This is the cost incurred when concluding the conditional (local) quantities.

---

[3] Given two functions l and u, the interval $[l, u]$ represents the set of all functions f such that $l \leq f \leq u$. An ε-bracket is defined as $[l, u]$ with $||u - l|| < \varepsilon$. The bracketing number $N_{[\,]}(\mathcal{F}, ||\cdot||, \varepsilon)$ corresponds to the minimum number of ε-brackets required to encompass the class $\mathcal{F}$. The entropy with bracketing is expressed as the logarithm of the bracketing number. It's important to note that, in the definition of the bracketing number, the upper and lower bounds u and l of the brackets need not be part of $\mathcal{F}$ itself, but they are assumed to have finite norms, refer to Definition 2.1.6 in [131].

6. CONCLUSIONS

We employ general empirical process methods to establish, under mild regularity conditions, the functional moderate deviations of kernel-type function estimators that are dependent on an infinite-dimensional covariate. We present a valuable moderate deviation principle for a function-indexed process by leveraging intricate exponential contiguity arguments. The moderate deviation principles are a useful tool in analyzing the behavior of the estimators in question. This paper aims to make several noteworthy contributions to the existing literature on functional data analysis. Specifically, we focus on establishing functional moderate deviation principles for the Nadaraya−Watson estimators, the conditional distribution processes, the kernel quantile regression, the kernel conditional density function, and the kernel conditional copula function. Our findings extend the current knowledge in the field and offer new avenues for future research in functional data analysis. Extending our work to encompass k-nearest neighbors estimators holds a significant interest. However, achieving this goal requires the development of new technical arguments, as it presently lies beyond reasonable expectations. Exploring the realm of k-nearest neighbors estimators would expand the scope of our research and provide valuable insights into their performance and properties. Additional extensions involve models about dimension reduction. While our paper presents fully nonparametric functional models, recent FDA literature has emphasized the interest in semi-parametric models as a bridge between flexible (but excessively dimensional) and low dimensional (but excessively restrictive) linear models. This field comprises, for example, functional single index models (refer to [37, 38, 74] for the most recent advancements), projection pursuit models (refer to [40]), and partial linear models (refer to [5, 92]). As far as we are aware, the majority of the literature proposed in these models relates to moderate deviations principals, and we hypothesize that our ideas and methodologies could likely be used successfully for deriving such results, thereby opening up new research avenues. Such an extension would require innovative approaches and advanced theoretical frameworks to effectively tackle the challenges associated with projection pursuit regression and projection pursuit conditional distribution processes. By embarking on this path, we can further enrich the existing literature and contribute to advancing functional data analysis. Finally, the extensions of our ideas would concern dependent statistical samples with possible time series applications. The literature on dependent kernel functional estimators has been rather developed (cf. [24, 28, 62, 100, 128]) but always with moderate deviation results. Note that this extension should be harder to get than the previous ones because one main difficulty in developing such a dependent extension of our work would be the statement of new probabilistic results since those used herein (cf. results in [6, 8]) are specific to iid samples. In conclusion, our ongoing exploration involves extending our ideas to encompass dependent statistical samples, with potential applications in time series analysis. The existing body of work on dependent kernel functional estimators has seen significant development, as documented in sources such as [24, 28, 62, 100]. However, it is worth noting that these developments have generally been investigated without moderate deviation results. It is important to acknowledge that extending our research in this direction presents a more formidable challenge compared to our previous endeavors. The primary obstacle lies in the necessity of formulating novel probabilistic results, as the ones employed in our current work, as demonstrated in [6−7], are tailored specifically for independent and identically distributed (i.i.d.) samples.

7. PROOFS

In this section, we present the proofs for all the theoretical findings outlined in this study. The previously introduced notation will be consistently applied throughout the ensuing discussion. We now explore slightly more generalized processes compared to those defined in (2.2). Specifically, we offer a broader framework that does not necessitate $\mathcal{M}_{x,l}(\cdot) = c_l(x)l(\cdot) + d_l(x)$. The proof of Theorems 3.1 and 3.2 is intricate and will be dissected into multiple lemmas, each elucidated in Section 8.

Proof of Theorem 3.1

The proof utilizes the Gärtner−Ellis theorem as a principal tool. For any integer $k \geq 1$ and arbitrary $(x, l, \varrho) \in \mathcal{E} \times \mathcal{L} \times \mathcal{H}_0$, it is asserted that, for any nonnegative real function $M_{x,l}(\cdot)$ defined over the domain $\mathbb{R}$, one may discern the following

$$\mathbf{E}[M_{x,l}(Y)\Delta_1(x, \varrho h)] = \int\limits_0^1 \int M_{x,l}(v)K(u)\, d\mathbb{P}\left(\frac{d(x, X_1)}{\varrho h} \leq u, Y \leq v\right)$$

$$= \int\limits_0^1 \int M_{x,l}(v)K(u)\, d\mathbb{P}\left(\frac{d(x,X_1)}{\varrho h} \leq u|Y=v\right)g(v)\, dv$$

$$:= \int M_{x,l}(v)\mathcal{S}_{x,\varrho}(v,h)g(v)\, dv.$$

Given the existence of $K'(\cdot)$, it follows

$$K(u) = K(0) + \int\limits_0^u K'(t)\, dt,$$

which, by using a Condition (A2)(i), implies that

$$\mathcal{S}_{x,\varrho}(v,h) = K(0)F_x^v(\varrho h) + \int\limits_0^1 \left(\int\limits_0^t K'(u)\, du\right)d\mathbb{P}\left(\frac{d(x,X_1)}{\varrho h} \leq t|Y=v\right)$$

$$= K(0)F_x^v(\varrho h) + \int\limits_0^1 K'(u)\mathbb{P}\left(u \leq \frac{d(x,X_1)}{\varrho h} \leq 1|Y=v\right)du$$

$$= K(1)F_x^v(\varrho h) - \int\limits_0^1 K'(u)F_x^v(u\varrho h)\, du$$

$$= K(1)\Big(\phi(\varrho h)f_v(x) + g_{x,v}(\varrho h)\Big) - \int\limits_0^1 K'(u)\Big(\phi(u\varrho h)f_v(x) + g_{x,v}(u\varrho h)\Big)\, du$$

$$= \phi(\varrho h)\left\{K(1)\Big(f_v(x) + \frac{g_{x,v}(\varrho h)}{\phi(\varrho h)}\Big) - \int\limits_0^1 K'(u)\frac{\phi(u\varrho h)}{\phi(\varrho h)}\Big(f_v(x) + \frac{g_{x,v}(u\varrho h)}{\phi(u\varrho h)}\Big)\, du\right\}$$

$$:= \phi(\varrho h)L_{x,\varrho}(v,h).$$

Therefore, for any $(x,l,\varrho) \in \mathcal{E} \times \mathcal{L} \times \mathcal{H}_0$, we infer

$$\mathbf{E}[M_{x,l}(Y)\Delta_1(x,\varrho h)] = \phi(\varrho h)\int L_{x,\varrho}(v,h)M_{x,l}(v)g(v)\, dv. \tag{7.1}$$

Let $y_1 := (x,l_1)$ and $y_2 := (x,l_2)$ in $\mathcal{E} \times \mathcal{L}$, and $(\varrho_1,\varrho_2) \in (\mathcal{H}_0)^2$. For any real function $\mathcal{M}_{y_1,y_2}(\cdot)$ defined on $\mathbb{R}$ such that $\mathbf{E}[|\mathcal{M}_{y_1,y_2}(Y)|] < \infty$, observe that

$$\mathbf{E}[\mathcal{M}_{y_1,y_2}(Y)\Delta_1(x,\varrho_1 h)\Delta_1(x,\varrho_2 h)]$$

$$= \int\limits_0^1 \int \mathcal{M}_{y_1,y_2}(v)K(\frac{\varrho}{\varrho_1}u)K(\frac{\varrho}{\varrho_2}u)\, d\mathbb{P}\left(\frac{d(x,X_1)}{\varrho h} \leq u|Y=v\right)g(v)\, dv$$

$$:= \int \mathcal{M}_{y_1,y_2}(v)\mathcal{S}_{x,\varrho_1,\varrho_2}(v,h)g(v)\, dv.$$

Similarly, we have

$$K\left(\frac{\varrho}{\varrho_1}u\right)K\left(\frac{\varrho}{\varrho_2}u\right) = K(0)^2 + \int\limits_0^u \left(K\left(\frac{\varrho}{\varrho_1}t\right)K\left(\frac{\varrho}{\varrho_2}t\right)\right)'\, dt,$$

which, by using the Condition (A2)(i), implies that

$$\mathcal{S}_{x,\varrho_1,\varrho_2}(v,h) = \phi(\varrho h)\left\{K\left(\frac{\varrho}{\varrho_1}\right)K\left(\frac{\varrho}{\varrho_2}\right)\left(f_v(x) + \frac{g_{x,v}(\varrho h)}{\phi(\varrho h)}\right)\right.$$

$$-\int_0^1 \left(K\left(\frac{\varrho}{\varrho_1}u\right)K\left(\frac{\varrho}{\varrho_2}u\right)\right)'\frac{\phi(u\varrho h)}{\phi(\varrho h)}\left(f_v(x)+\frac{g_{x,v}(u\varrho h)}{\phi(u\varrho h)}\right)du\right\}$$

$$:=\phi(\varrho h)\mathcal{L}_{x,\varrho_1,\varrho_2}(v,h).$$

Therefore, we have

$$\mathbf{E}[\mathcal{M}_{y_1,y_2}(Y)\Delta_1(x,\varrho_1 h)\Delta_1(x,\varrho_2 h)]=\phi(\varrho h)\int\mathcal{L}_{x,\varrho_1,\varrho_2}(v,h)\mathcal{M}_{y_1,y_2}(v)g(v)\,dv. \qquad (7.2)$$

For some $\gamma\in\mathcal{H}_0$, set $\beta_n=n\phi(\gamma h)/w_n$. For any tuple $(\theta_1,\ldots,\theta_m)\in\mathbb{R}^m$, the Laplace transform corresponding to $\beta_n(W_n(x,z_1),\ldots,W_n(x,z_m))$ is explicitly defined as follows

$$\Phi_n^{x,z_1,\ldots,z_m}(\theta_1,\ldots,\theta_m)$$

$$=\mathbf{E}\left[\exp\left\{\left\langle(\theta_1,\ldots,\theta_m),\beta_n\big(W_n(x,z_1),\ldots,W_n(x,z_m)\big)\right\rangle\right\}\right]$$

$$=\mathbf{E}\left[\exp\left\{\sum_{i=1}^n\sum_{j=1}^m\frac{\beta_n\theta_j}{n\mathbf{E}[\Delta_1(x,\varrho_j h)]}\Big(\mathcal{M}_{x,l_j}(Y_i)\Delta_i(x,\varrho_j h)-\mathbf{E}\big[\mathcal{M}_{x,l_j}(Y_i)\Delta_i(x,\varrho_j h)\big]\Big)\right\}\right]$$

$$=\left(\mathbf{E}\left[\exp\left\{\sum_{j=1}^m\frac{\beta_n\theta_j}{n\mathbf{E}[\Delta_1(x,\varrho_j h)]}\Big(\mathcal{M}_{x,l_j}(Y)\Delta_1(x,\varrho_j h)-\mathbf{E}\big[\mathcal{M}_{x,l_j}(Y)\Delta_1(x,\varrho_j h)\big]\Big)\right\}\right]\right)^n$$

$$:=\left(\varphi_n^{x,z_1,\ldots,z_m}(\theta_1,\ldots,\theta_m)\right)^n. \qquad (7.3)$$

Let us now evaluate the quantity $\varphi_n^{x,z_1,\ldots,z_m}(\theta_1,\ldots,\theta_m)$. We remark that

$$\varphi_n^{x,z_1,\ldots,z_m}(\theta_1,\ldots,\theta_m)=1+\frac{1}{2}\mathcal{T}_1+\mathcal{T}_2, \qquad (7.4)$$

where

$$\mathcal{T}_1=\mathbf{E}\left[\left\{\sum_{j=1}^m\frac{\theta_j\beta_n}{n\mathbf{E}[\Delta_1(x,\varrho_j h)]}\Big(\mathcal{M}_{x,l_j}(Y)\Delta_1(x,\varrho_j h)-\mathbf{E}[\mathcal{M}_{x,l_j}(Y)\Delta_1(x,\varrho_j h)]\Big)\right\}^2\right]$$

and

$$\mathcal{T}_2=\sum_{k=3}^\infty\frac{1}{k!}\left(\frac{\beta_n}{n}\right)^k\mathbf{E}\left[\left|\sum_{j=1}^m\frac{\theta_j}{\mathbf{E}[\Delta_1(x,\varrho_j h)]}\Big\{\mathcal{M}_{x,l_j}(Y)\Delta_1(x,\varrho_j h)-\mathbf{E}\big[\mathcal{M}_{x,l_j}(Y)\Delta_1(x,\varrho_j h)\big]\Big\}\right|^k\right].$$

Now, observe that

$$\mathcal{T}_1=\sum_{j=1}^m\sum_{p=1}^m\frac{\theta_j\theta_p\beta_n^2}{n^2\mathbf{E}[\Delta_1(x,\varrho_j h)]\mathbf{E}[\Delta_1(x,\varrho_p h)]}\mathbf{E}\left[\mathcal{M}_{x,l_j}(Y))\mathcal{M}_{x,l_p}(Y)\Delta_1(x,\varrho_j h)\Delta_1(x_p,\varrho_p h)\right]$$

$$-\sum_{j=1}^m\sum_{p=1}^m\frac{\theta_j\theta_p\beta_n^2}{n^2\mathbf{E}[\Delta_1(x,\varrho_j h)]\mathbf{E}[\Delta_1(x,\varrho_p h)]}\mathbf{E}\left[\mathcal{M}_{x,l_j}(Y)\Delta_1(x,\varrho_j h)\right]\mathbf{E}\left[\mathcal{M}_{x,l_p}(Y)\Delta_1(x,\varrho_p h)\right]$$

$$:=\frac{\beta_n^2}{n^2}(\mathcal{T}_{1,1}-\mathcal{T}_{1,2}). \qquad (7.5)$$

Let $\varrho=\min(\varrho_j,\varrho_p)$ be defined. Referring to (7.2) and under the conditions (A1)–(A2) and (A4)(ii) we deduce that

$$\lim_{n\to\infty}\frac{1}{\phi(\varrho h)}\mathbf{E}\left[\mathcal{M}_{x,l_j}(Y))\mathcal{M}_{x,l_p}(Y)\Delta_1(x,\varrho_j h)\Delta_1(x,\varrho_p h)\right]$$

$$= \alpha_1(\varrho_j, \varrho_p) \int \mathcal{M}_{x,l_j}(v)) \mathcal{M}_{x,l_p}(v) f_v(x) g(v) \, dv, \tag{7.6}$$

where

$$\alpha_1(\varrho_j, \varrho_p) = K\left(\frac{\varrho}{\varrho_j}\right) K\left(\frac{\varrho}{\varrho_p}\right) - \int\limits_0^1 \left(K\left(\frac{\varrho}{\varrho_j} u\right) K\left(\frac{\varrho}{\varrho_p} u\right) \right)' \tau_0(u) du,$$

Using the equality (7.1) with $M_{x,l}(v) = 1$, in accordance with conditions (A1)–(A2), we find

$$\lim_{n \to \infty} \frac{1}{\phi(\varrho_j h)} \mathbf{E}[\Delta_1(x, \varrho_j h)] = \left(K(1) - \int\limits_0^1 K'(u) \tau_0(u) \, du \right) \int f_v(x) g(v) \, dv$$

$$:= \alpha_0 \int f_v(x) g(v) \, dv. \tag{7.7}$$

Now, by Condition (A2)(ii), we have

$$\lim_{n \to \infty} \frac{\phi(\gamma h) \phi(\varrho h)}{\phi(\varrho_j h) \phi(\varrho_p h)} = \frac{\tau(\varrho, \gamma)}{\tau(\varrho_j, \gamma) \tau(\varrho_p, \gamma)}.$$

Therefore, we obtain

$$\lim_{n \to \infty} \phi(\gamma h) \mathcal{T}_{1,1} = \sum_{j=1}^m \sum_{p=1}^m \theta_j \theta_p \frac{\alpha_1(\varrho_j, \varrho_p) \tau(\varrho, \gamma)}{\alpha_0^2 \tau(\varrho_j, \gamma) \tau(\varrho_p, \gamma)} \cdot \frac{\int \mathcal{M}_{x,l_j}(v)) \mathcal{M}_{x,l_p}(v) f_v(x) g(v) \, dv}{\left(\int f_v(x) g(v) \, dv \right) \left(\int f_v(x) g(v) \, dv \right)}. \tag{7.8}$$

Making use of the condition (A3), we infer that

$$\lim_{n \to \infty} \frac{\beta_n^2}{n^2} \mathcal{T}_{1,1} = 0. \tag{7.9}$$

Likewise, by the condition (A3), we derive

$$\lim_{n \to \infty} \phi(\gamma h) \mathcal{T}_{1,2} = 0 \quad \text{and} \quad \lim_{n \to \infty} \frac{\beta_n^2}{n^2} \mathcal{T}_{1,2} = 0. \tag{7.10}$$

Using the C_r-inequality and the boundedness of $K(\cdot)$, we get, for some $\kappa_0 > 0$,

$$\mathcal{T}_2 \leq \sum_{j=1}^m \sum_{k=3}^\infty \frac{1}{k!} \left(\frac{2^m \beta_n}{n} \right)^k \frac{\kappa_0^k |\theta_j|^k \mathbf{E}\left[|\mathcal{M}_{x,l_j}(Y)|^k \Delta_1(x, \varrho_j h) \right]}{\left(\mathbf{E}[\Delta_1(x, \varrho_j h)] \right)^k}. \tag{7.11}$$

By (7.1), and making use of the conditions (A1)–(A2), there exists a constant $\kappa_1 > 0$ such that

$$\mathbf{E}\left[|\mathcal{M}_{x,l_j}(Y)|^k \Delta_1(x, \varrho_j h) \right]$$

$$< \kappa_1^k \phi(\varrho_j h) \left(\int |\mathcal{M}_{x,l_j}(v)|^k f_v(x) g(v) dv + \int |\mathcal{M}_{x,l_j}(v)|^k g(v) dv \right). \tag{7.12}$$

Again, by the equality (7.1) with $M_{x,l}(\cdot) = 1$, we have

$$\mathbf{E}[\Delta_1(x, \varrho_j h)] = \phi(\varrho_j h) \int L_{x,\varrho_j}(v, h) g(v) \, dv := \phi(\varrho_j h) T(x, \varrho_j h). \tag{7.13}$$

For each $j = 1, \ldots, m$, under conditions (A1)–(A2) and considering the fact that $f(x) > 0$, we obtain

$$\lim_{h \to 0} T(x, \varrho_j h) > 0. \tag{7.14}$$

For sufficiently large n and for some $\gamma \in \mathcal{H}_0$, there exists $t_0 > 0$ such that $2^m \beta_n / (n\phi(\gamma h)) = 2^m / w_n \leq t_0$. Now, by employing (7.11)–(7.14), we derive

$$\mathcal{T}_2 \leq \left(\frac{2^m \beta_n}{t_0 n\phi(\gamma h)} \right)^3 \phi(\gamma h) \sum_{j=1}^{m} \left(\tau(\varrho_j, \gamma) + o(1) \right) \left(\int \left[\exp \left\{ \frac{t_0 \kappa_0 |\theta_j \mathcal{M}_{x,l_j}(v)|}{(\tau(\varrho_j, \gamma) + o(1)) T(x, \varrho_j h)} \right\} f_v(x) \right. \right.$$
$$\left. \left. + \exp \left\{ \frac{t_0 C |\theta_j \mathcal{M}_{x,l_j}(v)|}{(\tau(\varrho_j, \gamma) + o(1)) T(x, \varrho_j h)} \right\} \right] g(v) dv \right).$$

Making use of (7.14), we readily infer

$$\mathcal{T}_2 \leq \kappa_2 \left(\frac{2^m \beta_n}{t_0 n\phi(\gamma h)} \right)^3 \phi(\gamma h) \sum_{j=1}^{m} \left(\int \left[\exp \left(t_0 \kappa_3 |\theta_j \mathcal{M}_{x,l_j}(v)| \right) f_v(x) \right. \right.$$
$$\left. \left. + \exp \left(t_0 \kappa_3 |\theta_j \mathcal{M}_{x,l_j}(v)| \right) \right] g(v) dv \right),$$

where $\kappa_2, \kappa_3 > 0$. By Conditions (A4)(ii)–(iii), we get

$$\mathcal{T}_2 = O\left(\left(\frac{\beta_n}{n\phi(\gamma h)} \right)^3 \phi(\gamma h) \right). \tag{7.15}$$

Thus, based on the Condition (A3), we deduce

$$\lim_{n \to \infty} \mathcal{T}_2 = 0 \quad \text{and} \quad \lim_{n \to \infty} \frac{n^2 \phi(\gamma h)}{\beta_n^2} \mathcal{T}_2 = 0. \tag{7.16}$$

By combining (7.3)–(7.10) with (7.15), we readily obtain

$$\lim_{n \to \infty} \frac{n\phi(\gamma h)}{\beta_n^2} \log \Phi_n^{x,z_1,\dots,z_m}(\theta_1, \dots, \theta_m) = \frac{1}{2} \sum_{j=1}^{m} \sum_{p=1}^{m} \theta_j \theta_p R_\gamma(x, z_j, z_p)$$
$$:= \Psi_\gamma^{x,z_1,\dots,z_m}(\theta_1 \dots, \theta_m),$$

where the function R_γ is defined in the Statement (3.1). Note that the Condition (A4) implies that the function $\Psi_\gamma^{x,z_1,\dots,z_m}$ is finite and differentiable everywhere. The Fenchel–Legendre transform of $\Psi_\gamma^{x,z_1,\dots,z_m}$ is given by

$$\Gamma_{x,z_1\dots,z_m}^\gamma(\lambda_1, \dots, \lambda_m) = \sup_{(\theta_1,\dots,\theta_m)} \left\{ \sum_{i=1}^{m} \lambda_j \theta_j - \Psi_\gamma^{x,z_1,\dots,z_m}(\theta_1, \dots, \theta_m) \right\}.$$

We need to establish the essential smoothness of the function $\Psi_\gamma^{x,z_1,\dots,z_m}$ and utilize the Gärtner–Ellis theorem for the proof, as outlined in [49] on page 44. Under the assumption (A4), it is evident that the interior of the set

$$D = \{ (\theta_1, \dots, \theta_m) : \Psi_\gamma^{x,z_1,\dots,z_m}(\theta_1, \dots, \theta_m) < \infty \}$$

is not empty. Furthermore, considering these conditions, the function $\Psi_\gamma^{z_1,\dots,z_m}$ is shown to be steep, establishing its essential smoothness. In conclusion, this ensures that for any $\theta_1, \dots, \theta_m \in \mathbb{R}$, the function holds the properties necessary for applying the Gärtner–Ellis theorem in our proof

$$\sum_{j,p}^{m} \theta_j \theta_p R_\gamma(x, z_j, z_p) \geq 0,$$

there exists a mean-zero Gaussian process $\{ \Xi_\gamma(x, z), z \in \mathcal{L} \times \mathcal{H}_0 \}$ such that

$$\mathbf{E}[\Xi_\gamma(x, z_1) \Xi_\gamma(x, z_2)] = R_\gamma(x, z_1, z_2) \text{ for each } z_1, z_2, \in \mathcal{L} \times \mathcal{H}_0.$$

Considering Eq. (3.3) and employing Lemma 4.1 from the work of [7], one can express the rate function as follows

$$\Gamma_{x,z_1\dots,z_m}^\gamma(\lambda_1, \dots, \lambda_m) = \inf \left\{ 2^{-1} \mathbf{E}[\xi^2] : \xi \in \mathcal{Z}_\gamma, \ \mathbf{E}[\Xi_\gamma(x, z_j)\xi] = \lambda_j \text{ for each } 1 \leq j \leq m \right\},$$

where $\mathcal{Z}_\gamma$ is defined in the introduction of this rate function [7, p. 6].   □

The proof of The demonstration of Theorem 3.2 requires several intermediary results, which we present in the following set of lemmas.

Lemma 7.1. *Under the assumptions* (A1)−(A2) *and* (B3) *we have, for any nonnegative function* $M_0(\cdot)$ *such that* $\mathbf{E}[M_0(Y)] < \infty$,

$$\lim_{n\to\infty} \sup_{(x,\varrho)\in\mathcal{J}\times\mathcal{H}_0} \left| \frac{\mathbb{E}[M_0(Y)\Delta_1(x,\varrho h)]}{\phi(\varrho h)} - \alpha_0 \int M_0(v) f_v(x) g(v)\, dv \right| = 0,$$

where α_0 *is given in* (3.2).

The proof of Lemma 7.1 is postponed until Section 8.

In the proof of Theorem 3.2 we shall apply Lemma 7.1 with $M_0(\cdot)$ belongs to a finite class $\mathcal{M}_0$ of nonnegative real valued functions such that $\mathbf{E}[M_0(Y)] < \infty$. Denote

$$\epsilon_n = \sup_{(x,\varrho)\in\mathcal{J}\times\mathcal{H}_0} \left| \frac{\mathbb{E}[M_0(Y)\Delta_1(x,\varrho h)]}{\phi(\varrho h)} - \alpha_0 \int M_0(v) f_v(x) g(v)\, dv \right|.$$

By Lemma 7.1

$$\epsilon_n \to 0 \quad \text{as} \quad n \to \infty.$$

By condition (B3), we have $\delta_f := \inf_v \inf_{x\in\mathcal{J}} f_v(x) > 0$ and therefore, for n large enough,

$$\epsilon_n \le \frac{\delta_f}{2} \min \left\{ \int M_0(v) g(v)\, dv : M_0 \in \mathcal{M}_0 \right\}.$$

Then for any $(x,\varrho) \in \mathcal{J} \times \mathcal{H}_0$ and $M_0(\cdot) \in \mathcal{M}_0$, we have

$$\frac{\phi(\varrho h)\alpha_0}{2} J(M_0) \le \mathbb{E}[M_0(Y)\Delta_1(x,\varrho h)] \le 2\phi(\varrho h)\alpha_0 J(M_0),$$

where $J(M_0) := \int M_0(v) f_v(x) g(v)\, dv$. By assumptions (A1) and (A2) we can see that $\alpha_0 > 0$. Then by condition (B4)(i) and the fact that $0 < \tau_0\left(\dfrac{\vartheta_1}{\vartheta_2}\right) \le \tau_0\left(\dfrac{\varrho}{\gamma}\right)$, we obtain, for any $(x,\varrho) \in \mathcal{J} \times \mathcal{H}_0$,

$$C_1\phi(\gamma h) \le \mathbb{E}[M_0(Y)\Delta_1(x,\varrho h)] \le C_2\phi(\gamma h), \tag{7.17}$$

where C_1 and C_2 are two strictly positive constants. For any $l \in \mathcal{L}$, set

$$\bar{\mathcal{M}}_{x,l}(v) = c_l(x)l(v)\mathbb{1}_{\{L(v)\le\mathcal{M}^{\text{inv}}(n)\}} + d_l(x). \tag{7.18}$$

Lemma 7.2. *Assuming that the conditions* (A1)−(A2), (B1), (B3), *and* (B4)(i) *or* (B′4)(i) *are satisfied, along with* $\tau_0\left(\dfrac{\vartheta_1}{\vartheta_2}\right) > 0$, *then for any given* $\epsilon > 0$, *there exists a finite subclass* $\mathcal{L}_\epsilon$ *of* $\mathcal{L}$ *such that, for sufficiently large* n, *for any* $l_1 \in \mathcal{L}$, *we have*

$$\min_{l_2\in\mathcal{L}_\epsilon} \sup_{x\in\mathcal{J}} \sup_{\varrho\in\mathcal{H}_0} \mathbf{E}\left[\left\{ \bar{\mathcal{M}}_{x,l_1}(Y) - \bar{\mathcal{M}}_{x,l_2}(Y) \right\}^2 \frac{(\Delta_1(x,\varrho h))^2}{(\mathbf{E}[\Delta_1(x,\varrho h)])^2/\phi(\gamma h)} \right] \le \epsilon. \tag{7.19}$$

The proof of Lemma 7.2 is postponed until Section 8.

Set $\epsilon > 0$ and select $n_0 > 0$ to be sufficiently large such that (7.19) holds for all $n \ge n_0$. For any $l_1, l_2 \in \mathcal{L}$, define

$$d_\mathcal{L}(l_1,l_2) = \sup_{n\ge n_0} \sup_{x\in\mathcal{J}} \sup_{\varrho\in\mathcal{H}_0} \mathbf{E}\left[\left\{ \bar{\mathcal{M}}_{x,l_1}(Y) - \bar{\mathcal{M}}_{x,l_2}(Y) \right\}^2 \frac{(\Delta_1(x,\varrho h))^2}{(\mathbf{E}[\Delta_1(x,\varrho h)])^2/\phi(\gamma h)} \right].$$

Consider any $\zeta_1, \zeta_2 > 0$ and define

$$\mathcal{F}_0(\zeta_1,\zeta_2) = \left\{ (z_1,z_2) \in \left(\mathcal{L}\times\mathcal{H}_0\right)^2 : d_\mathcal{L}(l_1,l_2) \le \zeta_1 \text{ and } |\varrho_1 - \varrho_2| \le \zeta_2 \right\},$$

where $z_i = (l_i, \varrho_i)$, $i = 1, 2$. For any $0 < \delta < 1$, let $l_n = \mathcal{N}(\delta\phi(\gamma h), \mathcal{J}, d)$.

Proposition 7.3. Under assumptions of Theorem 3.2, for any $\eta > 0$, we have, for any $x \in \mathcal{J}$

$$\lim_{(\delta, \zeta_1, \zeta_2) \to (0,0,0)} \limsup_{n \to \infty} \frac{w_n^2}{n\phi(\gamma h)}$$

$$\times \log \mathbb{P}\left(\sup_{y \in B(x, \delta\phi(\gamma h))} \sup_{(z_1, z_2) \in \mathcal{F}_0(\zeta_1, \zeta_2)} w_n |W_n(x, z_1) - W_n(y, z_2)| \geq \eta \right) = -\infty,$$

where $B(x, \delta\phi(\gamma h))$ is the open ball with the center x and the radius $\delta\phi(\gamma h)$. In order to prove Proposition 7.3, we need an exponential inequality for the empirical process. Let us introduce first some additional notation. Let $(\mathcal{X}, \mathcal{A})$ be a measurable space on which we consider a uniformly bounded collection of measurable functions $\mathcal{F}$. The class $\mathcal{F}$ is said to be a bounded measurable VC class of functions if it satisfies the Condition (B1). For any map T from $\mathcal{F}$ into $\mathbb{R}$, set

$$||T||_{\mathcal{F}} = \sup_{g \in \mathcal{F}} |T(g)|.$$

Let μ be any probability measure on $(\mathcal{X}, \mathcal{A})$ and $Pr = \prod_{i \in \mathbb{N}} \mu_i$ the probability measure product where, for $i \in \mathbb{N}$, $\mu_i = \mu$. Set $\pi_i : \mathcal{X}^{\mathbb{N}} \mapsto \mathcal{X}$, $i \in \mathbb{N}$, to be the coordinate functions. The following lemma is due to [71].

Lemma 7.4. *Let $\mathcal{F}$ be a measurable uniformly VC class of functions, and σ^2 and U be any numbers such that $\sigma^2 \geq \sup_{f \in \mathcal{F}} Var_{Pr}(f)$, $U \geq \sup_{f \in \mathcal{F}} ||f||_\infty$, and $0 < \sigma^2 \leq U/2$. Then there exist constants C and M depending only on the characteristic (C, ν) of the class $\mathcal{F}$, such that the inequality*

$$Pr\left(\left\| \sum_{i=1}^{n} g(\pi_i) - \mathbf{E}_{Pr}[g(\pi_i)] \right\|_{\mathcal{F}} > t \right)$$

$$\leq M \exp\left\{ -\frac{t}{MU} \log\left(1 + \frac{tU}{M(\sqrt{n}\sigma + U\sqrt{\log(U/\sigma)})^2} \right) \right\}, \qquad (7.20)$$

whenever

$$t \geq C\left(U \log\left(\frac{U}{\sigma} \right) + \sqrt{n}\sigma\sqrt{\log(U/\sigma)} \right).$$

The proof of Proposition 7.3 is split up into two cases, the unbounded case where the Assumption (B4) is assumed and the bounded case where we suppose that the Condition (B'4) is satisfied.

The unbounded case will follow as a consequence of a sequence of lemmas. Set, for any $(x, z) = (x, l, \varrho) \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0$

$$\tilde{\mathbf{W}}_n(x, z) = \frac{1}{n\mathbf{E}[\Delta_1(x, \varrho h)]} \sum_{i=1}^{n} \left\{ \bar{\mathcal{M}}_{x,l}(Y_i)\Delta_i(x, \varrho h) - \mathbf{E}\left[\bar{\mathcal{M}}_{x,l}(Y_i)\Delta_i(x, \varrho h) \right] \right\}.$$

We will first show that the processes

$$\left\{ w_n W_n(x, z) : z \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0 \right\} \quad \text{and} \quad \left\{ w_n \tilde{\mathbf{W}}_n(x, z) : z \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0 \right\}$$

are exponentially contiguous.

Lemma 7.5. *Under the assumptions* (A1)–(A3) *and* (B3–(B4)(ii), *for any $\eta > 0$, we have*

$$\limsup_{n \to \infty} \frac{w_n^2}{n\phi(\gamma h)} \log \mathbb{P}\left(\sup_{z \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0} w_n |W_n(x, z) - \tilde{\mathbf{W}}_n(x, z)| \geq \eta \right) = -\infty.$$

The proof of Lemma 7.5 is postponed until Section 8.

For any $z = (l, \varrho) \in \mathcal{L} \times \mathcal{H}_0$ and $x \in \mathcal{J}$, set

$$B_{n,x,z}(u, v) = \phi(\gamma h) \frac{\bar{\mathcal{M}}_{x,l}(v) K(d(u, x)/\varrho h)}{\mathbf{E}[\Delta_1(x, \varrho h)]}. \tag{7.21}$$

Lemma 7.6. *Assume that assumptions* (A1)–(A2) *and* (B1)–(B3) *hold true. Furthermore, consider the classes of continuous functions* $\mathcal{C}$ *and* $\mathcal{D}$ *given above. Then the class*

$$\mathfrak{M} := \left\{ (u, v) \mapsto B_{n,x_1,z_1}(u, v) - B_{n,x_2,z_2}(u, v) : (x_1, x_2) \in \mathcal{J}^2, (z_1, z_2) \in \left(\mathcal{L} \times \mathcal{H}_0\right)^2 \right\}$$

is a pointwise measurable class of functions with the envelope function

$$G(x, v) := C_0(C_{\mathcal{L}} L(v) + D_{\mathcal{L}}),$$

and satisfying the condition

$$N(\epsilon, \mathfrak{M}) \leq C_1 \epsilon^{-\nu}, \quad 0 < \epsilon < 1,$$

where C_0, C_1, *and* ν *are suitable positive constants.*

The proof of Lemma 7.6 is postponed until Section 8.

Lemma 7.7. *Assuming that the conditions* (A1)–(A2), (B1), (B3), *and* (B4)(i) *or* (B$'$4)(i) *are satisfied, along with* $\tau_0 \left(\dfrac{\vartheta_1}{\vartheta_2}\right) > 0$, *then, for n large enough and any* $\epsilon > 0$, *there exist* $\delta, \zeta_1, \zeta_2 > 0$ *such that for any* $(z_1, z_2) \in \mathcal{F}_0(\zeta_1, \zeta_2)$, *we have, for any* $x_1, x_2 \in \mathcal{J}$ *such that* $d(x_1, x_2) \leq \delta \phi(\gamma h)$,

$$\mathrm{Var}(B_{n,x_1,z_1}(X, Y) - B_{n,x_2,z_2}(X, Y)) \leq C \epsilon \phi(\gamma h)$$

for suitable constant $C > 0$.

The proof of Lemma 7.7 is postponed until Section 8.

Proof of Proposition 7.3

The proof uses Lemma 7.4 as a device. By Lemma 7.6 the following classes

$$\mathcal{F}_{n,x}(\delta, \zeta_1, \zeta_2) = \left\{ (u, v) \mapsto B_{n,x,z_1}(u, v) - B_{n,y,z_2}(x, v) : y \in B(x, \delta\phi(\gamma h)), (z_1, z_2) \in \mathcal{F}_0(\zeta_1, \zeta_2) \right\}$$

are measurable VC classes of functions. Now, for n large enough, take $U = C_0(C_{\mathcal{L}} + D_{\mathcal{L}})\mathcal{M}^{\mathrm{inv}}(n)$ where C_0 is as in Lemma 7.6, $\mathcal{F} = \mathcal{F}_{n,x}(\delta, \zeta_1, \zeta_2)$ and by Lemma 7.7 $\sigma^2 = C\epsilon\phi(\gamma h)$. Then, for n large enough, we have $\sigma \leq U/2$, and by the Condition (B4)(iii)–(iv), we infer

$$\sqrt{n}\sigma\bar{C} \geq U\sqrt{\log(U/\sigma)}, \quad \text{for suitable constant } \bar{C}$$

and

$$\limsup_{n \to \infty} \frac{w_n\left(U\log(U/\sigma) + \sqrt{n}\sigma\sqrt{\log(U/\sigma)}\right)}{n\phi(\gamma h)} \leq \limsup_{n \to \infty} \frac{(1 + \bar{C})w_n\sqrt{n}\sigma\sqrt{\log(U/\sigma)}}{n\phi(\gamma h)} < \infty.$$

Consequently, there exists a positive integer number n_0 such that for any $n \geq n_0$

$$\frac{n\phi(\gamma h)}{w_n} \geq \bar{C}_1\left(U\log(U/\sigma) + \sqrt{n}\sigma\sqrt{\log(U/\sigma)}\right)$$

and

$$\sqrt{n}\sigma + U\sqrt{\log(U/\sigma)} \leq (1 + \bar{C})\sqrt{nC\epsilon\phi(\gamma h)}.$$

Applying Lemma 7.4, for any $n \geq n_0$, we obtain

$$\mathbb{P}\left(\sup_{(y,z_1,z_2) \in B(x,\delta\phi(\gamma h)) \times \mathcal{F}_0(\zeta_1,\zeta_2)} w_n |\tilde{\mathbf{W}}_n(x, z_1) - \tilde{\mathbf{W}}_n(y, z_2)| \geq \eta \right)$$

$$\leq M \exp \left\{ -\frac{n\phi(\gamma h)\eta}{w_n M C_0 (C_{\mathcal{L}} + D_{\mathcal{L}}) \mathcal{M}^{\text{inv}}(n)} \log \left(1 + \frac{\eta C_0 (C_{\mathcal{L}} + D_{\mathcal{L}}) \mathcal{M}^{\text{inv}}(n)}{w_n M (1 + \bar{C})^2 C \epsilon} \right) \right\}.$$

Therefore, by (B4)(iii)−(iv), we infer

$$\limsup_{n \to \infty} \frac{w_n^2}{n\phi(\gamma h)} \log \left(\mathbb{P} \left(\sup_{(y,z_1,z_2) \in B(x,\delta\phi(\gamma h)) \times \mathcal{F}_0(\zeta_1,\zeta_2)} w_n |\tilde{\mathbf{W}}_n(x,z_1) - \tilde{\mathbf{W}}_n(y,z_2)| \geq \eta \right) \right)$$

$$\leq -\frac{\eta^2}{M^2 (1 + \bar{C})^2 C \epsilon}. \tag{7.22}$$

Letting ϵ go to 0, we prove that the process

$$\left\{ \tilde{\mathbf{W}}_n(x,z) : (x,z) \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0 \right\}$$

fulfills the results of Proposition 7.3. Finally, the same conclusion holds for the process $\{W_n(x,z) : (x,z) \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0\}$ in the unbounded case in view of Lemma 7.5. In the Bounded case, under Condition (B′4), we take

$$U = C_0 (\mathcal{C}_{\mathcal{L}} L_0 + \mathcal{D}_{\mathcal{L}}) \quad \text{and} \quad \sigma^2 = C \epsilon \phi(\gamma h).$$

The same arguments as above yield, under the Condition (B′4), the same inequality as in (7.22). This achieves the proof. □

Proof of Theorem 3.2

By combining the findings from Theorem 3.1, Proposition 7.3, and Lemma 7.2, and by applying Theorem 3.1 in [6], we can establish that the process $\{w_n W_n(x,z) : z \in \mathcal{L} \times \mathcal{H}_0\}$ satisfies the large deviation principle within $l_\infty(\mathcal{L} \times \mathcal{H}_0)$, featuring the speed $n\phi(\gamma h)/w_n^2$ and the corresponding good rate function

$$I_x^\gamma(\psi) = \sup \left\{ \Gamma_{x,z_1\ldots,z_m}^\gamma(\psi(z_1), \ldots, \psi(z_m)) : z_i \in \mathcal{L} \times \mathcal{H}_0, i = 1, \ldots, m, m \geq 1 \right\}.$$

Finally, by Theorem 4.2 in [7] this rate function can be expressed as

$$I_x^\gamma(\psi) = \inf \left\{ 2^{-1} \mathbf{E}[\xi^2] : \xi \in \mathcal{Z}_{x,\gamma}, \, \varphi(\xi) = \psi \right\}.$$

Hence the proof is complete. □

By condition (J) there exists $x_{n,1}, \ldots, x_{n,l_n}$ in $\mathcal{J}$ such that

$$\mathcal{J} \subset \bigcup_{k=1}^{l_n} B_{n,k},$$

and there exists $\nu \geq 0$ such that $l_n \leq C(\delta\phi(\gamma h))^{-\nu}$, for some suitable positive constant C. Here $B_{n,k}$ denotes the open ball with center $x_{n,k}$ and radius $\delta\phi(\gamma h)$.

Proof of Theorem 3.3

The lower bound is easy. In fact, for any $x \in \mathcal{J}$, by Theorem 3.2 we have

$$\liminf_{n \to \infty} \frac{w_n^2}{n\phi(\gamma h)} \log \left(\mathbb{P} \left\{ w_n \|W_n\|_\infty > \lambda \right\} \right) \geq \liminf_{n \to \infty} \frac{w_n^2}{n\phi(\gamma h)} \log \left(\mathbb{P} \left\{ w_n \sup_{z \in \mathcal{L} \times \mathcal{H}_0} |W_n(x,z)| > \lambda \right\} \right)$$

$$\geq -I_x^\gamma(\lambda).$$

Hence

$$\liminf_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\Big\{w_n||W_n||_\infty > \lambda\Big\}\right) \geq -I^\gamma(\lambda).$$

Now we show the upper bound. By considering the condition (J) and applying Proposition 7.3 we obtain from the Condition $(B4)(iv)$ and inequality $\log(a+b) \leq \log 2 + \max(\log a, \log b)$, $a \geq 0$, $b \geq 0$, for any $\epsilon < \lambda$,

$$\limsup_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\Big\{w_n||W_n||_\infty > \lambda\Big\}\right)$$

$$\leq \limsup_{\delta\to 0}\limsup_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\Big\{w_n \max_{1\leq k\leq l_n} \sup_{x\in B_{n,k}} \sup_{z\in\mathcal{L}\times\mathcal{H}_0} |W_n(x,z) - W_n(x_{n,k},z)| > \epsilon\Big\}\right.$$

$$\left.+ \mathbb{P}\Big\{w_n \max_{1\leq k\leq l_n} \sup_{z\in\mathcal{L}\times\mathcal{H}_0} |W_n(x_{n,k},z)| > \lambda - \epsilon\Big\}\right)$$

$$\leq \limsup_{\delta\to 0}\limsup_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\Big\{w_n \max_{1\leq k\leq l_n} \sup_{z\in\mathcal{L}\times\mathcal{H}_0} |W_n(x_{n,k},z)| > \lambda - \epsilon\Big\}\right).$$

On the other hand

$$\limsup_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\Big\{w_n \max_{1\leq k\leq l_n} \sup_{z\in\mathcal{L}\times\mathcal{H}_0} |W_n(x_{n,k},z)| > \lambda - \epsilon\Big\}\right)$$

$$\leq \limsup_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(l_n\right) + \sup_{x\in\mathcal{J}}\limsup_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\Big\{w_n \sup_{z\in\mathcal{L}\times\mathcal{H}_0} |W_n(x,z)| > \lambda - \epsilon\Big\}\right).$$

It follows, by Condition $(B4)(iv)$ that

$$\limsup_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log\left(\mathbb{P}\Big\{w_n||W_n||_\infty > \lambda\Big\}\right) \leq -I^\gamma(\lambda - \epsilon).$$

The demonstration of Theorem 3.3 may be concluded through the asymptotic limit of the parameter ϵ converging towards zero since the function $I^\gamma(\cdot)$ is continuous. $\qquad\square$

To prove Corollary 4.1, we need some intermediate results. For any $x \in \mathcal{J}$ and any $(l,\varrho) \in \mathcal{L} \times \mathcal{H}_0$, set

$$\mathfrak{B}_n(x,l,\varrho) = c_l(x)\left(\widehat{r}^l_{n,2}(x,\varrho h) - r^l(x)\right) + d_l(x)\left(\widehat{r}^l_{n,1}(x,\varrho h) - 1\right). \tag{7.23}$$

Lemma 7.8. *Under the assumption of Theorem* 3.2, *assume that the condition* $(C1)$ *holds. Then the process*

$$\Big\{w_n\mathfrak{B}_n(x,l,\varrho) : (l,\varrho) \in \mathcal{L} \times \mathcal{H}_0\Big\}$$

satisfies a large deviation principle in $l_\infty(\mathcal{L} \times \mathcal{H}_0)$ *with the speed* $(n\phi(\gamma h)/w_n^2)$ *and the good rate function* $I^\gamma_{1,x}(\cdot)$.

The proof of Lemma 7.8 is postponed until Section 8.

Proof of Corollary 4.1

By choosing $c_l(x) = 1$ and $d_l(x) = -r^l(x)$, the sequence $\mathfrak{B}_n(x,l,\varrho h)$ in (7.23) may be then written as

$$\mathfrak{B}_n(x,l,\varrho h) = \left(\widehat{r}^l_{n,2}(x,\varrho h) - r^l(x)\right) - r^l(x)\left(\widehat{r}^l_{n,1}(x,\varrho h) - 1\right).$$

The subsequent statements show that the processes

$$\left\{ w_n(\widehat{r}_n^l(x, \varrho h) - r^l(x, \varrho h)) : (l, \varrho) \in \mathcal{L} \times \mathcal{H}_0 \right\}$$

and

$$\left\{ w_n \mathfrak{B}_n(x, l, \varrho h) : (l, \varrho) \in \mathcal{L} \times \mathcal{H}_0 \right\}$$

are exponentially contiguous. Indeed, we have

$$\mathbb{P}\left(w_n \sup_{(l,\varrho)\in\mathcal{L}\times\mathcal{H}_0} \left| \widehat{r}_n^l(x, \varrho h) - r^l(x, \varrho h) - \mathfrak{B}_n(x, l, \varrho h) \right| > \eta \right)$$

$$\leq \mathbb{P}\left(w_n \sup_{(l,\varrho)\in\mathcal{L}\times\mathcal{H}_0} \left| \mathfrak{B}_n(x, l, \varrho h)\left(\frac{1}{\widehat{r}_{n,1}^l(x, \varrho h)} - 1 \right) \right| > \eta, \inf_{\varrho\in\mathcal{H}_0} \widehat{r}_{n,1}^l(x, \varrho h) \geq 1/2 \right)$$

$$+ \mathbb{P}\left(\inf_{\varrho\in\mathcal{H}_0} \widehat{r}_{n,1}^l(x, \varrho h) < 1/2 \right)$$

$$\leq \mathbb{P}\left(\sqrt{w_n} \sup_{(l,\varrho)\in\mathcal{L}\times\mathcal{H}_0} |\mathfrak{B}_n(x, l, \varrho h)| > \sqrt{\eta} \right) + \mathbb{P}\left(\sqrt{w_n} \sup_{\varrho\in\mathcal{H}_0} |\widehat{r}_{n,1}^l(x, \varrho h) - 1| > \sqrt{\eta}/2 \right)$$

$$+ \mathbb{P}\left(\sup_{\varrho\in\mathcal{H}_0} |\widehat{r}_{n,1}^l(x, \varrho h) - 1| > \frac{1}{2} \right).$$

Since $\lim_{n\to\infty} w_n = \infty$, for n large enough, it follows that

$$\mathbb{P}\left(w_n \sup_{(l,\varrho)\in\mathcal{L}\times\mathcal{H}_0} \left| \widehat{r}_n^l(x, \varrho h) - r^l(x, \varrho h) - \mathfrak{B}_n(x, l, \varrho h) \right| > \eta \right)$$

$$\leq 3 \max \left\{ \mathbb{P}\left(\sqrt{w_n} \sup_{(l,\varrho)\in\mathcal{L}\times\mathcal{H}_0} |\mathfrak{B}_n(x, l, \varrho h)| > \sqrt{\eta} \right), \mathbb{P}\left(\sqrt{w_n} \sup_{\varrho\in\mathcal{H}_0} |\widehat{r}_{n,1}^l(x, \varrho h) - 1| > \sqrt{\eta}/2 \right) \right\}.$$

Now, by Lemma 7.8 the sequence

$$\left\{ w_n \mathfrak{B}_n(x, l, \varrho h) : (l, \varrho) \in \mathcal{L} \times \mathcal{H}_0 \right\}$$

satisfies a large deviation principle with the speed $(n\phi(\gamma h)/w_n^2)$ and the good rate function $I_{1,x}^\gamma(\cdot)$, there exists a constant $c_1 > 0$ such that

$$\limsup_{n\to\infty} \frac{w_n}{n\phi(\gamma h)} \log \mathbb{P}\left(\sqrt{w_n} \sup_{(l,\varrho)\in\mathcal{L}\times\mathcal{H}_0} |\mathfrak{B}_n(x, l, \varrho h)| > \sqrt{\eta} \right) < -c_1.$$

Moreover, an application of Theorem 3.2 guarantees the existence of a real $c_2 > 0$ such that

$$\limsup_{n\to\infty} \frac{w_n}{n\phi(\gamma h)} \log \mathbb{P}\left(\sqrt{w_n} \sup_{\varrho\in\mathcal{H}_0} |\widehat{r}_{n,1}^l(x, \varrho h) - 1| > \sqrt{\eta}/2 \right) < -c_2.$$

We then deduce that

$$\limsup_{n\to\infty} \frac{w_n^2}{n\phi(\gamma h)} \log \mathbb{P}\left(w_n \sup_{(l,\varrho)\in\mathcal{L}\times\mathcal{H}_0} \left| \widehat{r}_n^l(x, \varrho h) - r^l(x, \varrho h) - \mathfrak{B}_n(x, l, \varrho h) \right| > \eta \right) = -\infty,$$

which means that the process

$$\left\{ w_n(\widehat{r}^l_n(x, \varrho h) - r^l(x, \varrho h)) : (l, \varrho) \in \mathcal{L} \times \mathcal{H}_0 \right\}$$

and

$$\left\{ w_n \mathfrak{B}_n(x, l, \varrho h) : (l, \varrho) \in \mathcal{L} \times \mathcal{H}_0 \right\}$$

are exponentially contiguous. Thus, Corollary 4.1 follows by making use of Lemma 7.8.   □

Proof of Corollary 4.2

To see how Corollary 4.2 follows from Theorem 3.2, take in this case

$$\mathcal{L} = \left\{ \mathbb{1}_{\{y \leq t\}} : t \in \mathbb{R} \right\}, \quad \mathcal{D} = \left\{ -F(t|\cdot) : t \in \mathbb{R} \right\} \quad \text{and} \quad c_l(x) = 1,$$

for each $x \in \mathcal{J}$. Under the Condition (C2), it becomes evident that the set of functions $\mathcal{D}$ constitutes a collection of uniformly equicontinuous functions in $(\mathcal{J}, d)$. Consequently, $\mathcal{D}$ forms a set of uniformly bounded functions. Applying the Arzéla−Ascoli theorem (refer to, for instance, Theorem IV.6.7 in [56]), it follows that $\mathcal{D}$ is a compact set within $l_\infty(\mathcal{J})$. It is worth noting that we can employ this theorem even when $(\mathcal{J}, d)$ is a totally bounded pseudo-metric space, not necessarily a compact one. Now, employing the same line of reasoning as in the proof of Corollary 4.1, the proof is successfully established.   □

Proof of Corollary 4.3

By Lemma 3.9.23 of [131], it follows that the inverse map $\Phi : G \mapsto G^{-1}$ as a map $D_1[F^{-1}(p|x) - \varepsilon, F^{-1}(q|x) + \varepsilon] \mapsto l_\infty[p, q]$ is Hadamard differentiable at $F(\cdot|x)$ tangentially to $C[F^{-1}(p|x) - \varepsilon, F^{-1}(q|x) + \varepsilon]$, and the derivative is the map

$$\psi \mapsto -\psi \left(F^{-1}(\cdot|x) \right) / f \left(F^{-1}(\cdot|x) \right).$$

Therefore, by Theorem 3.1 of [67] and Corollary 4.2, we conclude that

$$\{ w_n \left(\widehat{q}_{n,\alpha}(x, \varrho) - q_\alpha(x) \right), (\alpha, \varrho) \in l_\infty([p, q] \times \mathcal{H}_0) \}$$

satisfies the LDP in $l_\infty([p, q] \times \mathcal{H}_0)$ with speed $n\phi(\gamma h)/w_n^2$ and the rate function $I_{\gamma,x}^{EQ}(\cdot)$.   □

8. PROOF OF THE TECHNICAL LEMMAS

Proof of Lemma 7.1

Note that, the condition (A1) and (A2)(i) implies the equality (7.1), and this last equality with $M_{x,l}(v) = M_0(v)$ gives that for any $(x, \varrho) \in \mathcal{J} \times \mathcal{H}_0$,

$$\frac{\mathbf{E}[M_0(Y_1)\Delta_1(x, \varrho h)]}{\phi(\varrho h)} = \int L_{x,\varrho}(v, h) M_0(v) g(v) \, dv,$$

where

$$L_{x,\varrho}(v, h) = K(1)\left(f_v(x) + \frac{g_{x,v}(\varrho h)}{\phi(\varrho h)} \right) - \int\limits_0^1 K'(u) \frac{\phi(u\varrho h)}{\phi(\varrho h)} \left(f_v(x) + \frac{g_{x,v}(u\varrho h)}{\phi(u\varrho h)} \right) du.$$

Now, observe that, for any $(x, \varrho) \in \mathcal{J} \times \mathcal{H}_0$,

$$\left| \int L_{x,\varrho}(v, h) M_0(v) g(v) \, dv - \alpha_0 \int M_0(v) f_v(x) g(v) \, dv \right|$$

$$\leq K(1) \int \frac{g_{x,v}(\varrho h)}{\phi(\varrho h)} M_0(v)g(v)\, dv$$

$$+ \left(\int M_0(v)f_v(x)g(v)\, dv \right) \int_0^1 |K'(u)| \left| \frac{\phi(u\varrho h)}{\phi(\varrho h)} - \tau_0(u) \right| du$$

$$+ \int \int_0^1 |K'(u)| \frac{g_{x,v}(u\varrho h)}{\phi(u\varrho h)} M_0(v)g(v)\, du\, dv.$$

Making use of the conditions (A1)–(A2) and (B3), we derive

$$\lim_{n\to\infty} \sup_{(x,\varrho)\in\mathcal{J}\times\mathcal{H}_0} \left| \frac{\mathbf{E}[M_0(Y_1)\Delta_1(x,\varrho h)]}{\phi(\varrho h)} - \alpha_0 \int M_0(v)f_v(x)g(v)\, dv \right| = 0.$$

This completes the proof of Lemma 7.1. $\square$

Proof of Lemma 7.2

Making use of the conditions (A1)–(A2) and (B3)–(B4)(i), and applying Lemma 7.1, there exists a positive constant C_1 such that, for every pair $l, l' \in \mathcal{L}$, every x in $\mathcal{J}$, every ϱ in $\mathcal{H}_0$, and for sufficiently large n, the following holds:

$$\mathbf{E}\left[\left\{ (c_l(x)l(Y) - c_{l'}(x)l'(Y))\mathbb{1}_{\{L(Y)\leq\mathcal{M}^{\mathrm{inv}}(n)\}} + (d_l(x) - d_{l'}(x)) \right\}^2 \frac{(\Delta_1(x,\varrho h))^2}{(\mathbf{E}[\Delta_1(x,\varrho h)])^2/\phi(\gamma h)} \right]$$

$$\leq C_1 \frac{\phi(\gamma h)}{\phi(\varrho h)} \int \left\{ (c_l(x)l(v) - c_{l'}(x)l'(v))\mathbb{1}_{\{L(v)\leq\mathcal{M}^{\mathrm{inv}}(n)\}} + (d_l(x) - d_{l'}(x)) \right\}^2 g(v)dv.$$

According to (A2)(ii), we obtain, uniformly for ϱ in $\mathcal{H}_0$,

$$\frac{\phi(\varrho h)}{\phi(\gamma h)} = \tau(\varrho, \gamma) + o(1).$$

By combining the last equation with the observation that $0 < \tau_0\left(\frac{\vartheta_1}{\vartheta_2}\right) \leq \tau_0\left(\frac{\varrho}{\gamma}\right)$, we can deduce the existence of a positive constant C_2 such that, for sufficiently large n,

$$C_2\phi(\gamma h) \leq \inf_{\varrho\in\mathcal{H}_0} \phi(\varrho h).$$

Henceforth, there exists a positive constant $C > 0$ such that, for every pair $l, l' \in \mathcal{L}$, every ϱ in $\mathcal{H}_0$, and for sufficiently large n,

$$\mathbf{E}\left[\left\{ (c_l(x)l(Y) - c_{l'}(x)l'(Y))\mathbb{1}_{\{L(Y)\leq\mathcal{M}^{\mathrm{inv}}(n)\}} + (d_l(x) - d_{l'}(x)) \right\}^2 \frac{(\Delta_1(x,\varrho h))^2}{(\mathbf{E}[\Delta_1(x,\varrho h)])^2/\phi(\gamma h)} \right]$$

$$\leq C \int \left\{ (c_l(x)l(v) - c_{l'}(x)l'(v))\mathbb{1}_{\{L(v)\leq\mathcal{M}^{\mathrm{inv}}(n)\}} + (d_l(x) - d_{l'}(x)) \right\}^2 g(v)dv. \qquad (8.1)$$

Given the Condition (B1) on the class $\mathcal{L}$, which states that it is totally bounded with respect to the distance $d_Q(\cdot, \cdot)$, where $Q(\cdot)$ is a distribution with density $g(\cdot)$, it follows that for any $\delta > 0$, there exists a finite subclass $\mathcal{L}_1 \subset \mathcal{L}$ such that

$$\sup_{l\in\mathcal{L}} \min_{l'\in\mathcal{L}_1} \int (l(v) - l'(v))^2 g(v)dv < \delta.$$

Furthermore, due to the compactness of the function classes $\mathcal{C}$ and $\mathcal{D}$, we can identify finite subclasses $\mathcal{L}_2 \subset \mathcal{L}$ and $\mathcal{L}_3 \subset \mathcal{L}$ such that

$$\sup_{l\in\mathcal{L}} \min_{l'\in\mathcal{L}_2} ||c_l - c_{l'}||_{\mathcal{J}} \vee \sup_{l\in\mathcal{L}} \min_{l'\in\mathcal{L}_3} ||d_l - d_{l'}||_{\mathcal{J}} < \delta.$$

Therefore, employing the observation that both

$$\sup_{l \in \mathcal{L}} \|c_l\|_{\mathcal{J}} < \infty \quad \text{and} \quad \sup_{l \in \mathcal{L}} \int (l(v))^2 g(v) dv < \infty.$$

After a brief arithmetic manipulation and selecting a sufficiently small $\delta > 0$, we obtain

$$\sup_{l \in \mathcal{L}} \min_{l'_1, l'_2, l'_3} \sup_{x \in \mathcal{J}} \int \left\{ (c_l(x)l(v) - c_{l'_2}(x)l'_1(v)) \mathbb{1}_{\{L(v) \le \mathcal{M}^{\mathrm{inv}}(n)\}} + (d_l(x) - d_{l'_3}(x)) \right\}^2 g(v) dv \le \epsilon/2C,$$

where the minimum is taken over $\mathcal{L}_1 \times \mathcal{L}_2 \times \mathcal{L}_3$. Now, consider any triple $(l'_1, l'_2, l'_3) \in \mathcal{L}_1 \times \mathcal{L}_2 \times \mathcal{L}_3$ for which there exists $l' \in \mathcal{L}$ such that

$$\sup_{x \in \mathcal{J}} \int \left\{ (c_{l'}(x)l'(v) - c_{l'_2}(x)l'_1(v)) \mathbb{1}_{\{L(v) \le \mathcal{M}^{\mathrm{inv}}(n)\}} + (d_{l'}(x) - d_{l'_3}(x)) \right\}^2 g(v) dv \le \epsilon/2C,$$

select one of them to form the desired subclass $\mathcal{L}_\epsilon$. To conclude the proof, it is enough to apply the triangle inequality. $\qquad \square$

Proof of Lemma 7.5

Set, for $z = (x, l, \varrho) \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0$,

$$\mathbf{W}''_n(z) = \frac{1}{n\mathbf{E}[\Delta_1(x, \varrho h)]} \sum_{i=1}^n c_l(x)l(Y_i) \mathbb{1}_{\{L(Y_i) > \mathcal{M}^{\mathrm{inv}}(n)\}} \Delta_i(x, \varrho h).$$

Observe first that $s/\mathcal{M}(s)$ is a decreasing function. In turn, this implies, for any v such that $L(v) \ge \mathcal{M}^{\mathrm{inv}}(n)$, that

$$|l(v)| = \frac{|l(v)|}{L(v)} \frac{L(v)}{\mathcal{M}(L(v))} \mathcal{M}(L(v)) \le \left\{ \frac{\mathcal{M}^{\mathrm{inv}}(n)}{n} \right\} \mathcal{M}(L(v)).$$

Thus, for n large enough, uniformly in $(x, l, \varrho) \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0$, we have

$$\mathbf{E}\left[|\mathbf{W}''_n(z)| \right] \le \left\{ \frac{\mathcal{M}^{\mathrm{inv}}(n)}{n} \right\} C_{\mathcal{L}} \mathbf{E}\left\{ \mathcal{M}(L(Y)) \frac{\Delta_1(x, \varrho h)}{\mathbf{E}[\Delta_1(x, \varrho h)]} \right\}.$$

Thus, by conditions (A1)–(A2) and (B3)–(B4)(i) and using Lemma 7.1, there exists two strictly positive constants C_1, C_2 such that, for any $x \in \mathcal{J}$ and $\varrho \in \mathcal{H}_0$, we have

$$C_1 \phi(\varrho h) \le \mathbf{E}[\Delta_1(x, \varrho h)] \tag{8.2}$$

and

$$\mathbf{E}[\mathcal{M}(L(Y))\Delta_1(x, \varrho h)] \le C_2 \phi(\varrho h). \tag{8.3}$$

Hence, we readily infer

$$\sup_{z \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0} w_n \mathbf{E}\left[|\mathbf{W}''_n(z)| \right] \le \frac{w_n \mathcal{M}^{\mathrm{inv}}(n)}{n} \frac{C_2 C_{\mathcal{L}}}{C_1},$$

which by (A3) converges to 0 as $n \to \infty$. Considering now the inequality (8.2) and the boundedness of the kernel $K(\cdot)$ in (A1), for any for $z \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0$, we have

$$|\mathbf{W}''_n(z)| \le \frac{C_{\mathcal{L}}}{C_1} \frac{1}{n\phi(\varrho h)} \sum_{i=1}^n L(Y_i) \mathbb{1}_{\{L(Y_i) > \mathcal{M}^{\mathrm{inv}}(n)\}}.$$

By (A2)(ii), we have, uniformly in $\varrho \in \mathcal{H}_0$,

$$\frac{\phi(\varrho h)}{\phi(\gamma h)} = \tau(\varrho, \gamma) + o(1).$$

Combining this with the fact that $0 < \tau_0(\frac{\vartheta_1}{\vartheta_2}) \le \tau_0(\frac{\varrho}{\gamma})$ implies that there exists a constant $C_2 > 0$ such that, for n large enough,

$$C_2 \phi(\gamma h) \le \inf_{\varrho \in \mathcal{H}_0} \phi(\varrho h).$$

Therefore, for any $\eta > 0$ and n large enough, we have

$$\mathbb{P}\left(\sup_{z \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0} w_n |\mathbf{W}_n(z) - \tilde{\mathbf{W}}_n(z)| \ge \eta \right) \le \mathbb{P}\left(\sup_{z \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0} w_n |\mathbf{W}_n''(z)| \ge \eta/2 \right)$$

$$\le \mathbb{P}\left(\frac{C_{\mathcal{L}}}{C_1} \frac{w_n}{nC_2\phi(\gamma h)} \sum_{i=1}^{n} L(Y_i) \mathbb{1}_{\{L(Y_i) > \mathcal{M}^{\text{inv}}(n)\}} \ge \eta/2 \right)$$

$$\le \mathbb{P}\left(\max_{1 \le i \le n} L(Y_i) > \mathcal{M}^{\text{inv}}(n) \right)$$

$$\le n\mathbb{P}\left(L(Y) > \mathcal{M}^{\text{inv}}(n) \right).$$

The application of the exponential Tchebychev inequality with t chosen as in the condition (B4)(i) yields to

$$\lim_{n \to \infty} \frac{w_n^2}{n\phi(\gamma h)} \log \left(n\mathbb{P}\left(L(Y) > \mathcal{M}^{\text{inv}}(n) \right) \right)$$

$$\le \lim_{n \to \infty} \left(\frac{w_n^2}{n\phi(\gamma h)} \log n - t \frac{w_n^2}{n\phi(\gamma h)} \mathcal{M}^{\text{inv}}(n) + \frac{w_n^2}{n\phi(\gamma h)} \log \left(\mathbf{E}\left[e^{tL(Y)} \right] \right) \right),$$

$$\le \lim_{n \to \infty} \left((1-t) \frac{w_n^2}{n\phi(\gamma h)} \max \left(\mathcal{M}^{\text{inv}}(n), \log n \right) + \frac{w_n^2}{n\phi(\gamma h)} \log \left(\mathbf{E}\left[e^{tL(Y)} \right] \right) \right),$$

which by Assumptions (A3) and (B4)(ii) converges to $-\infty$ as $n \to \infty$. $\qquad\square$

Proof of Lemma 7.6

Similarly as in the proof of Lemma 5 in [58], it follows that the class

$$\tilde{\mathfrak{M}}_1 = \left\{ (u, v) \to \mathbb{1}_{\{L(v) \le t\}} K(d(x, u)/\varrho h), \varrho \in \mathcal{H}_0, x \in \mathcal{J}, t > 0 \right\}$$

satisfies for any probability measure Q on Borel subsets of $\mathcal{J} \times \mathbb{R}$ the condition

$$N(\epsilon\kappa, \tilde{\mathfrak{M}}_1, d_Q) \le \tilde{C}\epsilon^{-\tilde{\nu}}, \quad 0 < \epsilon < 1,$$

where $\tilde{C}$, $\tilde{\nu} > 0$ are suitable positive constants. Next, consider the function class

$$\tilde{\mathfrak{M}}_2 = \left\{ (u, v) \mapsto \phi(\gamma h) \frac{c_l(x) l(v)}{\mathbf{E}[\Delta_1(x, \varrho h)]} : \varrho \in \mathcal{H}_0, l \in \mathcal{L}, x \in \mathcal{J} \right\}.$$

Applying the same reasoning as presented in the proof of Lemma 5 in [58] and leveraging the Vapnik−Červonenkis property of $\mathcal{L}$, along with the inequality (8.2) and the bounded nature of the classes $\mathcal{C}$, it follows that the class $\tilde{\mathfrak{M}}_2$ possesses a polynomial covering number. Consequently, by referring to Lemma A.1 in [58], it can be deduced that the product class $\tilde{\mathfrak{M}}_1 \cdot \tilde{\mathfrak{M}}_2$ exhibits a polynomial covering number. Now, let's consider the class

$$\tilde{\mathfrak{M}}_3 = \left\{ (u, v) \mapsto \phi(\gamma h) \frac{d_l(x)}{\mathbf{E}[\Delta_1(x, \varrho h)]} : \varrho \in \mathcal{H}_0, l \in \mathcal{L}, x \in \mathcal{J} \right\}.$$

Utilizing inequality (8.2), the bounded nature of the classes $\mathcal{D}$, and invoking Lemma A.1 from [58], it can be established that the product class $\mathcal{K} \cdot \tilde{\mathfrak{M}}_3$ possesses a polynomial covering number. Consequently, the class resulting from the summation of $\tilde{\mathfrak{M}}_1 \cdot \tilde{\mathfrak{M}}_2$ and $\mathcal{K} \cdot \tilde{\mathfrak{M}}_3$ also exhibits a polynomial covering number. Finally, it can be concluded that the class $\mathfrak{M}$ satisfies this covering property as well. The measurability is straightforwardly derived from the kernel function's continuity, the separability of $(\mathcal{J}, d)$, and the fact that the functions belong to the classes $\mathcal{C}$ and $\mathcal{D}$.   □

Proof of Lemma 7.7

Recall the Eq. (7.21). Note that, for any $(x_1, x_2) \in \mathcal{J}^2$ and any $(z_1, z_2) \in \left(\mathcal{L} \times \mathcal{H}_0\right)^2$,

$$\mathrm{Var}(B_{n,x_1,z_1}(X,Y) - B_{n,x_2,z_2}(X,Y))$$

$$\leq 2\mathbb{E}\left[\phi(\gamma h)\left\{\bar{\mathcal{M}}_{x_1,l_1}(Y) - \bar{\mathcal{M}}_{x_1,l_2}(Y)\right\}\frac{\Delta_1(x_1, \varrho_1 h)}{\mathbf{E}[\Delta_1(x_1, \varrho_1 h)]}\right]^2$$

$$+ 2\mathbb{E}\left[\phi(\gamma h)\left\{\frac{\bar{\mathcal{M}}_{x_1,l_2}(Y)\Delta_1(x_1, \varrho_1 h)}{\mathbf{E}[\Delta_1(x_1, \varrho_1 h)]} - \frac{\bar{\mathcal{M}}_{x_2,l_2}(Y)\Delta_1(x_2, \varrho_2 h)}{\mathbf{E}[\Delta_1(x_2, \varrho_2 h)]}\right\}\right]^2$$

$$:= I + II.$$

According to Lemma 7.2, for sufficiently large n, there exists a constant $C > 0$ such that

$$I \leq 2C d_{\mathcal{L}}(l_1, l_2)\phi(\gamma h).$$

Now, observe that

$$II \leq 4\phi(\gamma h)^2 \mathbb{E}\left[\left([c_{l_2}(x_1) - c_{l_2}(x_2)]l_2(Y)\mathbb{1}_{\{L(Y)\leq\mathcal{M}^{\mathrm{inv}}(n)\}} + [d_{l_2}(x_1) - d_{l_2}(x_2)]\right)\frac{\Delta_1(x_1, \varrho_1 h)}{\mathbf{E}[\Delta_1(x_1, \varrho_1 h)]}\right]^2$$

$$+ 4\phi(\gamma h)^2 \mathbb{E}\left[\left(c_{l_2}(x_2)l_2(Y)\mathbb{1}_{\{L(Y)\leq\mathcal{M}^{\mathrm{inv}}(n)\}} + d_{l_2}(x_2)\right)\left[\frac{\Delta_1(x_2, \varrho_2 h)}{\mathbf{E}[\Delta_1(x_2, \varrho_2 h)]} - \frac{\Delta_1(x_1, \varrho_1 h)}{\mathbf{E}[\Delta_1(x_1, \varrho_1 h)]}\right]\right]^2$$

$$:= III + IV.$$

Using the fact that the classes $\mathcal{C}$ and $\mathcal{D}$ are uniformly equicontinuous, it follows, for any $\epsilon > 0$, that there exists $\delta > 0$ such that for any $x_1, x_2 \in \mathcal{J}$, with $d(x_1, x_2) \leq \delta$, and

$$\sup_{l\in\mathcal{L}}|c_l(x_1) - c_l(x_2)| \vee \sup_{l\in\mathcal{L}}|d_l(x_1) - d_l(x_2)| \leq \epsilon.$$

Hence, we have

$$III \leq 4(\epsilon\phi(\gamma h))^2 \mathbb{E}\left[\left(L(Y) + 1\right)\frac{\Delta_1(x_1, \varrho_1 h)}{\mathbf{E}[\Delta_1(x_1, \varrho_1 h)]}\right]^2.$$

Using the same arguments as in (8.1), by the Conditions (A1)–(A2) and (B3)–(B4)(i), and applying Lemma 7.1 with the observation that $0 < \tau_0\left(\frac{\vartheta_1}{\vartheta_2}\right) \leq \tau_0\left(\frac{\varrho_1}{\gamma}\right)$, there exists a positive constant C_1 such that

$$III \leq C_1\epsilon^2\phi(\gamma h).$$

Similarly, and using the fact that the classes $\mathcal{C}$ and $\mathcal{D}$ are uniformly bounded, for suitable finite constants C_1 and C_2, we have

$$IV \leq 8\phi(\gamma h)^2 \mathbb{E}\left[\left(c_{l_2}(x_2)l_2(Y)\mathbb{1}_{\{L(Y)\leq\mathcal{M}^{\mathrm{inv}}(n)\}} + d_{l_2}(x_2)\right)\left[\frac{\Delta_1(x_2, \varrho_2 h))}{\mathbf{E}[\Delta_1(x_2, \varrho_2 h)]} - \frac{\Delta_1(x_1, \varrho_1 h)}{\mathbf{E}[\Delta_1(x_2, \varrho_2 h)]}\right]\right]^2$$

$$+ 8\phi(\gamma h)^2 \mathbb{E}\left[\left(c_{l_2}(x_2)l_2(Y)\mathbb{1}_{\{L(Y)\leq\mathcal{M}^{\mathrm{inv}}(n)\}} + d_{l_2}(x_2)\right)\left[\frac{\Delta_1(x_1, \varrho_1 h))}{\mathbf{E}[\Delta_1(x_2, \varrho_2 h)]} - \frac{\Delta_1(x_1, \varrho_1 h)}{\mathbf{E}[\Delta_1(x_1, \varrho_1 h)]}\right]\right]^2.$$

By (7.17) and (B4)(i), and by using the fact that the classes $\mathcal{C}$ and $\mathcal{D}$ are uniformly bounded, for suitable finite constants C_1 and C_2, we have

$$IV$$

$$\leq C_1 \mathbf{E}\left[\left(L(Y)+1\right)\left[\Delta_1(x_2,\varrho_2 h)-\Delta_1(x_1,\varrho_1 h)\right]\right]^2 + \frac{C_2}{\phi(\gamma h)}\left[\mathbf{E}\left[\Delta_1(x_2,\varrho_2 h)-\Delta_1(x_1,\varrho_1 h)\right]\right]^2$$

$$\leq 2C_1\left\{\mathbf{E}\left[\left(L(Y)+1\right)\left[\Delta_1(x_2,\varrho_2 h)-\Delta_1(x_2,\varrho_1 h)\right]\right]^2 + \mathbf{E}\left[\left(L(Y)+1\right)\left[\Delta_1(x_2,\varrho_1 h)-\Delta_1(x_1,\varrho_1 h)\right]\right]^2\right\}$$

$$+ 2\frac{C_2}{\phi(\gamma h)}\left\{\left[\mathbf{E}\left[\Delta_1(x_2,\varrho_2 h)-\Delta_1(x_2,\varrho_1 h)\right]\right]^2 + \left[\mathbf{E}\left[\Delta_1(x_2,\varrho_1 h)-\Delta_1(x_1,\varrho_1 h)\right]\right]^2\right\}.$$

Now, by (A1)(i), we obtain, for some constant C,

$$\mathbf{E}\left[\left(L(Y)+1\right)\left[\Delta_1(x_2,\varrho_2 h)-\Delta_1(x_2,\varrho_1 h)\right]\right]^2$$

$$= \int_0^1 \int (L(v)+1)^2 \left(K\left(\frac{u}{\varrho_2}\right)-K\left(\frac{u}{\varrho_1}\right)\right)^2 d\mathbb{P}\left(\frac{d(x_2,X_1)}{h}\leq u, Y\leq v\right)$$

$$\leq C\frac{(\varrho_2-\varrho_1)^2}{\vartheta_1^2}\int_0^1 \int (L(v)+1)^2 u^2 \, d\mathbb{P}\left(\frac{d(x_2,X_1)}{h}\leq u|Y=v\right)g(v)\, dv.$$

Using the same arguments as in (7.1), we get

$$\mathbf{E}\left[\left(L(Y)+1\right)\left[\Delta_1(x_2,\varrho_2 h)-\Delta_1(x_2,\varrho_1 h)\right]\right]^2 \leq C\,(\varrho_2-\varrho_1)^2\phi(h)\int_0^1 \Gamma_{x_2}(v,h)\,(L(v)+1)^2\,g(v)\,dv,$$

where

$$\Gamma_x(v,h) = f_v(x) + \frac{g_{x,v}(h)}{\phi(h)} - 2\int_0^1 u\frac{\phi(uh)}{\phi(h)}\left(f_v(x)+\frac{g_{x,v}(uh)}{\phi(uh)}\right)du.$$

By the Conditions (A1)–(A2) and (B4)(i), It follows that

$$\mathbf{E}\left[\left(L(Y)+1\right)\left[\Delta_1(x_2,\varrho_2 h)-\Delta_1(x_2,\varrho_1 h)\right]\right]^2 \leq C\,|\varrho_2-\varrho_1|\phi(\gamma h). \tag{8.4}$$

By using the same arguments as above, it follows that

$$\mathbf{E}\left[\left(L(Y)+1\right)\left[\Delta_1(x_2,\varrho_1 h)-\Delta_1(x_1,\varrho_1 h)\right]\right]^2 \leq C\left(\frac{d(x_1,x_2)}{h}\right)^2\phi(\gamma h),$$

which implies, by using the fact that $d(x_1,x_2)\leq \delta\phi(\gamma h)$ and $\phi(\gamma h)/h = O(1)$,

$$\mathbf{E}\left[\left(L(Y)+1\right)\left[\Delta_1(x_2,\varrho_1 h)-\Delta_1(x_1,\varrho_1 h)\right]\right]^2 \leq C\,\delta\phi(\gamma h). \tag{8.5}$$

Similarly

$$\mathbf{E}\left|\Delta_1(x,\varrho_2 h)-\Delta_1(x,\varrho_1 h)\right| \leq C\,|\varrho_2-\varrho_1|\phi(\gamma h) \tag{8.6}$$

and

$$\mathbf{E}\Big|\Delta_1(x_2, \varrho_1 h) - \Delta_1(x, \varrho_1 h)\Big| \leq C\,\delta\phi(\gamma h). \tag{8.7}$$

Finally, from (8.4)–(8.7), we deduce

$$IV \leq C\,\phi(\gamma h)\,(\delta + |\varrho_2 - \varrho_1|),$$

which completes the proof of the Lemma 7.7.   □

Proof of Lemma 7.8

Observe, for any $(x, z) \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0$, that

$$\mathfrak{B}_n(x, l, \varrho h) = W_n(x, z) + c_l(x)(\mathbf{E}[\widehat{r}^l_{n,2}(x, \varrho h)] - r^l(x))$$
$$=: W_n(x, z) + V_n(x, z),$$

where $z = (l, \varrho)$. To prove Lemma 7.8, we have to show that

$$\lim_{n \to \infty} w_n \sup_{z \in \mathcal{L} \times \mathcal{H}_n} V_n(x, z) = 0.$$

Now, for any $(x, l, \varrho) \in \mathcal{J} \times \mathcal{L} \times \mathcal{H}_0$, observe that

$$\mathbf{E}[\widehat{r}^l_{n,2}(x, \varrho h) - r^l(x)]$$

$$= \sum_{i=1}^n \frac{\mathbf{E}[l(Y_i)\Delta_i(x, \varrho h)]}{n\mathbf{E}[\Delta_1(x, \varrho h)]} - r^l(x) = \frac{\mathbf{E}[l(Y_1)\Delta_1(x, \varrho h)]}{\mathbf{E}[\Delta_1(x, \varrho h)]} - r^l(x)$$

$$= \frac{\mathbf{E}[\mathbf{E}[l(Y_1)\Delta_1(x, \varrho h)|X_1]]}{\mathbf{E}[\Delta_1(x, \varrho h)]} - r^l(x) = \frac{\mathbf{E}[r^l(X_1)\Delta_1(x, \varrho h)]}{\mathbf{E}[\Delta_1(x, \varrho h)]} - r^l(x)$$

$$= \frac{\mathbf{E}\left[\left(r^l(X_1) - r^l(x)\right)\Delta_1(x, \varrho h)\right]}{\mathbf{E}[\Delta_1(x, \varrho h)]}.$$

Since the kernel function $K(\cdot)$ is $[0, 1]$-supported by condition (A1), it follows that

$$\Big|r^l(X_1) - r^l(x)\Big|\Delta_1(x, \varrho h)) \leq \sup_{\{x': d(x', x) \leq \varrho h\}} \Big|r^l(x') - r^l(x)\Big|\Delta_1(x, \varrho h).$$

The use of the Condition (C1) allows us to obtain the claimed result.   □

APPENDIX A

This appendix contains supplementary information that is essential to providing a more comprehensive understanding of the paper.

Theorem A.1 (Theorem 3.1 [6]). *Let $\{U_n(t) : t \in T\}$ be a sequence of stochastic processes, where T is an index set. Let $\{\varepsilon_n\}$ be a sequence of positive numbers that converge to zero. Let $I : l_\infty(T) \to [0, \infty]$ and let $I_{t_1, \ldots, t_m} : \mathbb{R}^m \to [0, \infty]$ be a function for each $t_1, \ldots, t_m \in T$. Let $d(\cdot, \cdot)$ be a pseudometric in T. Consider the following conditions:*

(a.1) *(T, d) is totally bounded;*

(a.2) *for each $t_1, \ldots, t_m \in T, (U_n(t_1), \ldots, U_n(t_m))$ satisfies the LDP with the rate ε_n^{-1} and good rate function $I_{t_1, \ldots, t_m}$;*

(a.3) *for each $\tau > 0$,*

$$\lim_{\eta \to 0} \limsup_{n \to \infty} \varepsilon_n \log\left(\mathbb{P}^*\left\{\sup_{d(s,t) \leq \eta} |U_n(t) - U_n(s)| \geq \tau\right\}\right) = -\infty;$$

(b.1) *for each $0 \leq c < \infty$, $\left\{ z \in l_\infty(T) : I(z) \leq c \right\}$ is a compact set of $l_\infty(T)$;*

(b.2) *for each $A \subset l_\infty(T)$,*

$$- \inf_{z \in A^\circ} I(z) \leq \liminf_{n \to \infty} \varepsilon_n \log \left(\mathbb{P}_* \left\{ U_n \in A \right\} \right)$$
$$\leq \limsup_{n \to \infty} \varepsilon_n \log \left(\mathbb{P}^* \left\{ U_n \in A \right\} \right) \leq - \inf_{z \in \bar{A}} I(z).$$

If the set of conditions (a) is satisfied, then the set of conditions (b) holds with $I(\cdot)$ given by

$$I(z) = \sup \left\{ I_{t_1, \ldots, t_m} \left(z(t_1), \ldots, z(t_m) \right) : t_1, \ldots, t_m \in T, m \geq 1 \right\}.$$

If the set of conditions (b) *is satisfied, then the set of conditions* (a) *holds with*

$$I_{t_1, \ldots, t_m} (u_1, \ldots, u_m) = \inf \left\{ I(z) : z \in l_\infty(T), \left(z(t_1), \ldots, z(t_m) \right) = (u_1, \ldots, u_m) \right\}$$

and the pseudometric $d(\cdot, \cdot)$ is defined

$$d(s, t) = \sum_{k=1}^{\infty} k^{-2} \min \left(d_k(s, t), 1 \right),$$

where

$$d_k(s, t) = \sup \left\{ |u_2 - u_1| : I_{s,t}(u_1, u_2) \leqq k \right\}.$$

ACKNOWLEDGEMENTS

FUNDING

CONFLICTS OF INTEREST

The authors of this work declare that they have no conflicts of interest.

REFERENCES

1. I. M. Almanjahie, S. Bouzebda, Z. Chikr Elmezouar, and A. Laksaci, "The functional kNN estimator of the conditional expectile: Uniform consistency in number of neighbors," Stat. Risk Model. **38** (3−4), 47−63 (2022).

2. I. M. Almanjahie, S. Bouzebda, Z. Kaid, and A. Laksaci, "Nonparametric estimation of expectile regression in functional dependent data," J. Nonparametr. Stat. **34** (1), 250−281 (2022).

3. I. M. Almanjahie, S. Bouzebda, Z. Kaid, and A. Laksaci, "The local linear functional kNN estimator of the conditional expectile: Uniform consistency in number of neighbors," Metrika **34** (1), 1−29 (2024).

4. G. Aneiros, R. Cao, R. Fraiman, and P. Vieu, "Editorial for the special issue on functional data analysis and related topics," J. Multivariate Anal. **170**, 1−2 (2019).

5. G. Aneiros-Pérez and P. Vieu, "Nonparametric time series prediction: A semi-functional partial linear modeling," J. Multivariate Anal. **99** (5), 834−857 (2008).

6. M. A. Arcones, "The large deviation principle for stochastic processes. I," Teor. Veroyatnost. i Primenen. **47** (4), 727−746 (2002).

7. M. A. Arcones, "The large deviation principle for stochastic processes. II," Teor. Veroyatnost. i Primenen. **48** (1), 122–150 (2003).

8. M. A. Arcones, *Moderate Deviations of Empirical Processes*. In Stochastic Inequalities and Applications, Vol. 56 of Progr. Probab. (Birkhäuser, Basel, 2003), pp. 189–212.

9. R. R. Bahadur, Some limit theorems in statistics, No. 4: *Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics*, Society for Industrial and Applied Mathematics (Philadelphia, PA, 1971).

10. R. R. Bahadur and S. L. Zabell, "Large deviations of the sample mean in general vector spaces," Ann. Probab. **7** (4), 587–621 (1979).

11. N. Berrahou, "Principe de grandes déviations uniforme pour l'estimateur de la densité par la méthode des delta-suites," C. R. Math. Acad. Sci. Paris **343** (9), 595–600 (2006).

12. N. Berrahou, "Large deviations probabilities for a symmetry test statistic based on delta-sequence density estimation," Statist. Probab. Lett. **78** (3), 238–248 (2008).

13. V. I. Bogachev, *Gaussian Measures*, Vol. 62: Mathematical Surveys and Monographs. American Mathematical Society (Providence, RI, 1998).

14. D. Bosq, *Linear Processes in Function Spaces*, Vol. 149: Lecture Notes in Statistics (Springer-Verlag, New York. Theory and Applications, 2000).

15. S. Bouzebda, "On the strong approximation of bootstrapped empirical copula processes with applications," Math. Methods Statist. **21** (3), 153–188 (2012).

16. S. Bouzebda, "General tests of conditional independence based on empirical processes indexed by functions," Jpn. J. Stat. Data Sci. **6** (1), 115–177 (2023).

17. S. Bouzebda, "On the weak convergence and the uniform-in-bandwidth consistency of the general conditional U-processes based on the copula representation: Multivariate setting," Hacet. J. Math. Stat. **52** (5), 1303–1348 (2023).

18. S. Bouzebda and M. Chaouch, "Uniform limit theorems for a class of conditional Z-estimators when covariates are functions," J. Multivariate Anal. **189** (104872), 21 (2022).

19. S. Bouzebda and M. Cherfi, "General bootstrap for dual ϕ-divergence estimates," J. Probab. Stat., Art. ID 834107, 33 (2012).

20. S. Bouzebda and S. Didi, "Some results about kernel estimators for function derivatives based on stationary and ergodic continuous time processes with applications," Comm. Statist. Theory Methods **51** (12), 3886–3933 (2022).

21. S. Bouzebda and I. Elhattab, "Uniform in bandwidth consistency of the kernel-type estimator of the Shannon's entropy," C. R. Math. Acad. Sci. Paris **348** (5–6), 317–321 (2010).

22. S. Bouzebda and I. Elhattab, "Uniform-in-bandwidth consistency for kernel-type estimators of Shannon's entropy," Electron. J. Stat. **5**, 440–459 (2011).

23. S. Bouzebda and N. Limnios, "The uniform CLT for the empirical estimator of countable state space semi-Markov kernels indexed by functions with applications," J. Nonparametr. Stat. **34** (4), 758–788 (2022).

24. S. Bouzebda and B. Nemouchi, "Weak-convergence of empirical conditional processes and conditional U-processes involving functional mixing data," Stat. Inference Stoch. Process. **26** (1), 33–88 (2023).

25. S. Bouzebda and A. Nezzal, "Uniform consistency and uniform in number of neighbors consistency for nonparametric regression estimates and conditional U-statistics involving functional data," Jpn. J. Stat. Data Sci. **5** (2), 431–533 (2022).

26. S. Bouzebda and A. Nezzal, "Asymptotic properties of conditional U-statistics using delta sequences," Comm. Statist. Theory Methods, 1–56 (2024).
https://doi.org/10.1080/03610926.2023.2179887

27. S. Bouzebda and A. Nezzal, "Uniform in number of neighbors consistency and weak convergence of kNN empirical conditional processes and kNN conditional U-processes involving functional mixing data," AIMS Math. **9** (2), 4427–4550 (2024).

28. S. Bouzebda and I. Soukarieh, "Non-parametric conditional U-processes for locally stationary functional random fields under stochastic sampling design," Mathematics **11** (1), 1–70 (2023).

29. S. Bouzebda and N. Taachouche, On the variable bandwidth kernel estimation of conditional U-statistics at optimal rates in sup-norm. Phys. A **625** (129000), 72 (2023).

30. S. Bouzebda and N. Taachouche, "Rates of the strong uniform consistency for the kernel-type regression function estimators with general kernels on manifolds," Math. Methods Statist. **32** (1), 27−80 (2023).

31. S. Bouzebda and N. Taachouche, "Rates of the strong uniform consistency with rates for conditional U-statistics estimators with general kernels on manifolds," Math. Methods Statist. **33** (1), 1−55 (2024).

32. S. Bouzebda and T. Zari, "Strong approximation of multidimensional $\mathbb{P}−\mathbb{P}$ plots processes by Gaussian processes with applications to statistical tests," Math. Methods Statist. **23** (3), 210−238 (2014).

33. S. Bouzebda, I. Elhattab, and C. T. Seck, "Uniform in bandwidth consistency of nonparametric regression based on copula representation," Statist. Probab. Lett. **137**, 173−182 (2018).

34. S. Bouzebda, I. Elhattab, and B. Nemouchi, "On the uniform-in-bandwidth consistency of the general conditional U-statistics based on the copula representation," J. Nonparametr. Stat. **33** (2), 321−358 (2021).

35. S. Bouzebda, M. Chaouch, and S. Didi Biha, "Asymptotics for function derivatives estimators based on stationary and ergodic discrete time processes," Ann. Inst. Statist. Math. **74** (4), 737−771 (2022).

36. S. Bouzebda, I. Elhattab, and A. A. Ferfache, "General M-estimator processes and their m out of n bootstrap with functional nuisance parameters," Methodol. Comput. Appl. Probab. **24** (4), 2961−3005 (2022b).

37. S. Bouzebda, A. Laksaci, and M. Mohammedi, "Single index regression model for functional quasi-associated time series data," REVSTAT **20** (5), 605−631 (2022c).

38. S. Bouzebda, A. Laksaci, and M. Mohammedi, "The k-nearest neighbors method in single index regression model for functional quasi-associated time series data," Rev. Mat. Complut. **36** (2), 361−391 (2023).

39. J. E. Chacón and T. Duong, *Multivariate Kernel Smoothing and Its Applications*, Vol. 160: Monographs on Statistics and Applied Probability (CRC Press, Boca Raton, FL, 2018).

40. D. Chen, P. Hall, and H.-G. Müller, "Single and multiple index functional regression models with nonparametric link," Ann. Statist. **39** (3), 1720−1747 (2011).

41. M. Cherfi, "Large deviations theorems in nonparametric regression on functional data," C. R. Math. Acad. Sci. Paris **349** (9−10), 583−585 (2011).

42. H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," Ann. Math. Statistics **23**, 493−507 (1952).

43. K. Chokri and S. Bouzebda, "Uniform-in-bandwidth consistency results in the partially linear additive model components estimation," Comm. Statist. Theory Methods, 1−42 (2023).

44. J. A. Clarkson and C. R. Adams, "On definitions of bounded variation for functions of two variables," Trans. Amer. Math. Soc. **35** (4), 824−854 (1933).

45. S. Dabo-Niang and A. Laksaci, "Nonparametric quantile regression estimation for functional dependent data," Comm. Statist. Theory Methods **41** (7), 1254−1268 (2012).

46. P. Deheuvels, "One bootstrap suffices to generate sharp uniform bounds in functional estimation," Kybernetika (Prague) **47** (6), 855−865 (2011).

47. P. Deheuvels, "Uniform-in-bandwidth functional limit laws for multivariate empirical processes," in: *High dimensional probability VIII−The Oaxaca volume*, Vol. 74 of Progr. Probab. (Birkhäuser/Springer, Cham, 2019), pp. 201−239.

48. P. Deheuvels and D. M. Mason, "General asymptotic confidence bands based on kernel-type function estimators," Stat. Inference Stoch. Process. **7** (3), 225−277 (2004).

49. A. Dembo and O. Zeitouni, *Large Deviations Techniques and Applications*, Vol. 38: Applications of Mathematics, 2nd ed. (Springer-Verlag, New York, 1998).

50. J.-D. Deuschel and D. W. Stroock, *Large Deviations*, Vol. 137: Pure and Applied Mathematics (Academic Press, Inc., Boston, MA, 1989).

51. L. Devroye and L. Györfi, *Nonparametric Density Estimation*. Wiley Series in Probability and Mathematical Statistics: Tracts on Probability and Statistics (John Wiley and Sons, Inc., New York, The L_1 View, 1985).

52. L. Devroye and G. Lugosi, *Combinatorial Methods in Density Estimation*, Springer Series in Statistics (Springer-Verlag, New York, 2001).

53. J. Dony and U. Einmahl, "Uniform in bandwidth consistency of kernel regression estimators at a fixed point," in: *High dimensional probability V: The Luminy volume*, Vol. 5: Inst. Math. Stat. (IMS) Collect. Inst. Math. Statist (Beachwood, OH, 2009), p. 308−325.

54. L. Douge, "Théorèmes limites pour des variables quasi-associées hilbertiennes," Ann. I.S.U.P. **54** (1−2), 51−60 (2010).

55. R. M. Dudley, *Uniform Central Limit Theorems*, Vol. 63: Cambridge Studies in Advanced Mathematics (Cambridge University Press, Cambridge, 1999).

56. N. Dunford and J. T. Schwartz, *Linear Operators, I: General Theory*, Pure and Applied Mathematics, Vol. 7 (Interscience Publishers, Inc., New York; Interscience Publishers Ltd., London. With the assistance of W. G. Bade and R. G. Bartle, 1958).

57. P. P. B. Eggermont and V. N. LaRiccia, *Maximum Penalized Likelihood Estimation, Vol. II* (Springer Series in Statistics. Springer, Dordrecht, 2009).

58. U. Einmahl and D. M. Mason, "An empirical process approach to the uniform consistency of kernel-type function estimators," J. Theoret. Probab. **13** (1), 1−37 (2000).

59. U. Einmahl and D. M. Mason, "Uniform in bandwidth consistency of kernel-type function estimators," Ann. Statist. **33** (3), 1380−1403 (2005).

60. R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, Classics in Mathematics (Springer-Verlag, Berlin, Reprint of the 1985 original, 2006).

61. F. Ferraty and P. Vieu, "Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés," C. R. Acad. Sci. Paris Sér. I Math. **330** (2), 139−142 (2000).

62. F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis*, Springer Series in Statistics (Springer, New York, Theory and Practice, 2006).

63. F. Ferraty, A. Laksaci, and P. Vieu, "Estimating some characteristics of the conditional distribution in nonparametric functional models," Stat. Inference Stoch. Process. **9** (1), 47−76 (2006).

64. F. Ferraty, A. Mas, and P. Vieu, "Nonparametric regression on functional data: Inference and practical aspects," Australian and New Zealand Journal of Statistics **49**, 267−286 (2007).

65. F. Ferraty, A. Laksaci, A. Tadj, and P. Vieu, "Rate of uniform consistency for nonparametric estimates with functional variables," J. Statist. Plann. Inference **140** (2), 335−352 (2010).

66. J. C. Fu, "Large sample point estimation: A large deviation theory approach," Ann. Statist. **10** (3), 762−771 (1982).

67. F. Gao and X. Zhao, "Delta method in large deviations and moderate deviations for estimators," Ann. Statist. **39** (2), 1211−1240 (2011).

68. T. Gasser, P. Hall, and B. Presnell, "Nonparametric estimation of the mode of a distribution of random curves," J. R. Stat. Soc. Ser. B Stat. Methodol. **60** (4), 681−691 (1998).

69. I. Gijbels, M. Omelka, and N. Veraverbeke, "Multivariate and functional covariates and conditional copulas," Electron. J. Stat. **6**, 1273−1306 (2012).

70. R. D. Gill, "Non- and semi-parametric maximum likelihood estimators and the von Mises method. I," Scand. J. Statist. **16** (2), 97−128. With a discussion by J. A. Wellner and J. Præstgaard and a reply by the author (1989).

71. E. Giné and A. Guillou, "On consistency of kernel density estimators for randomly censored data: Rates holding uniformly over adaptive intervals," Ann. Inst. H. Poincaré Probab. Statist. **37** (4), 503−522 (2001).

72. E. Giné and R. Nickl, *Mathematical Foundations of Infinite-Dimensional Statistical Models*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, New York, 2016).

73. E. Giné, V. Koltchinskii, and J. Zinn, "Weighted uniform consistency of kernel density estimators," Ann. Probab. **32** (3B), 2570−2605 (2004).

74. A. Goia and P. Vieu, "A partitioned single functional index model," Comput. Statist. **30** (3), 673−692 (2015).

75. P. Groeneboom, J. Oosterhoff, and F. H. Ruymgaart, "Large deviation theorems for empirical probability measures," Ann. Probab. **7** (4), 553−586 (1979).

76. L. Györfi, M. Kohler, A. Krzyżak, and H. Walk, *A Distribution-Free Theory of Nonparametric Regression*, Springer Series in Statistics (Springer-Verlag, New York, 2002).

77. P. Hall, "Asymptotic properties of integrated square error and cross-validation for kernel estimation of a regression function," Z. Wahrsch. Verw. Gebiete **67** (2), 175−196 (1984).

78. W. Härdle, *Applied Nonparametric Regression*, Vol. 19: Econometric Society Monographs (Cambridge University Press, Cambridge, 1990).

79. W. Härdle and J. S. Marron, "Optimal bandwidth selection in nonparametric regression function estimation," Ann. Statist. **13** (4), 1465−1481 (1985).

80. G. H. Hardy, "On double fourier series and especially those which represent the double zeta-function with real and incommensurable parameters," Quart. J. Math **37** (1), 53−79 (1905).

81. E. W. Hobson, *The theory of functions of a real variable and the theory of Fourier's series,* Vol. II (Dover Publications, Inc., New York, N.Y., 1958).

82. L. Horváth and P. Kokoszka, *Inference for Functional Data with Applications* (Springer Series in Statistics. Springer, New York, 2012).

83. P. J. Huber, "Robust estimation of a location parameter," Ann. Math. Statist. **35**, 73−101 (1964).

84. W. C. M. Kallenberg, "Chernoff efficiency and deficiency," Ann. Statist. **10** (2), 583−594 (1982).

85. W. C. M. Kallenberg, "Intermediate efficiency, theory and examples," Ann. Statist. **11** (1), 170−182 (1983a).

86. W. C. M. Kallenberg, "On moderate deviation theory in estimation," Ann. Statist. **11** (2), 498−504 (1983b).

87. L.-Z. Kara, A. Laksaci, M. Rachdi, and P. Vieu, "Data-driven kNN estimation in nonparametric functional data analysis," J. Multivariate Anal. **153**, 176−188 (2017).

88. A. D. M. Kester and W. C. M. Kallenberg, "Large deviations of estimators," Ann. Statist. **14** (2), 648−664 (1986).

89. M. R. Kosorok, *Introduction to Empirical Processes and Semiparametric Inference* (Springer Series in Statistics, Springer, New York, 2008).

90. M. Krause, "Über Mittelwertsätze im Gebiete der Doppelsummen und Doppelintegrale," Leipz. Ber. **55**, 239−263 (1903).

91. W. V. Li and Q.-M. Shao, *Gaussian Processes: Inequalities, Small Ball Probabilities, and Applications*, in: Stochastic Processes: Theory and Methods, Vol. 19: Handbook of Statist. (North-Holland, Amsterdam, 2001), p. 533−597.

92. H. Lian, "Functional partial linear model," J. Nonparametr. Stat. **23** (1), 115−128 (2011).

93. N. Ling and P. Vieu, "Nonparametric modelling for functional data: Selected survey and tracks for future," Statistics **52** (4), 934−949 (2018).

94. Q. Liu and S. Zhao, "Pointwise and uniform moderate deviations for nonparametric regression function estimator on functional data," Statist. Probab. Lett. **83** (5), 1372−1381 (2013).

95. D. Louani and S. M. Ould Maouloud, "Large deviation results for the nonparametric regression function estimator on functional data," Math. Methods Statist. **21** (4), 298−313 (2012).

96. D. M. Mason, "Proving consistency of non-standard kernel estimators," Stat. Inference Stoch. Process. **15** (2), 151−176 (2012).

97. D. M. Mason and M. A. Newton, "A rank statistics approach to the consistency of a general bootstrap," Ann. Statist. **20** (3), 1611−1624 (1992).

98. D. M. Mason and J. W. H. Swanepoel, "A general result on the uniform in bandwidth consistency of kernel-type function estimators," TEST **20** (1), 72−94 (2011).

99. D. M. Mason and J. W. H. Swanepoel, "Uniform in bandwidth consistency of kernel estimators of the density of mixed data," Electron. J. Stat. **9** (1), 1518−1539 (2015).

100. E. Masry, "Nonparametric regression estimation for dependent functional data: asymptotic normality," Stochastic Process. Appl. **115** (1), 155−177 (2005).

101. E. Mayer-Wolf and O. Zeitouni, "The probability of small Gaussian ellipsoids and associated conditional moments," Ann. Probab. **21** (1), 14−24 (1993).

102. M. Mohammedi, S. Bouzebda, and A. Laksaci, "On the nonparametric estimation of the functional expectile regression," C. R. Math. Acad. Sci. Paris **358** (3), 267−272 (2020).

103. M. Mohammedi, S. Bouzebda, and A. Laksaci, "The consistency and asymptotic normality of the kernel type expectile regression estimator for functional data," J. Multivariate Anal. **181** (104673), 24 (2021).

104. A. Mokkadem and M. Pelletier, "Moderate deviations principles for the kernel estimator of nonrandom regression functions," Afr. Stat. **11** (2), 995−1021 (2016).

105. A. Mokkadem, M. Pelletier, and B. Thiam, "Large and moderate deviations principles for kernel estimators of the multivariate regression," Math. Methods Statist. **17** (2), 146−172 (2008).

106. H.-G. Müller, *Nonparametric Regression Analysis of Longitudinal Data*, Vol. 46: Lecture Notes in Statistics (Springer-Verlag, Berlin, 1988).

107. E. A. Nadaraja, "On a regression estimate," Teor. Verojatnost. i Primenen. **9**, 157−159 (1964).

108. E. A. Nadaraja, *Nonparametric Estimation of Probability Densities and Regression Curves*, Vol. 20: Mathematics and Its Applications (Soviet Series) (Kluwer Academic Publishers Group, Dordrecht, 1989).

109. R. B. Nelsen, *An introduction to copulas*. Springer Series in Statistics (Springer, New York, 2nd ed., 2006).

110. W. K. Newey and J. L. Powell, "Asymmetric least squares estimation and testing," Econometrica **55** (4), 819−847 (1987).

111. H. Niederreiter, *Random Number Generation and Quasi-Monte-Carlo Methods*, Vol. 63: CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM) (Philadelphia, PA, 1992).

112. Y. Nikitin, *Asymptotic Efficiency of Nonparametric Tests* (Cambridge University Press, Cambridge, 1995).

113. D. Nolan and D. Pollard, "U-processes: Rates of convergence," Ann. Statist. **15** (2), 780−799 (1987).

114. S. M. Ould Maouloud, "Some uniform large deviation results in nonparametric function estimation," J. Nonparametr. Stat. **20** (2), 129−152 (2008).

115. E. Parzen, "On estimation of a probability density function and mode," Ann. Math. Statist. **33**, 1065−1076 (1962).

116. D. Pollard, *Convergence of Stochastic Processes*, Springer Series in Statistics (Springer-Verlag, New York, 1984).

117. A. Puhalskii and V. Spokoiny, "On large-deviation efficiency in statistical inference," Bernoulli **4** (2), 203−272 (1998).

118. M. Rachdi and P. Vieu, "Nonparametric regression for functional data: Automatic smoothing parameter selection," J. Statist. Plann. Inference **137** (9), 2784−2801 (2007).

119. M. E. Radavichyus, "Probabilities of large deviations for maximum likelihood estimators," Dokl. Akad. Nauk SSSR **268** (3), 551−556 (1983).

120. J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, Springer Series in Statistics (Springer, New York, 2nd ed., 2005).

121. G. G. Roussas, "Nonparametric estimation of the transition distribution function of a Markov process," Ann. Math. Statist. **40**, 1386−1400 (1969).

122. M. Samanta, "Nonparametric estimation of conditional quantiles," Statist. Probab. Lett. **7** (5), 407−412 (1989).

123. I. N. Sanov, "On the probability of large deviations of random magnitudes," Mat. Sb. (N.S.) **42(84)**, 11−44 (1957).

124. D. W. Scott, *Multivariate Density Estimation*, Wiley Series in Probability and Statistics (John Wiley and Sons, Inc., Hoboken, NJ, 2nd ed., 2015).

125. A. Sieders and K. Dzhaparidze, "A large deviation result for parameter estimators and its application to nonlinear regression analysis," Ann. Statist. **15** (3), 1031−1049 (1987).

126. B. W. Silverman, *Density Estimation for Statistics and Data Analysis*, Monographs on Statistics and Applied Probability (Chapman and Hall, London, 1986).

127. I. Soukarieh and S. Bouzebda, "Renewal type bootstrap for increasing degree U-process of a Markov chain," J. Multivariate Anal. **195** (105143), 25 (2023).

128. I. Soukarieh and S. Bouzebda, "Weak convergence of the conditional U-statistics for locally stationary functional time series," Stat. Inference Stoch. Process. 1−78 (2024).
https://doi.org/10.1007/s11203-023-09305-y

129. W. Stute, "Conditional empirical processes," Ann. Statist. **14** (2), 638−647 (1986).

130. R. A. Tapia and J. R. Thompson, *Nonparametric Probability Density Estimation*, Vol. 1: Johns Hopkins Series in the Mathematical Sciences (Johns Hopkins University Press, Baltimore, Md., 1978).

131. A. W. van der Vaart and J. A. Wellner, *Weak Convergence and Empirical Processes*, Springer Series in Statistics (Springer-Verlag, New York, 1996).

132. S. R. S. Varadhan, Large deviations and applications, in: École d'Été de Probabilités de Saint-Flour XV−XVII, 1985−1987, Vol. 1362: Lecture Notes in Math. (Springer, Berlin, 1988), pp. 1−49.

133. S. R. S. Varadhan, *Large Deviations*, Vol. 27: Courant Lecture Notes in Mathematics (Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2016).

134. G. Vitali, "Sui gruppi di punti e sulle funzioni di variabili reali," Torino Atti **43**, 229−246 (1908).

135. A. G. Vituškin, *O mnogomernykh variatsiyakh* (Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow, 1955).

136. M. P. Wand and M. C. Jones, *Kernel Smoothing*, Vol. 60: Monographs on Statistics and Applied Probability (Chapman and Hall, Ltd., London, 1995).

137. G. S. Watson, "Smooth regression analysis," Sankhyā Ser. A **26**, 359−372 (1964).

138. W. Wertz, *Statistical Density Estimation: A Survey*, Vol. 13: Angewandte Statistik und Ökonometrie [Applied Statistics and Econometrics] (Vandenhoeck and Ruprecht, Göttingen, With German and French Summaries, 1978).

139. C. Wu, N. Ling, P. Vieu, and W. Liang, "Partially functional linear quantile regression model and variable selection with censoring indicators MAR," J. Multivariate Anal. **197**, Paper No. 105189 (2023).