Estimation of a regular conditional functional by conditional U-statistic regression

Derumigny, Alexis

DOI
10.1111/stan.12350

Publication date
2024

Document Version
Final published version

Published in
Statistica Neerlandica

Received: 7 June 2024 | Accepted: 12 June 2024
DOI: 10.1111/stan.12350

ORIGINAL ARTICLE

WILEY

Estimation of a regular conditional functional by conditional U-statistic regression

Alexis Derumigny

Department of Applied Mathematics, Delft University of Technology, CD Delft, The Netherlands

Correspondence
Alexis Derumigny, Department of Applied Mathematics, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands.
Email: A.F.F.Derumigny@tudelft.nl

Funding information
Labex Ecodec, French Agence Nationale de la Recherche, Grant/Award Number: ANR-11-LABEX-0047

U-statistics constitute a large class of estimators, generalizing the empirical mean of a random variable $\mathbf{X}$ to sums over every k-tuple of distinct observations of $\mathbf{X}$. They may be used to estimate a regular functional $\theta(\mathbb{P}_{\mathbf{X}})$ of the law of $\mathbf{X}$. When a vector of covariates $\mathbf{Z}$ is available, a conditional U-statistic describes the effect of $\mathbf{z}$ on the conditional law of $\mathbf{X}$ given $\mathbf{Z} = \mathbf{z}$, by estimating a regular conditional functional $\theta(\mathbb{P}_{\mathbf{X}|\mathbf{Z}=\cdot})$. We state nonasymptotic bounds of general conditional U-statistics and study their asymptotics too. Assuming a parametric model of the conditional functional of interest, we propose a regression-type estimator based on conditional U-statistics. Its theoretical properties are derived, first in a nonasymptotic framework and then in two different asymptotic regimes. Some examples are given to illustrate our methods.

KEYWORDS
conditional distribution, penalized regression, regression-type models, U-statistics

1 | INTRODUCTION

1.1 | Our setting

Let $\mathbf{X}$ be a random element with values in a measurable space $(\mathcal{X}, \mathcal{A})$, and denote by $\mathbb{P}_{\mathbf{X}}$ its law. Often, we will use as a typical example the special case $\mathcal{X} = \mathbb{R}^{p_{\mathbf{X}}}$, for a fixed dimension $p_{\mathbf{X}} > 0$.

In statistics, many parameters of interest can be seen as a regular functional of $\mathbb{P}_{\mathbf{X}}$. A functional $\mathbb{P}_{\mathbf{X}} \mapsto \theta(\mathbb{P}_{\mathbf{X}})$ is called *regular* if it is of the form

$$\theta(\mathbb{P}_{\mathbf{X}}) = \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k)\right] = \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) d\mathbb{P}_{\mathbf{X}}(\mathbf{x}_1) \dots d\mathbb{P}_{\mathbf{X}}(\mathbf{x}_k), \tag{1}$$

for a fixed $k > 0$, a function $g : \mathcal{X}^k \to \mathbb{R}$ and $\mathbf{X}_1, \dots \mathbf{X}_k \overset{\text{i.i.d.}}{\sim} \mathbb{P}_{\mathbf{X}}$. Following Hoeffding (1948), a natural estimator of $\theta(\mathbb{P}_{\mathbf{X}})$ is the U-statistics $\hat{\theta}$, defined by

$$\hat{\theta} := |\mathfrak{I}_{k,n}|^{-1} \sum_{\sigma \in \mathfrak{I}_{k,n}} g\left(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}\right),$$

where $\mathfrak{I}_{k,n}$ is the set of injective functions from $\{1, \dots, k\}$ to $\{1, \dots, n\}$. Remark that $\mathfrak{I}_{k,n}$ is isomorph to the set of variations of k distinct elements of $\{1, \dots, n\}$. Therefore, $|\mathfrak{I}_{k,n}| = \binom{n}{k} k! = n(n-1) \dots (n-k+1)$. For an introduction to the theory of U-statistics, we refer to Korolyuk and Borovskich (1994) and (Serfling, 1980, chapter 5).

In our framework, we assume that we are actually interested in a random pair $(\mathbf{X}, \mathbf{Z})$ where $\mathbf{Z}$ is a p-dimensional covariate. For this, we assume that we observe n i.i.d. pairs $(\mathbf{X}_i, \mathbf{Z}_i)$, $i = 1, \dots, n$. We want to use the information of the covariate $\mathbf{Z}$ to obtain knowledge on the conditional distribution of $\mathbf{X}$. In other words, the object of interest is the mapping $\mathbf{z} \in \mathbb{R}^p \mapsto \mathbb{P}_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}$. This mapping will be denoted by $\mathbb{P}_{\mathbf{X}|\mathbf{Z}=\cdot}$.

> **Assumption 1.** As obtaining results valid uniformly over $\mathbb{R}^p$ may not be feasible, we restrict ourselves in this article to the inference of the mapping $\mathbf{z} \in \mathcal{Z} \mapsto \mathbb{P}_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}$, for some compact subset $\mathcal{Z}$ of $\mathbb{R}^p$.

Extending the unconditional framework presented above to the conditional framework, we say that θ is a *regular conditional functional* if it is of the form

$$\begin{aligned}
\theta(\mathbf{z}_1, \dots, \mathbf{z}_k)(\mathbb{P}_{\mathbf{X}|\mathbf{Z}=\cdot}) &:= \theta(\mathbb{P}_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_1}, \dots, \mathbb{P}_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_k}) \\
&= \mathbb{E}_{\otimes_{i=1}^k \mathbb{P}_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_i}} \left[g(\mathbf{X}_1, \dots, \mathbf{X}_k)\right] = \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_i = \mathbf{z}_i, \forall i = 1, \dots, k\right] \\
&= \int g(\mathbf{x}_1, \dots, \mathbf{x}_k) d\mathbb{P}_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_1}(\mathbf{x}_1) \cdots d\mathbb{P}_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_k}(\mathbf{x}_k).
\end{aligned}$$

This can be seen as a generalization of $\theta(\mathbb{P}_{\mathbf{X}})$ to the conditional case. Indeed, when $\mathbf{X}$ and $\mathbf{Z}$ are independent, the new functional $\theta(\mathbf{z}_1, \dots, \mathbf{z}_k)(\mathbb{P}_{\mathbf{X}|\mathbf{Z}=\cdot})$ is equal to the unconditional functional $\theta(\mathbb{P}_{\mathbf{X}})$. For convenience, we will use the notation $\theta(\mathbf{z}_1, \dots, \mathbf{z}_k) := \theta(\mathbf{z}_1, \dots \mathbf{z}_k)(\mathbb{P}_{\mathbf{X}|\mathbf{Z}=\cdot})$, treating the law of $(\mathbf{X}, \mathbf{Z})$ as fixed (but unknown).

The goal of this paper is to study estimators of $\theta(\mathbf{z}_1, \dots, \mathbf{z}_k)$, first for fixed $\mathbf{z}_1, \dots, \mathbf{z}_k$, and then as a function $\theta : \mathcal{Z} \to \mathbb{R}$, given a parametric form.

Stute (1991) defined a kernel-based estimator $\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)$ of the conditional functional $\theta(\mathbf{z}_1, \dots, \mathbf{z}_k)$ by

$$\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k) := \frac{\sum_{\sigma \in \mathfrak{I}_{k,n}} g\left(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}\right) \prod_{j=1}^k K_h\left(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}\right)}{\sum_{\sigma \in \mathfrak{I}_{k,n}} \prod_{j=1}^k K_h\left(\mathbf{z}_j - \mathbf{Z}_{\sigma(j)}\right)}, \tag{2}$$

where $h > 0$ is the bandwidth, $K(\cdot)$ a kernel on $\mathbb{R}^p$, $K_h(\cdot) := h^{-p}K(\cdot/h)$, and $(\mathbf{X}_i, \mathbf{Z}_i) \overset{\text{i.i.d.}}{\sim} \mathbb{P}_{\mathbf{X},\mathbf{Z}}$. Stute (1991) proved the asymptotic normality of $\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)$ and its weak and strong consistency. Dony and Mason (2008) derived its uniform in bandwidth consistency under VC-type conditions over a class of possible functions g.

1.2 | Two-step estimation of $\theta(\mathbf{z}_1, \dots, \mathbf{z}_k)$ based on a parametric model

The estimator (2) has several weaknesses. First, the interpretation of the whole hypersurface $(\mathbf{z}_1, \dots, \mathbf{z}_k) \mapsto \hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)$ can be difficult. Indeed, the latter curve is of dimension $1 + p \times k$, and it is rather challenging to visualize it even for small values of p and k. Second, for each new k-tuples $(\mathbf{z}_1, \dots, \mathbf{z}_k)$, the computation of $\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)$ has a cost of $O(n^k)$. Then, if we want to estimate N values of the conditional functional at N new conditioning tuples, that is, we want $\hat{\theta}(\mathbf{z}_1^{(i)}, \dots, \mathbf{z}_k^{(i)}), i = 1, \dots, N$, where $\left(\mathbf{z}_1^{(1)}, \dots, \mathbf{z}_k^{(1)}, \dots, \mathbf{z}_1^{(N)}, \dots, \mathbf{z}_k^{(N)}\right) \in \mathcal{Z}^{k\times N}$, then the total computational cost of that estimation by (2) is $O(Nn^k)$. Third, it is well-known that kernel estimators are not very smooth, in the sense that they usually present many spurious local minima and maxima, and this can be a problem in some applications. Therefore, we may want to build estimators which are more regular with respect to the conditioning variables $\mathbf{z}_1, \dots \mathbf{z}_k$, and have a simple functional form.

To do so, we decompose the function $(\mathbf{z}_1, \dots, \mathbf{z}_k) \mapsto \theta(\mathbf{z}_1, \dots, \mathbf{z}_k)$ on a basis $(\psi_i)_{i\geq 0}$, generalizing the work of Derumigny and Fermanian (2020). This may not be always easy if the range of the function $\theta(\cdot, \cdots, \cdot)$ is a strict subset of $\mathbb{R}$. In that case, it is always possible to use a "link function" Λ, that would be strictly increasing and continuously differentiable and such that the range $\Lambda\circ\theta(\cdot, \cdots, \cdot)$ is exactly $\mathbb{R}$. Whatever the choice of Λ (including the identity function), we can decompose the latter function on any basis $(\psi_i)_{i\geq 0}$. If only a finite number $r > 0$ of elements of this basis are necessary to represent the whole function $\Lambda\circ\theta(\cdot, \cdots, \cdot)$ over $\mathcal{Z}^k$, then we have the following parametric model:

$$\forall(\mathbf{z}_1, \dots, \mathbf{z}_k) \in \mathcal{Z}^k,\ \Lambda\left(\theta(\mathbf{z}_1, \dots, \mathbf{z}_k)\right) = \boldsymbol{\psi}(\mathbf{z}_1, \dots, \mathbf{z}_k)^T\beta^*, \tag{3}$$

where $\beta^* \in \mathbb{R}^r$ is the true parameter and $\boldsymbol{\psi}(\cdot) := (\psi_1(\cdot), \dots, \psi_r(\cdot))^T \in \mathbb{R}^r$. In most applications, finding an appropriate basis $\boldsymbol{\psi}$ is not easy. This will depend on the choice of the (conditional) functional θ. Therefore, the most simple solution consists in choosing a concatenation of several well-known basis such as polynomials, exponentials, sinuses and cosinuses, indicator functions, etc. They allow to take into account potential nonlinearities and even discontinuities of the function $\Lambda\circ\theta(\cdot, \cdots, \cdot)$. For the sake of inference, a necessary condition is the linear independence of such functions, as seen in the following proposition (whose straightforward proof is omitted). We say that Model (3) is identifiable if for all pairs of distribution $\mathbb{P}_{\mathbf{X},\mathbf{Z}}$, $\widetilde{\mathbb{P}}_{\mathbf{X},\mathbf{Z}}$ on $\mathcal{X} \times \mathcal{Z}$ satisfying Model (3) with corresponding parameters $\beta, \widetilde{\beta} \in \mathbb{R}^r$, $\beta \neq \widetilde{\beta}$ implies that $\mathbb{P}_{\mathbf{X},\mathbf{Z}} \neq \widetilde{\mathbb{P}}_{\mathbf{X},\mathbf{Z}}$. Conversely, Model (3) is not identifiable if the same distribution $\mathbb{P}_{\mathbf{X},\mathbf{Z}}$ can be represented by two different vectors β^*.

Proposition 1. *The parameter β^* is identifiable in Model (3) if and only if the functions $(\psi_1(\cdot), \dots, \psi_r(\cdot))$ are linearly independent $\mathbb{P}_Z^{\otimes k}$-almost everywhere in the sense that, for all vectors $\mathbf{t} = (t_1, \dots, t_r) \in \mathbb{R}^r$, $\mathbb{P}_Z^{\otimes k}\left(\boldsymbol{\psi}(Z_1, \dots, Z_k)^T\mathbf{t} = 0\right) = 1 \Rightarrow \mathbf{t} = 0$.*

With such a choice of a wide and flexible class of functions, it is likely that not all these functions are relevant. This is know as sparsity, that is, the number of nonzero coefficients of β^*, denoted by $|\mathcal{S}| = |\beta^*|_0$ is less than s, for some $s \in \{1, \dots, r\}$. Here, $|\cdot|_0$ denotes the number of nonzero components of a vector of $\mathbb{R}^r$ and $\mathcal{S}$ is the set of nonzero components of β^*. Note that, in this framework, r can be moderately large, for example 30 or 50, while the original dimension p is small, for example $p = 1$ or 2. This corresponds to the decomposition of a function, defined on a small-dimension domain, in a mildly large basis.

Remark 1. At first sight, in Model (3), there seems to be no noise perturbing the variable of interest. In fact, this can be seen as a simple consequence of our formulation of the model. In the same way, the classical linear model $Y = \mathbf{X}^T\beta^* + \varepsilon$ can be rewritten as $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}] = \mathbf{x}^T\beta^*$ without any explicit noise. By definition, $\mathbb{E}[Y|\mathbf{X} = \mathbf{x}]$ is a deterministic function of a given $\mathbf{x}$. In our case, the corresponding fact is: $\Lambda(\theta(\mathbf{z}_1, \dots, \mathbf{z}_k))$ is a deterministic function of the variables $(\mathbf{z}_1, \dots, \mathbf{z}_k)$. This means that we cannot write formally a model with noise, such as $\Lambda(\theta(\mathbf{z}_1, \dots, \mathbf{z}_k)) = \boldsymbol{\psi}(\mathbf{z}_1, \dots, \mathbf{z}_k)^T\beta^* + \varepsilon$ where ε is independent of the choice of $(\mathbf{z}_1, \dots, \mathbf{z}_k)$ since the left-hand side of the latter equality is a $(\mathbf{z}_1, \dots, \mathbf{z}_k)$-measurable quantity, unless ε is constant almost surely.

Contrary to more usual models, the explained variable $\Lambda(\theta(\mathbf{z}_1, \dots, \mathbf{z}_k))$, is not observed in Model (3). Therefore, a direct estimation of the parameter β^* (e.g., by the ordinary least squares, or by the Lasso) is unfeasible. In other words, even if the function $(\mathbf{z}_1, \dots, \mathbf{z}_k) \mapsto \Lambda(\theta(\mathbf{z}_1, \dots, \mathbf{z}_k))$ is deterministic (by definition of conditional probabilities), finding the best β in Model (3) is far from being a numerical analysis problem since the function to be decomposed is unknown. Nevertheless, we will replace $\Lambda(\theta(\mathbf{z}_1, \dots, \mathbf{z}_k))$ by the nonparametric estimate $\Lambda\left(\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)\right)$, and use it as an approximation of the explained variable.

More precisely, in our setting, the statistician chooses a finite collection of points $\mathbf{z}'_1, \dots, \mathbf{z}'_{n'} \in \mathcal{Z}^{n'}$ used as reference values, and a collection $\mathfrak{I}_{k,n'}$ of injective functions $\sigma : \{1, \dots, k\} \to \{1, \dots, n'\}$. Note that we are not forced to include *all* the injective functions in $\mathfrak{I}_{k,n'}$, reducing its number of elements. This will allow us to decrease the computational cost of the procedure. For every $\sigma \in \mathfrak{I}_{k,n'}$, we estimate $\hat{\theta}(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)})$. Finally, the estimator $\hat{\beta}$ is defined as the minimizer of the following l_1-penalized criteria

$$\hat{\beta} := \arg\min_{\beta \in \mathbb{R}^r} \left[\frac{(n'-k)!}{n'!} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \left(\Lambda\left(\hat{\theta}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)\right) - \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)^T \beta \right)^2 + \lambda|\beta|_1 \right], \quad (4)$$

where λ is a positive tuning parameter (that may depend on n and n'), and $|\cdot|_q$ denotes the l_q norm, for $1 \le q \le \infty$. Note that the first term in Equation (4) is an incomplete U-statistics since only the injective functions that belong to $\mathfrak{I}_{k,n'}$ are used in order to reduce the computational cost. This procedure is summed up in the following Algorithm 1. Note that even if we study the general case with any $\lambda \ge 0$, the corresponding properties of the unpenalized estimator can be derived by choosing the particular case $\lambda = 0$.

Once an estimator $\hat{\beta}$ of β^* has been computed, the prediction of all the conditional functionals is reduced to the computation of $\Lambda^{(-1)}\left(\boldsymbol{\psi}(\mathbf{z}_1^{(i)}, \dots, \mathbf{z}_k^{(i)})^T\hat{\beta}\right) := \tilde{\theta}(\mathbf{z}_1^{(i)}, \dots, \mathbf{z}_k^{(i)})$, for every $i = 1, \dots, N$. The total computational cost of this new method is therefore $O(|\mathfrak{I}_{k,n'}|n'^k + |\mathfrak{I}_{k,n'}|r + Ns)$ operations. The first term corresponds to the cost of evaluating each nonparametric estimator (2). The second term corresponds to the minimization of the convex optimization program (4),

Algorithm 1. Two-step estimation of β and prediction of the conditional parameters $\theta(\mathbf{z}_1^{(i)}, \dots, \mathbf{z}_k^{(i)})$, for $i = 1, \dots, N$

Input: A dataset $(X_{i,1}, X_{i,2}, \mathbf{Z}_i)$, $i = 1, \dots, n$
Input: A finite collection of points $\mathbf{z}'_1, \dots, \mathbf{z}'_{n'} \in \mathcal{Z}^{n'}$, selected for estimation
Input: A collection of N k-tuples for prediction $\left(\mathbf{z}_1^{(1)}, \dots, \mathbf{z}_k^{(1)}, \dots, \mathbf{z}_1^{(N)}, \dots, \mathbf{z}_k^{(N)}\right) \in \mathcal{Z}^{k \times N}$
for $\sigma \in \mathfrak{I}_{k,n'}$ **do**
  Compute the estimator $\hat{\theta}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)$ using the sample $(\mathbf{X}_i, \mathbf{Z}_i)$, $i = 1, \dots, n$
end
Compute the minimizer $\hat{\beta}$ of (4) using the $\hat{\theta}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)$, $j = 1, \dots, n'$, estimated in the above step **for** $i \leftarrow 1$ **to** N **do**
  Compute the prediction $\tilde{\theta}(\mathbf{z}_1^{(i)}, \dots, \mathbf{z}_k^{(i)}) := \Lambda^{(-1)}\left(\boldsymbol{\psi}(\mathbf{z}_1^{(i)}, \dots, \mathbf{z}_k^{(i)})^T \hat{\beta}\right)$
end
Output: An estimator $\hat{\beta}$ and N predictions $\tilde{\theta}(\mathbf{z}_1^{(i)}, \dots, \mathbf{z}_k^{(i)})$, $i = 1, \dots, N$.

and the last one is the prediction cost. Note that the procedure described in Algorithm 1 can provide a huge improvement compared to the previously available estimator with a cost in $O(Nn^k)$ when $N \to \infty$, i.e. when we want to recover the full function $\theta(\cdot, \cdots, \cdot)$. Moreover, the speed-up given by Algorithm 1 compared to the original conditional U-statistics (2) even increases with the sample size n, for moderate choices of n'.

A similar model, called *functional response*, has already been studied: see, for example, Kowalski and Tu (2008, chapter 6.2). They provide a method to estimate the parameter β^*, using generalized estimating equations. However, they only provide asymptotic results for their estimator, and their algorithm needs to solve a multidimensional equation which has no reason to be convex.

In Section 2, we provide nonasymptotic bounds for the nonparametric estimator $\hat{\theta}$. Then Section 3 is devoted to the statement of corresponding bounds, as well as asymptotic properties for the parametric estimator $\hat{\beta}$. Finally, a few examples are presented in Section 4. All proofs have been postponed to the Appendix.

2 | THEORETICAL PROPERTIES OF THE NONPARAMETRIC ESTIMATOR $\hat{\theta}(\cdot)$

2.1 | Nonasymptotic bounds for N_k

We remark that the estimator $\hat{\theta}$ is well-defined if and only if $N_k(\mathbf{z}_1, \dots, \mathbf{z}_k) > 0$, where

$$N_k(\mathbf{z}_1, \dots, \mathbf{z}_k) := \frac{k!(n-k)!}{n!} \sum_{\sigma \in \mathfrak{I}^{\uparrow}_{k,n}} K_h\left(\mathbf{Z}_{\sigma(1)} - \mathbf{z}_1\right) \dots K_h\left(\mathbf{Z}_{\sigma(k)} - \mathbf{z}_k\right). \tag{5}$$

To prove that our estimator $\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)$ exists with a probability that tends to 1, we will therefore study the behavior of N_k. We will need the following assumptions to control the kernel K and the density of $\mathbf{Z}$.

Assumption 2. The kernel $K(\cdot)$ is bounded, that is, there exists a finite constant C_K such that $K(\cdot) \leq C_K$ and $\int K(\mathbf{u})d\mathbf{u} = 1$. The kernel is of order α for some $\alpha > 0$, that is, for all $j = 1, \dots, \alpha - 1$ and all $1 \leq i_1, \dots, i_\alpha \leq p$, $\int K(\mathbf{u})\, u_{i_1} \dots u_{i_j} d\mathbf{u} = 0$.

Assumption 3. $f_{\mathbf{Z}}$ is α-times continuously differentiable on $\mathcal{Z}$, and there exists finite constants, $C_{f,\alpha}$, $C_{K,\text{int}}$ such that $\forall i = 1, \dots, \alpha$, $\forall j_1, \dots, j_i \in \{1, \dots, p\}$, $\sup_{\mathbf{z}} \left| \frac{\partial^i f_{\mathbf{Z}}}{\partial z_{j_1} \cdots \partial z_{j_i}} (\mathbf{z}) \right| \leq C_{f,\alpha}$ and $\int \left| K(\mathbf{u})\, u_{j_1} \dots u_{j_i} \right| d\mathbf{u} \leq C_{K,\text{int}}$.

Assumption 4. $f_{\mathbf{Z}}(\cdot) \leq f_{\mathbf{Z},\max}$ for some finite constant $f_{\mathbf{Z},\max}$.

Lemma 1. *Under Assumptions 2, 3, and 4, we have for any $t > 0$,*

$$\mathbb{P}\left(\left| N_k(\mathbf{z}_1, \dots, \mathbf{z}_k) - \prod_{i=1}^{k} f_{\mathbf{Z}}(\mathbf{z}_i) \right| \leq \frac{C_{K,\alpha}}{\alpha!} h^\alpha + t \right) \geq 1 - 2\exp\left(-\frac{[n/k]t^2}{h^{-kp}C_1 + h^{-kp}C_2 t} \right),$$

where $C_1 := 2f^k_{\mathbf{Z},\max}||K||_2^{2k}$, $C_2 := (4/3)C_K^k$, $||K||_2^2 := \int K^2$ and $C_{K,\alpha} := C^k_{K,\text{int}} C^k_{f,\alpha} k^\alpha p^\alpha$.

This lemma is proved in Section D.1 under a weaker condition than Assumption 3, that in general leads to a better constant $C_{K,\alpha}$. More can be said if the density $f_{\mathbf{Z}}$ is bounded below. Therefore, we will use the following assumption.

Assumption 5. There exists a constant $f_{\mathbf{Z},\min} > 0$ such that for every $\mathbf{z} \in \mathcal{Z}$, $f_{\mathbf{Z}}(\mathbf{z}) > f_{\mathbf{Z},\min}$.

If for some $\epsilon > 0$, we have $C_{K,\alpha}h^\alpha/\alpha! + t \leq f_{\mathbf{Z},\min} - \epsilon$, then $\hat{f}(\mathbf{z}) \geq \epsilon > 0$ with probability larger than on the event whose probability is bound in Lemma 1. We should therefore choose the largest t possible, which yields the following corollary.

Corollary 1. *Under Assumptions 2–5, if $C_{K,\alpha}h^\alpha/\alpha! < f_{\mathbf{Z},min}$, then the random variable $N_k(\mathbf{z}_1, \dots, \mathbf{z}_k)$ is strictly positive with a probability larger than $1 - 2\exp\left(-\frac{[n/k]h^{kp}\left(f_{\mathbf{Z},min} - C_{K,\alpha}h^\alpha/\alpha!\right)^2}{C_1 + C_2\left(f_{\mathbf{Z},min} - C_{K,\alpha}h^\alpha/\alpha!\right)} \right)$, guaranteeing the existence of the estimator $\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)$ on this event.*

2.2 | Nonasymptotic bounds in probability for $\hat{\theta}$

In this section, we generalize the bounds given in Derumigny and Fermanian (2019) for the conditional Kendall's tau to any conditional U-statistics. To establish bounds on $\hat{\theta}$ for every fixed n, we will need some assumptions on the joint law of $(\mathbf{X}, \mathbf{Z})$.

Assumption 6. There exists a probability measure μ on $(\mathcal{X}, \mathcal{A})$ such that $\mathbb{P}_{\mathbf{X},\mathbf{Z}}$ is absolutely continuous with respect to $\mu \otimes \text{Leb}_p$, where Leb_p is the Lebesgue measure on $\mathbb{R}^p$.

Assumption 7. For every $\mathbf{x} \in \mathcal{X}$, $\mathbf{z} \mapsto f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})$ is differentiable almost everywhere up to the order $\alpha + 1$, and there exists finite constants $\widetilde{C}_{f,\alpha}$, $\widetilde{C}_{K,\text{int}}$ such that $\forall i = 1, \dots, \alpha + 1$, $\forall j_1, \dots, j_i \in \{1, \dots, p\}$, $\sup_{\mathbf{x},\mathbf{z}} \left| \frac{\partial^i f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \cdots \partial z_{j_i}} (\mathbf{x}, \mathbf{z}) \right| \leq C_{f,\alpha}$ and $\int \left| K(\mathbf{u})\, u_{j_1} \dots u_{j_i} \right| d\mathbf{u} \leq C_{K,\text{int}}$.

Assumption 8. There exists a constant C_g such that $||g||_\infty \leq C_g < +\infty$.

The following proposition is proved in Section D.2.

Proposition 2 (Exponential bound for the estimator $\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)$, with fixed $\mathbf{z}_1, \dots \mathbf{z}_k \in \mathcal{Z}^k$). *Under Assumptions 2–8, there exists positive constants $C_3, \dots, C_7$ such that, for every $t, t' > 0$ such that $C_{K,\alpha} h^\alpha / \alpha! + t < f_{\mathbf{Z},\min}/2$, we have*

$$\begin{aligned} &\mathbb{P}\left(\left| \hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k) - \theta(\mathbf{z}_1, \dots, \mathbf{z}_k) \right| < (1 + C_3 h^\alpha + C_4 t) \times \left(C_5 h^{k+\alpha} + t' \right) \right) \\ &\geq 1 - 2\exp\left(-\frac{[n/k] t^2 h^{kp}}{C_1 + C_2 t} \right) - 2\exp\left(-\frac{[n/k] t'^2 h^{kp}}{C_6 + C_7 t'} \right). \end{aligned} \tag{6}$$

The precise expression of the constants can be found in the Appendix, where a more general result is proved under weaker but more sophisticated assumptions. In particular, it is possible to weaken Assumption 8 and to obtain a similar constant for unbounded g under only (conditional) moment conditions on g.

3 | THEORETICAL PROPERTIES OF THE ESTIMATOR $\hat{\beta}$

Let us define the matrix $\mathbb{Z}$ of dimension $|\mathfrak{I}_{k,n'}| \times r$ by $[\mathbb{Z}']_{i,j} := \psi_j \left(\mathbf{z}'_{\sigma_i(1)}, \dots, \mathbf{z}'_{\sigma_i(k)} \right)$, where $1 \leq i \leq |\mathfrak{I}_{k,n'}|$, $1 \leq j \leq r$ and σ_i is the ith element of $\mathfrak{I}_{k,n'}$. The chosen order of $\mathfrak{I}_{k,n'}$ is arbitrary and has no impact in practice. In the same way, we define the vector $\mathbf{Y}$ of dimension $|\mathfrak{I}_{k,n'}|$ defined by $Y_i := \Lambda\left(\hat{\theta}\left(\mathbf{z}'_{\sigma_i(1)}, \dots, \mathbf{z}'_{\sigma_i(k)} \right) \right)$, such that the criterion (4) is in the standard Lasso form $\hat{\beta} := \arg\min_{\beta \in \mathbb{R}^r} \left[||\mathbf{Y} - \mathbb{Z}'\beta||^2 + \lambda |\beta|_1 \right]$, where for any vector $\mathbf{v}$ of size $|\mathfrak{I}_{k,n'}|$, its scaled norm is defined by $||\mathbf{v}|| := |\mathbf{v}|_2 / \sqrt{|\mathfrak{I}_{k,n'}|}$. Following Derumigny and Fermanian (2020), we define $\xi_{i,n}$, for $1 \leq i \leq |\mathfrak{I}_{k,n'}|$, by $\xi_{i,n} = \xi_{\sigma_i,n} := \Lambda\left(\hat{\theta}\left(\mathbf{z}'_{\sigma_i(1)}, \dots, \mathbf{z}'_{\sigma_i(k)} \right) \right) - \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma_i(1)}, \dots, \mathbf{z}'_{\sigma_i(k)} \right)^T \beta^*$.

3.1 | Nonasymptotic bounds on $\hat{\beta}$

We will also use the *Restricted Eigenvalue* (RE) condition, introduced by Bickel, Ritov, and Tsybakov (2009). For $c_0 > 0$ and $s \in \{1, \dots, p\}$, it is defined as follows:

RE(s, c_0) **condition:** *The design matrix $\mathbb{Z}'$ satisfies*

$$\kappa(s, c_0) := \min\left\{ \frac{||\mathbb{Z}'\delta||}{|\delta|_2} : \delta \neq 0,\ |\delta_{J_0^C}|_1 \leq c_0 |\delta_{J_0}|_1,\ J_0 \subset \{1, \dots, r\},\ |J_0| \leq s \right\} > 0.$$

Note that this condition is very mild, and is satisfied with a high probability for a large class of random matrices: see Bellec et al. (2018, section 8.1) for references and a discussion. A (strong) sufficient condition for the RE(p, c_0) condition to hold is the following: all the singular values of $\mathbb{Z}$ are positive. We refer to van de Geer and Bühlmann (2009) for a discussion of the relationship between the different assumptions used to prove bounds on the Lasso. We will also need the following regularity assumption on the function $\Lambda(\cdot)$.

Assumption 9. The function $\mathbf{z} \mapsto \boldsymbol{\psi}(\mathbf{z})$ are bounded on $\mathcal{Z}$ by a constant $C_{\boldsymbol{\psi}}$. Moreover, $\Lambda(\cdot)$ is continuously differentiable. Let $\mathcal{T}$ be the range of θ, from $\mathcal{Z}^k$ toward $\mathbb{R}$. On an open neighborhood of $\mathcal{T}$, the derivative of $\Lambda(\cdot)$ is bounded by a constant $C_{\Lambda'}$.

The following theorem is proved in Section D.3 where the expression for the constants $C_{6,\sigma}, C_{7,\sigma}$ is given.

Theorem 1. *Assume that Assumption 9 and Equation (6) hold and that the design matrix $\mathbb{Z}'$ satisfies the* RE$(s,3)$ *condition. Choose the tuning parameter as $\lambda = \gamma t$, with $\gamma \geq 4$ and $t > 0$, and assume that we choose h small enough such that*

$$h \leq \min\left(\left(\frac{f_{Z,min}\alpha!}{4\ C_{K,\alpha}}\right)^{1/\alpha}, \left(\frac{t}{2C_5C_8}\right)^{1/(k+\alpha)} \right), \tag{7}$$

where $C_8 := C_{\boldsymbol{\psi}} C_{\Lambda'} \left(1 + C_4 f_{\mathbf{Z},\min}/2\right)$. *Then, we have*

$$\begin{aligned}
&\mathbb{P}\left(||\mathbb{Z}'(\hat{\beta} - \beta^*)|| \leq \frac{4(\gamma+1)t\sqrt{s}}{\kappa(s,3)} \text{ and } |\hat{\beta} - \beta^*|_q \leq \frac{4^{2/q}(\gamma+1)ts^{1/q}}{\kappa^2(s,3)}, \text{ for every } 1 \leq q \leq 2 \right) \\
&\geq 1 - 2\sum_{\sigma \in \mathfrak{I}_{k,n'}} \left[\exp\left(-\frac{[n/k] f^2_{\mathbf{Z},min} h^{kp}}{16C_1 + 4C_2 f_{\mathbf{Z},min}} \right) + \exp\left(-\frac{[n/k] t^2 h^{kp}}{4C_8^2 C_{6,\sigma} + 2C_8 C_{7,\sigma} t} \right) \right],
\end{aligned} \tag{8}$$

for some constant $C_{6,\sigma}, C_{7,\sigma}$.

The latter theorem gives some bounds that hold in probability for the prediction error $||\mathbb{Z}'(\hat{\beta} - \beta^*)||_{n'}$ and for the estimation error $|\hat{\beta} - \beta^*|_q$ with $1 \leq q \leq 2$ under the specification (3). Note that the influence of n' and r is hidden through the Restricted Eigenvalue number $\kappa(s,3)$.

3.2 | Asymptotic properties of $\hat{\beta}$ when $n \to \infty$ and for fixed n'

In this part, n' is still assumed to be fixed and we state the consistency and the asymptotic normality of $\hat{\beta}$ as $n \to \infty$. Remember that the points $\mathbf{z}'_i$ are our design points. As above, we adopt a fixed design: the $\mathbf{z}'_i$ are arbitrarily fixed or, equivalently, our reasoning are made conditionally on the second sample. In this section, we follow section 3 of Derumigny and Fermanian (2020) which gives similar results for the conditional Kendall's tau, a particular conditional U-statistic of order 2. Proofs are identical and therefore omitted. Nevertheless, asymptotic properties of $\hat{\beta}$ require corresponding results on the first-step estimators $\hat{\theta}$. These results are state in Stute (1991) and recalled for convenience in Section C. For $n, n' > 0$, denote by $\hat{\beta}_{n,n'}$ the estimator (4) with $h = h_n$ and $\lambda = \lambda_{n,n'}$.

Lemma 2. *We have* $\hat{\beta}_{n,n'} = \arg\min_{\beta \in \mathbb{R}^{p'}} \mathbb{G}_{n,n'}(\beta)$, *where*

$$\begin{aligned}
\mathbb{G}_{n,n'}(\beta) &:= \frac{2(n'-k)!}{n'!} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \xi_{\sigma,n} \boldsymbol{\psi}\left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}\right)^T (\beta^* - \beta) \\
&\quad + \frac{(n'-k)!}{n'!} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \left\{ \boldsymbol{\psi}\left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}\right)^T (\beta^* - \beta) \right\}^2 + \lambda_{n,n'} |\beta|_1.
\end{aligned} \tag{9}$$

Theorem 2 (Consistency of $\hat{\beta}$). *Under Assumption 10, if n' is fixed and $\lambda = \lambda_{n,n'} \to \lambda_0$, then, given $z'_1, \dots, z'_{n'}$ and as n tends to the infinity, $\hat{\beta}_{n,n'} \xrightarrow{\mathbb{P}} \beta^{**} := \inf_\beta \mathbb{G}_{\infty,n'}(\beta)$, where*

$$\mathbb{G}_{\infty,n'}(\beta) := \frac{1}{n'} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \left(\boldsymbol{\psi}\left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}\right)^T (\beta^* - \beta) \right)^2 + \lambda_0 |\beta|_1.$$

In particular, if $\lambda_0 = 0$ and $< \{\boldsymbol{\psi}\left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}\right) : \sigma \in \mathfrak{I}_{k,n'}\} >= \mathbb{R}^r$, then $\hat{\beta}_{n,n'} \xrightarrow{\mathbb{P}} \beta^*$.

Theorem 3 (Asymptotic law of the estimator). *Under Assumption 11, and if $\lambda_{n,n'}(nh^p_{n,n'})^{1/2}$ tends to ℓ when $n \to \infty$, we have $(nh^p_{n,n'})^{1/2}(\hat{\beta}_{n,n'} - \beta^*) \xrightarrow{D} \boldsymbol{u}^* :=$ $\arg\min_{\boldsymbol{u} \in \mathbb{R}^r} \mathbb{F}_{\infty,n'}(\boldsymbol{u})$, given $\boldsymbol{z}'_1, \dots, \boldsymbol{z}'_{n'}$, where*

$$\begin{aligned}\mathbb{F}_{\infty,n'}(\boldsymbol{u}) \\ := \frac{2(n'-k)!}{n'!} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{j=1}^{r} W_\sigma \psi_j \left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}\right) \boldsymbol{u}_j + \frac{(n'-k)!}{n'!} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \left(\boldsymbol{\psi}\left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}\right)^T \boldsymbol{u} \right)^2 \\ + \ell \sum_{i=1}^{r} \left(|\boldsymbol{u}_i| \mathbb{1}_{\{\beta^*_i = 0\}} + \boldsymbol{u}_i \mathrm{sign}(\beta^*_i) \mathbb{1}_{\{\beta^*_i \neq 0\}} \right),\end{aligned}$$

with $\boldsymbol{W} = (W_\sigma)_{\sigma \in \mathfrak{I}_{k,n'}} \sim \mathcal{N}(0, \tilde{\mathbb{H}})$ where

$$\begin{aligned}[\tilde{\mathbb{H}}]_{\sigma,\varsigma} := \sum_{j,l=1}^{k} \mathbb{1}_{\left\{\boldsymbol{z}'_{\sigma(j)} = \boldsymbol{z}'_{\varsigma(l)}\right\}} \frac{||K||_2^2}{f_Z\left(\boldsymbol{z}'_{\sigma(j)}\right)} \Lambda'\left(\theta\left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}\right)\right) \Lambda'\left(\theta\left(\boldsymbol{z}'_{\varsigma(1)}, \dots, \boldsymbol{z}'_{\varsigma(k)}\right)\right) \\ \cdot \left(\tilde{\theta}_{j,l}\left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}, \boldsymbol{z}'_{\varsigma(1)}, \dots, \boldsymbol{z}'_{\varsigma(k)}\right) - \theta\left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}\right) \theta\left(\boldsymbol{z}'_{\varsigma(1)}, \dots, \boldsymbol{z}'_{\varsigma(k)}\right) \right),\end{aligned}$$

and $\tilde{\theta}_{j,l}$ is as defined in Equation (C1).

Moreover, $\limsup_{n \to \infty} \mathbb{P}(\mathcal{S}_n = \mathcal{S}) = c < 1$, *where $\mathcal{S}_n := \{j : \hat{\beta}_j \neq 0\}$ and $\mathcal{S} := \{j : \beta_j \neq 0\}$.*

This result is mainly of theoretical interest. Nevertheless, one could estimate the matrix $\tilde{\mathbb{H}}$ and construct confidence intervals or hypothesis test from the asymptotic distribution of $(nh^p_{n,n'})^{1/2}(\hat{\beta}_{n,n'} - \beta^*)$.

A usual way of obtaining the oracle property is to modify our estimator in an "adaptive" way. Following Zou (2006), consider a preliminary "rough" estimator of β^*, denoted by $\tilde{\beta}_n$, or more simply $\tilde{\beta}$. Moreover $\nu_n(\tilde{\beta}_n - \beta^*)$ is assumed to be asymptotically normal, for some deterministic sequence (ν_n) that tends to the infinity. Now, let us consider the same optimization program as in (4) but with a random tuning parameter given by $\lambda_{n,n'} := \tilde{\lambda}_{n,n'}/|\tilde{\beta}_n|^\delta$, for some constant $\delta > 0$ and some positive deterministic sequence $(\tilde{\lambda}_{n,n'})$. The corresponding adaptive estimator (solution of the modified Equation 4) will be denoted by $\check{\beta}_{n,n'}$, or simply $\check{\beta}$. Hereafter, we still set $\mathcal{S}_n = \{j : \check{\beta}_j \neq 0\}$.

Theorem 4 (Asymptotic law of the adaptive estimator of β). *Under Assumption 11, if* $\tilde{\lambda}_{n,n'}(nh^p_{n,n'})^{1/2} \to \ell \geq 0$ *and* $\tilde{\lambda}_{n,n'}(nh^p_{n,n'})^{1/2} \nu_n^\delta \to \infty$ *when* $n \to \infty$, *we have* $(nh^p_{n,n'})^{1/2}(\check{\beta}_{n,n'} - \beta^*)_S \xrightarrow{D} \boldsymbol{u}^{**}_S := \arg\min_{\mathbf{u}_S \in \mathbb{R}^s} \check{\mathbb{F}}_{\infty,n'}(\boldsymbol{u}_S)$, *where*

$$\check{\mathbb{F}}_{\infty,n'}(\boldsymbol{u}_S) := \frac{2(n'-k)!}{n'!} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{j \in S} W_\sigma \psi_j(\boldsymbol{z}'_i)\boldsymbol{u}_j + \frac{(n'-k)!}{n'!} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \left(\sum_{j \in S} \psi_j(\boldsymbol{z}'_i)\boldsymbol{u}_j \right)^2 + \ell \sum_{i \in S} \frac{\boldsymbol{u}_i}{|\beta^*_i|^\delta} sign(\beta^*_i),$$

and $\boldsymbol{W} = (W_\sigma)_{\sigma \in \mathfrak{I}_{k,n'}} \sim \mathcal{N}\left(0, \tilde{\mathbb{H}}\right)$.

Moreover, when $\ell = 0$ *and the matrix* $\sum_{\sigma \in \mathfrak{I}_{k,n'}} \boldsymbol{\psi}\left(z'_{\sigma(1)}, \dots, z'_{\sigma(k)}\right) \boldsymbol{\psi}\left(z'_{\sigma(1)}, \dots, z'_{\sigma(k)}\right)^T$ *is invertible, the oracle property is fulfilled:* $\mathbb{P}(S_n = S) \to 1$ *as* $n \to \infty$.

3.3 | Asymptotic properties of $\hat{\beta}$ jointly in (n, n')

Now, we consider the framework in which both n and n' are going to infinity, while the dimensions p and r stay fixed. We now provide a consistency result for $\hat{\beta}_{n,n'}$. Note that the results of this section are also valid under weaker assumptions (as discussed in the Appendix, Sections D.1 and D.2).

Theorem 5 (Consistency of $\hat{\beta}_{n,n'}$, jointly in (n, n')). *Assume that Assumptions 2–9 are satisfied. Moreover, assume that* $\sum_{\sigma \in \mathfrak{I}_{k,n'}} \boldsymbol{\psi}\left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}\right) \boldsymbol{\psi}\left(\boldsymbol{z}'_{\sigma(1)}, \dots, \boldsymbol{z}'_{\sigma(k)}\right)^T / n'$ *converges to a matrix* $M_{\psi,\mathbf{z}'}$, *as* $n' \to \infty$. *Assume that* $\lambda_{n,n'} \to \lambda_0$ *and* $n' \exp(-Anh^{2kp}) \to 0$ *for every* $A > 0$, *when* $(n, n') \to \infty$. *Then* $\hat{\beta}_{n,n'} \xrightarrow{\mathbb{P}} \arg\min_{\beta \in \mathbb{R}^r} \mathbb{G}_{\infty,\infty}(\beta)$, *as* $(n, n') \to \infty$, *where* $\mathbb{G}_{\infty,\infty}(\beta) := (\beta^* - \beta) M_{\psi,\mathbf{z}'} (\beta^* - \beta)^T + \lambda_0 |\beta|_1$. *Moreover, if* $\lambda_0 = 0$ *and* $M_{\psi,\mathbf{z}'}$ *is invertible, then* $\hat{\beta}_{n,n'}$ *is consistent and tends to the true value* β^*.

Note that, since the sequence $(\mathbf{z}'_i)$ is deterministic, we only assume the convergence of the sequence of deterministic matrices $\sum_{\sigma \in \mathfrak{I}_{k,n'}} \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right) \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)^T / n'$ in $\mathbb{R}^{r^2}$. Moreover, if the "second subset" $(\mathbf{z}'_i)_{i=1,\dots,n'}$ were a random sample (drawn along the law $\mathbb{P}_\mathbf{Z}$), the latter convergence would be understood "in probability." And if $\mathbb{P}_\mathbf{Z}$ satisfies the identifiability condition (Proposition 1), then $M_{\psi,\mathbf{z}'}$ would be invertible and $\hat{\beta}_{n,n'} \to \beta^*$ in probability. Now, we want to go one step further and derive the asymptotic law of the estimator $\hat{\beta}_{n,n'}$.

Theorem 6 (Asymptotic law of $\hat{\beta}_{n,n'}$, jointly in (n, n')). *Under Assumptions 2–6 and under Assumption 16, we have*

$$(n \times n' \times h^p_{n,n'})^{1/2}(\hat{\beta}_{n,n'} - \beta^*) \xrightarrow{D} \mathcal{N}(0, \tilde{V}_{as}),$$

where $\tilde{V}_{as} := V_1^{-1} V_2 V_1^{-1}$, V_1 *is the matrix defined in Assumption 16(iv), and* V_2 *in Assumption 16(v).*

This theorem is proved in Section E where we state Assumption 16.

Remark 2. When the size n' is too small, estimation of β will not be good, even if the sample size n is large, because we only have too few "noisy observations" of θ. When the size n' is large enough, consistent estimation of β becomes possible. Furthermore, our estimator $\hat{\beta}$ converges at the (nonparametric) rate $(nh^p_{n,n'})^{1/2}$,

However, our model is parametric, and we should expect that faster rates are achievable. Indeed, when $n' \to +\infty$, our estimator $\hat{\beta}$ converges at the rate $(nn'h^p_{n,n'})^{1/2}$, which is strictly better. This can be interpreted as the following: the first-step estimation does not "filter out" information but gives enough so that we can even beat the naive first-step estimator. This is possible because we are exploiting the parametric nature of the model. Therefore, we should be (and are) able to improve on the non-parametric estimator $\hat{\theta}$.

4 | APPLICATIONS

4.1 | Examples

Following Example 4.4 in Stute (1991), we consider the function $g(x_1, x_2) := \mathbb{1}\{x_1 \leq x_2\}$, with $k = 2$. In this case $\theta(\mathbf{z}_1, \mathbf{z}_2) = \mathbb{P}(X_1 \leq X_2 | \mathbf{Z}_1 = \mathbf{z}_1, \mathbf{Z}_2 = \mathbf{z}_2)$. The parameter $\theta(\mathbf{z}_1, \mathbf{z}_2)$ quantifies the probability that the quantity of interest X be smaller if we knew that $\mathbf{Z} = \mathbf{z}_1$ than if we knew that $\mathbf{Z} = \mathbf{z}_2$.

To illustrate our methods, we choose a simple example, with the Epanechnikov kernel, defined by $K(x) := (3/4)(1 - u^2)\mathbb{1}|u| \leq 1$. It is a kernel of order $\alpha = 2$, with $\int K^2 = 3/5$. Assumption 2 is then satisfied with $C_K := 3/4$. Fix $p = 1$, $\mathcal{Z} = [-1, 1]$, $\mathcal{X} = \mathbb{R}$, $f_Z(z) = \phi(z)\mathbb{1}\{|z| \leq 1\}/(1 - 2\Phi(-1))$, where Φ and ϕ are, respectively, the cdf and the density of the standard Gaussian distribution and $X|Z = z \sim \mathcal{N}(z, 1)$, for every $z \in \mathcal{Z}$.

Assumption 3 is then satisfied with $C_{K,\alpha} = 0.2$. Assumption 4 is easily satisfied with $f_{Z,\max} = 1/\left(\sqrt{2\pi}(1 - 2\Phi(-1))\right) \leq 0.59$. Therefore, we can apply Lemma 1. We compute the constants $C_1 := 2f^k_{\mathbf{Z},\max}||K||^{2k}_2 = 2 \times 0.59^2 \times (3/5)^2 \leq 0.26$ and $C_2 := (4/3)C^k_K = (4/3) \times (3/4)^2 = 3/4$. Therefore, for any $n \geq 0$, $h, t > 0$, $z_1, z_2 \in \mathcal{Z}$, we have

$$\mathbb{P}\left(\left|N_2(z_1, z_2) - f_Z(z_1)f_Z(z_2)\right| \leq 0.1h^\alpha + t\right) \geq 1 - 2\exp\left(-\frac{[n/2]t^2}{0.26h^2 + 0.75h^2t}\right),$$

Assumption 5 is satisfied with $f_{Z,\min} = \phi(1)/(1 - 2\Phi(-1)) > 0.35$, so that we can apply Corollary 1. Therefore, the estimator $\hat{\theta}(z_1, z_2)$ exists with probability greater than $1 - 2\exp\left(-\frac{(n-1)h^2(0.35-0.1h^2)^2}{0.52+1.5\times(0.35-0.1h^2)}\right)$. Note that this probability is greater than 0.99 as soon as $n \geq 3\left(0.52 + 1.5 \times (0.35 - 0.1h^2)\right) / \left(h^2(0.35 - 0.1h^2)^2\right)$. For example, with $h = 0.2$, it means that the estimator $\hat{\theta}(z_1, z_2)$ exists with a probability greater than 99% as soon as n is greater than 651.

We list below other possible examples of applications. Conditional moments constitute also a natural class of U-statistics. They include the conditional variance ($p_{\mathbf{X}} = 1$, $k = 2$, $g(X_1, X_2) = X_1^2 - X_1 \cdot X_2$) and the conditional covariance ($p_{\mathbf{X}} = 2$, $k = 2$, $g(\mathbf{X}_1, \mathbf{X}_2) := X_{1,1} \times X_{1,2} - X_{1,1} \times X_{2,2}$). The conditional variance gives information about the volatility of X given the variable $\mathbf{Z}$. Conditional covariances can be used to describe how the dependence moves as a function of the conditioning variables $\mathbf{Z}$. Higher-order conditional moments (skewness, kurtosis, and so on) can also be estimated by higher-order conditional U-statistics, and they described, respectively, how the asymmetry and the behavior of the tails of X change as function of Z.

Gini's mean difference, an indicator of dispersion, can also be used in this framework. Formally, it is defined as the U-statistic with $p_{\mathbf{X}} = 1$, $k = 2$ and $g(X_1, X_2) := |X_1 - X_2|$. Its conditional version describes how two variables are far away, on average, given their conditioning variables $\mathbf{Z}$. for example, X could be the income of an individual, $\mathbf{Z}$ could be the position of their home, and $\theta(\mathbf{z}_1, \mathbf{z}_2)$ represents the average inequality between the income of two persons, one at point $\mathbf{z}_1$ and the other at point $\mathbf{z}_2$.

Other conditional dependence measures can also be written as conditional U-statistics, see for example, example 1.1.7 of Korolyuk and Borovskich (1994). They show how a U-statistic of order $k = 5$ can be used to estimated the dependence parameter

$$\theta = \iint \left(F_{1,2}(x, y) - F_{1,2}(x, \infty)F_{1,2}(\infty, y)\right) dF_{1,2}(x, y).$$

In our framework, we could consider a conditional version, given by

$$\theta(\mathbf{z}_1, \mathbf{z}_2) = \iint \left(F_{1,2|\mathbf{Z}=\mathbf{z}}(x, y) - F_{1,2|\mathbf{Z}=\mathbf{z}}(x, \infty)F_{1,2|\mathbf{Z}=\mathbf{z}}(\infty, y)\right) dF_{1,2|\mathbf{Z}=\mathbf{z}}(x, y),$$

where $\mathbf{X}$ is of dimension $p_{\mathbf{X}} = 2$.

4.2 | Simulation study: nonparametric estimation of the conditional covariance and correlation

In this section, we consider the estimation of the conditional covariance, which is a U-statistics for the kernel $g(x_1, x_2) = (x_1 - \mu_1)(x_2 - \mu_2)$ depending on the two means μ_1, μ_2. In the conditional case, this means that we want to estimate

$$\text{CondCov}(X_1, X_2|\mathbf{Z} = \mathbf{z}) = \mathbb{E}[(X_1 - \mu_1(\mathbf{z})) \times (X_2 - \mu_2(\mathbf{z}))|\mathbf{Z} = \mathbf{z}], \tag{10}$$

where $\mu_i := \mathbb{E}[X_i|\mathbf{Z} = \mathbf{z}]$. Note that (10) is a regular conditional functional of order 1 if the conditional means are assumed to be known. If the means are unknown, then this estimator becomes a regular conditional functional of order 2, since

$$\begin{aligned}\text{CondCov}(X_1, X_2|\mathbf{Z} = \mathbf{z}) &= \mathbb{E}[X_1X_2|\mathbf{Z} = \mathbf{z}] - \mathbb{E}[X_1|\mathbf{Z} = \mathbf{z}]\mathbb{E}[X_2|\mathbf{Z} = \mathbf{z}] \\ &= \mathbb{E}[X_{1,1}X_{2,1} - X_{1,1}X_{2,2}|\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}],\end{aligned} \tag{11}$$

for two i.i.d. replications $(X_{1,1}, X_{1,2}, \mathbf{Z}_1)$, $(X_{2,1}, X_{2,2}, \mathbf{Z}_2)$ of $(X_1, X_2, \mathbf{Z})$. These two representations of the conditional covariance naturally correspond to two kernel-based estimators: one is a conditional U-statistic of order 1 that assumes the knowledge of the true conditional mean while the second is a conditional U-statistics of order 2.

The conditional correlation between X_1 and X_2 given $\mathbf{Z} = \mathbf{z}$ can then be defined as

$$\text{CondCorr}(X_1, X_2|\mathbf{Z} = \mathbf{z}) := \frac{\text{CondCov}(X_1, X_2|\mathbf{Z} = \mathbf{z})}{\sqrt{\text{CondCov}(X_1, X_1|\mathbf{Z} = \mathbf{z})\text{CondCov}(X_2, X_2|\mathbf{Z} = \mathbf{z})}}.$$

This also gives two possible estimators for the conditional correlation: one which in which the conditional covariances are replaced by the corresponding conditional U-statistics of order 1 and another one for which the conditional covariances are replaced by conditional U-statistics of order 2.

In the first experiment, we choose Z univariate and uniform on the interval $[0, 1]$. We choose (X_1, X_2) given $Z = z$ following a Gaussian distribution with means $\mu_1(z), \mu_2(z)$, *SD* $\text{sd}_1(z), \text{sd}_2(z)$ and correlation $\rho(z)$. We fix $\mu_2(z) = 0$ and $\text{sd}_2(z)$ while we will change the conditional parameters μ_1, sd_1 and ρ. The performance of these estimators are evaluated by their integrated mean squared error (IMSE), defined by:

$$\text{IMSE}(\hat{\theta}) = \int_z \mathbb{E}\left[\left(\hat{\theta}(z) - \theta^*(z)\right)^2\right] dz,$$

when the true (conditional) parameter is θ^* and the estimator is $\hat{\theta}$. We will denote by $\hat{\theta}^{(1)}$ the estimator of the conditional covariance or conditional correlation that depends on the knowledge of the conditional mean (using Equation 10). Similarly, we will denote by $\hat{\theta}^{(2)}$ the estimator of the conditional covariance or conditional correlation that does not depend on the knowledge of the conditional mean (using Equation 11). This notation ensures that $\hat{\theta}^{(k)}$ is based on U-statistics of order k, for $k = 1, 2$.

It would seem natural to conjecture that $\hat{\theta}^{(1)}$ would have better performance than $\hat{\theta}^{(2)}$: it uses some already available knowledge, and is a U-statistics of lower order. Surprisingly and counter-intuitively, this is far from being true. In Figure A1, the IMSE of the competing estimators are displayed. For the covariance, it seems that the estimator that uses a U-statistic of order 2 achieves a better performance than the one that uses a U-statistic of order 1. For the correlation, the situation appears to be different, and the performance of both estimators seems rather close.

To confirm these findings, Figure A2 displays the ratio between the IMSE of the two estimators. In each situation, the estimator of the conditional covariance based on (11) is more efficient than the one that uses the knowledge of the true conditional mean. Such a situation can be related to the findings of Genest and Segers (2010) who showed that for some copula models, estimation of the marginal distributions (seen as nuisance parameters) leads to an improved efficiency compared to using the true values. On the contrary, for estimating the conditional correlation, it seems that knowing the true conditional mean may improve the estimation. Finally, this effect does not seem to disappear when increasing the sample size.

4.3 | Simulation study: estimation of the covariance regression

In this section, we continue to study the same model, but choosing $\text{sd}_1(z) = \text{sd}_2(z) = 1$ and $\mu_2(z) = 0$. We choose $\rho(z) = 0.1 + 0.8 \times z$, so that $\text{CondCov}(X_1, X_2 | Z = z) = 0.1 + 0.8 \times z$. We then estimate the model

$$\text{CondCov}(X_1, X_2 | Z = z) = \beta_0 + \beta_1 \times z,$$

using Algorithm 1. Depending on whether we use the U-statistic (10) or (11), we obtain two different conditional U-statistic estimators $(\hat{\beta}_0^{(1)}, \hat{\beta}_1^{(1)})$ and $(\hat{\beta}_0^{(2)}, \hat{\beta}_1^{(2)})$. The mean squared error of both estimators are displayed as a function of the bandwidth h in Figure A3.

We find that, in all scenarios, the estimator $\hat{\beta}_0^{(2)}$ using the best bandwidth h has the a lower mean square-error than the estimator $\hat{\beta}_0^{(1)}$ (which uses the knowledge of the true conditional mean). Surprisingly, the situation is different for the estimation of β_1 and the optimal bandwidth for both estimators is the same, and furthermore leads to the same MSE. These findings happened for all considered sample sizes.

FUNDING INFORMATION

Labex Ecodec, French Agence Nationale de la Recherche, Grant: ANR-11-LABEX-0047.

ACKNOWLEDGMENTS

The author thanks Professor Jean-David Fermanian for helpful comments and discussions on an earlier draft. The author also thanks the Associate Editor and the reviewer, whose comments helped to significantly increase the readability of this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Alexis Derumigny https://orcid.org/0000-0002-6163-8097

REFERENCES

Bellec, P. C., Lecué, G., & Tsybakov, A. B. (2018). Slope meets lasso: Improved oracle bounds and optimality. *The Annals of Statistics*, *46*(6B), 3603–3642.

Bickel, P. J., Ritov, Y., & Tsybakov, A. B. (2009). Simultaneous analysis of lasso and Dantzig selector. *The Annals of Statistics*, *37*(4), 1705–1732.

Derumigny, A., & Fermanian, J. D. (2019). On kernel-based estimation of conditional Kendall's tau: Finite-distance bounds and asymptotic behavior. *Dependence Modeling*, *7*(1), 292–321.

Derumigny, A., & Fermanian, J. D. (2020). On Kendall's regression. *Journal of Multivariate Analysis*, *178*, 104610.

Dony, J., & Mason, D. M. (2008). Uniform in bandwidth consistency of conditional U-statistics. *Bernoulli*, *14*(4), 1108–1133.

Genest, C., & Segers, J. (2010). On the covariance of the asymptotic empirical copula process. *Journal of Multivariate Analysis*, *101*(8), 1837–1845.

Hoeffding, W. (1948). A class of statistics with asymptotically normal distribution. *The Annals of Mathematical Statistics*, *19*(3), 293–325.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, *58*(301), 13–30.

Korolyuk, V. S., & Borovskich, Y. V. (1994). *Theory of U-statistics*. Dordrecht: Springer.

Kowalski, J., & Tu, X. M. (2008). *Modern applied U-statistics* (Vol. *714*). Hoboken, New Jersey: John Wiley & Sons.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics*. New York: John Wiley & Sons.

Stute, W. (1991). Conditional U-statistics. *The Annals of Probability*, *19*(2), 812–825.

van de Geer, S. A., & Bühlmann, P. (2009). On the conditions used to prove oracle results for the lasso. *Electronic Journal of Statistics*, *3*, 1360–1392.

Zou, H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association*, *101*(476), 1418–1429.

How to cite this article: Derumigny, A. (2024). Estimation of a regular conditional functional by conditional U-statistic regression. *Statistica Neerlandica*, 1–41. https://doi.org/10.1111/stan.12350

APPENDIX A. FIGURES

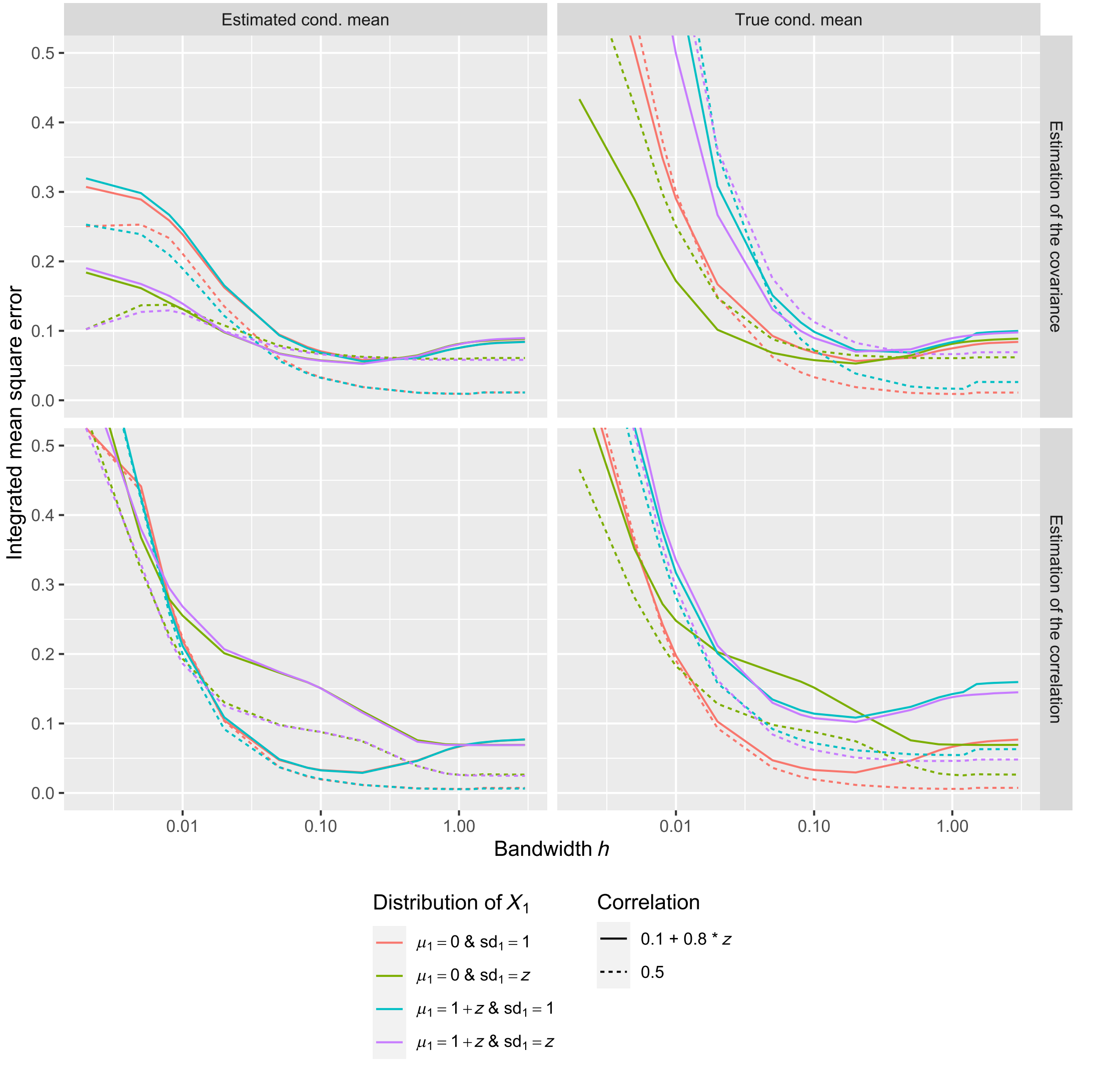

FIGURE A1 Integrated mean square error of the estimators of the conditional covariance and of the conditional correlation, without and with knowledge of the true conditional mean, for a sample size $n = 200$ and for different data-generating processes.

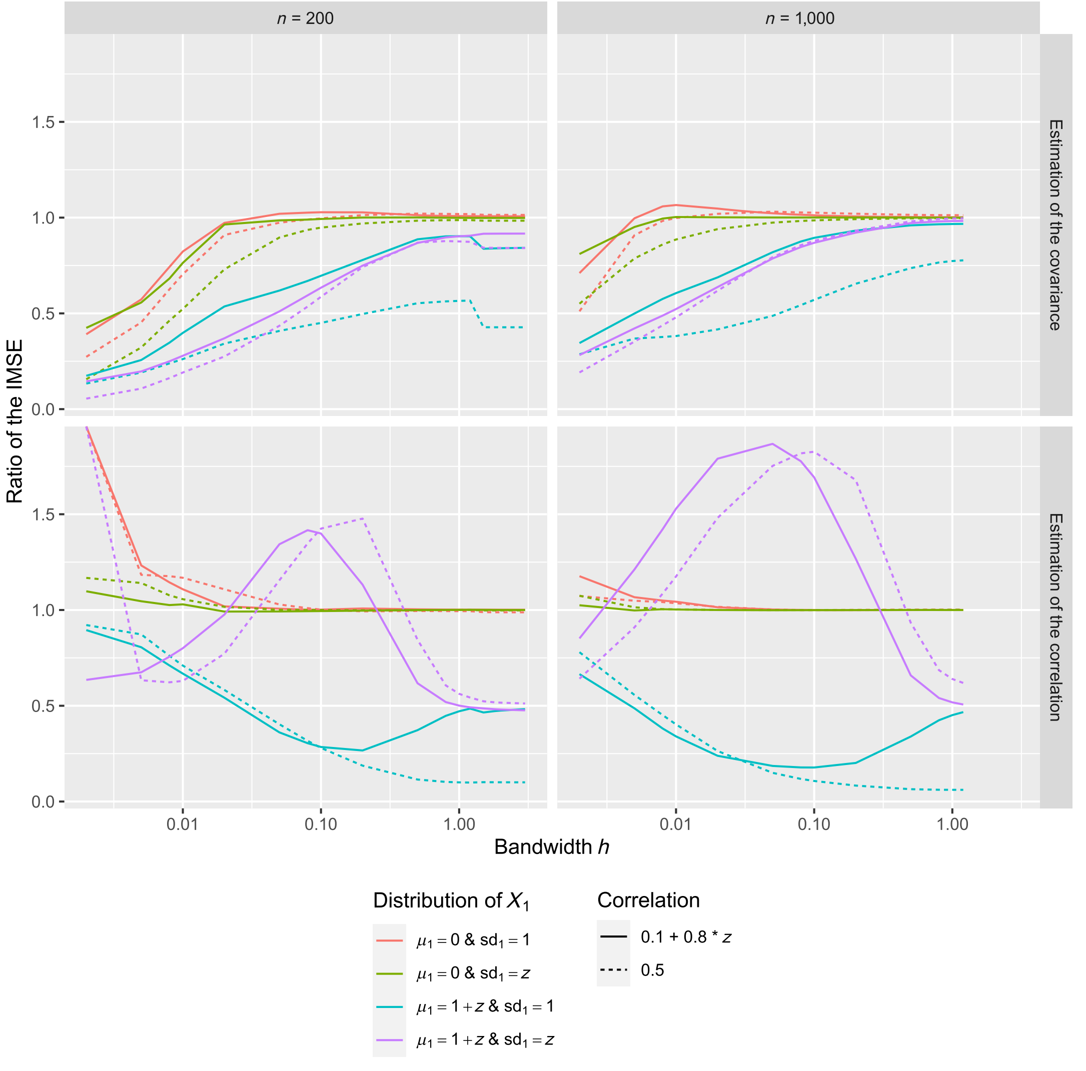

FIGURE A2 Ratio $\mathrm{IMSE}(\hat{\theta}^{(2)})/\mathrm{IMSE}(\hat{\theta}^{(1)})$ for different data-generating processes and for two sample sizes.

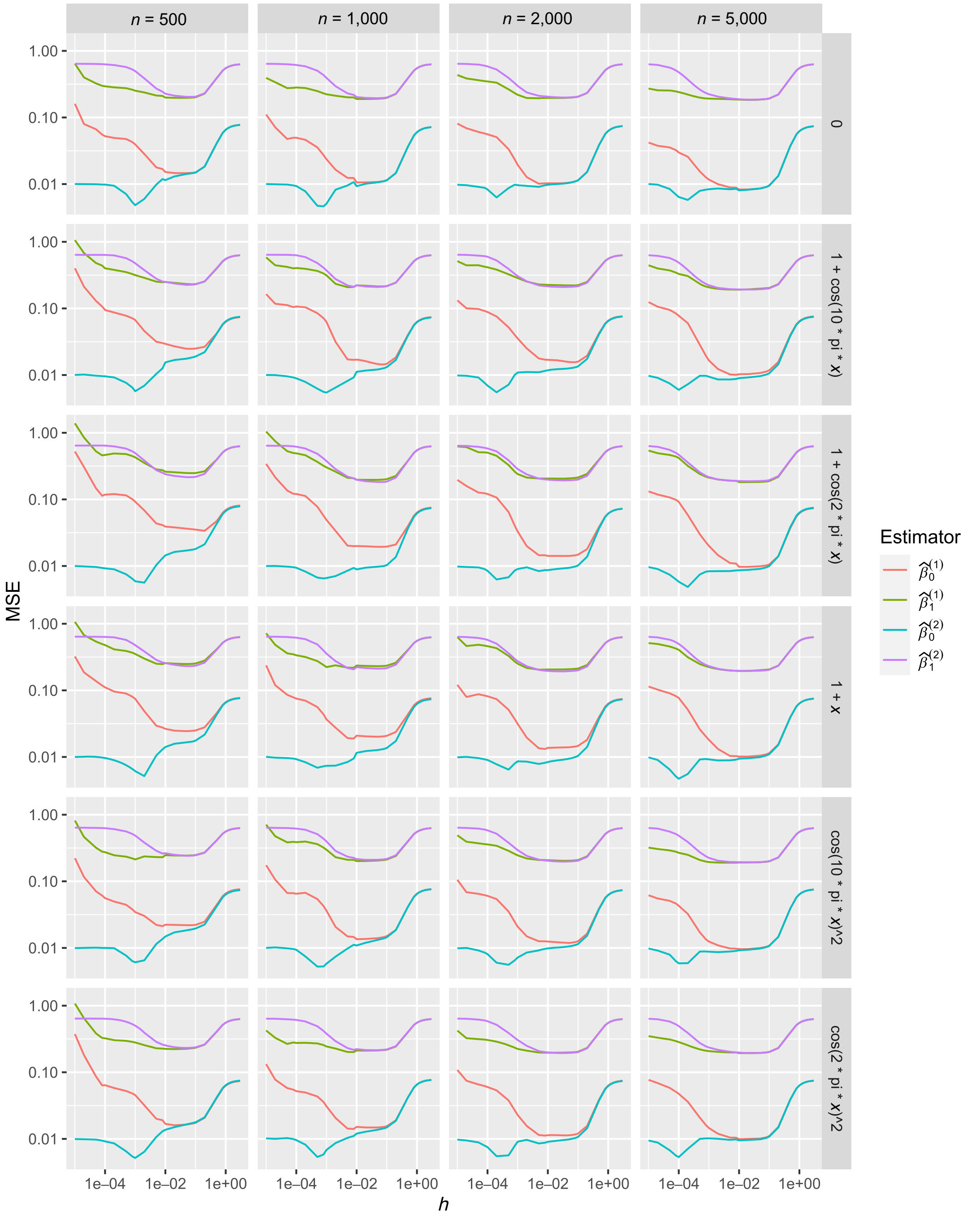

FIGURE A3 Mean squared error of $\widehat{\beta}_0^{(1)}$, $\widehat{\beta}_1^{(1)}$, $\widehat{\beta}_0^{(2)}$, $\widehat{\beta}_1^{(2)}$ as a function of h, for different sample sizes n and different specification of μ_1.

APPENDIX B. NOTATIONS

In the proofs, we will use the following shortcut notation. First, $\mathbf{x}_{1:k}$ denotes the k-tuple $(\mathbf{x}_1, \dots, \mathbf{x}_k) \in \mathcal{X}^k$. Similarly, for a function σ, $\sigma(1:k)$ denotes the tuple $(\sigma(1), \dots, \sigma(k))$, and $\mathbf{X}_{\sigma(1:k)}$ is the k-tuple $(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)})$. For any variable Y and any collection of given points $(\mathbf{z}_1, \dots \mathbf{z}_k)$, the conditional expectation $\mathbb{E}[Y|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}]$ denotes $\mathbb{E}[Y|\mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k]$. We denote by $\int \phi(\mathbf{z}_{1:k})d\mathbf{z}_{1:k}$ the integral $\int \phi(\mathbf{z}_1, \dots, \mathbf{z}_k)d\mathbf{z}_1 \cdots d\mathbf{z}_k$ for any integrable function $\phi : \mathbb{R}^{k\times p} \to \mathbb{R}$, and by $\int g(\mathbf{x}_{1:k})d\mu^{\otimes k}(\mathbf{x}_{1:k})$ the integral $\int g(\mathbf{z}_1, \dots, \mathbf{z}_k)d\mu(\mathbf{x}_1) \dots d\mu(\mathbf{x}_k)$ for any μ-integrable function $g : \mathcal{X}^k \to \mathbb{R}$.

APPENDIX C. ASYMPTOTIC RESULTS FOR $\hat{\theta}$

The estimator $\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)$ has been first studied by Stute, 1991. He proved the consistency and the asymptotic normality of $\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)$. We recall his results.

Assumption 10.

- ***i*** $h_n \to 0$ and $nh_n^p \to \infty$;
- ***ii*** $K(\mathbf{z}) \geq C_{K,1}\mathbb{1}_{\{|\mathbf{z}|_\infty \leq C_{K,2}\}}$ for some $C_{K,1}$, $C_{K,2} > 0$;
- ***iii*** there exists a decreasing function $H : \mathbb{R}_+ \to \mathbb{R}_+$, and positive constants c_1, c_2 such that $H(t)=o(t^{-1})$ and $c_1 H(|\mathbf{z}|_\infty) \leq K(\mathbf{z}) \leq c_2 H(|\mathbf{z}|_\infty)$.

Proposition 3 (Consistency of $\hat{\theta}$, theorem 2 in Stute, 1991). *Under Assumption 10, for $\mathbb{P}_{\mathbf{Z}}^{\otimes k}$-almost all $(\mathbf{z}_1, \dots, \mathbf{z}_k)$,*

$$\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k) \xrightarrow{\mathbb{P}} \theta(\mathbf{z}_1, \dots, \mathbf{z}_k) \text{ as } n \to \infty.$$

We introduce now a few more notations to state the asymptotic normality of $\hat{\theta}$. For $1 \leq j, l, m \leq k$ and $\mathbf{z}_1, \dots, \mathbf{z}_{3k} \in \mathcal{Z}^{3k}$, define

$$\begin{aligned}
\theta_{j,l}(\mathbf{z}_1, \dots, \mathbf{z}_k) &:= \mathbb{E}\big[g(\mathbf{X}_1, \dots, \mathbf{X}_{j-1}, \mathbf{X}, \mathbf{X}_{j+1}, \dots, \mathbf{X}_k)g(\mathbf{X}_{k+1}, \dots, \mathbf{X}_{k+l-1}, \mathbf{X}, \mathbf{X}_{k+l+1}, \dots, \mathbf{X}_{2k}) \\
&\quad |\mathbf{Z} = \mathbf{z}_j; \mathbf{Z}_i = \mathbf{z}_i, \forall i = 1, \dots, k, i \neq j; \mathbf{Z}_{k+i} = \mathbf{z}_i, \forall i = 1, \dots, k, i \neq l\big], \\
\tilde{\theta}_{j,l}(\mathbf{z}_1, \dots, \mathbf{z}_{2k}) &:= \mathbb{E}\big[g(\mathbf{X}_1, \dots, \mathbf{X}_{j-1}, \mathbf{X}, \mathbf{X}_{j+1}, \dots, \mathbf{X}_k)g(\mathbf{X}_{k+1}, \dots, \mathbf{X}_{k+l-1}, \mathbf{X}, \mathbf{X}_{k+l+1}, \dots, \mathbf{X}_{2k}) \\
&\quad |\mathbf{Z} = \mathbf{z}_j; \mathbf{Z}_i = \mathbf{z}_i, \forall i = 1, \dots, 2k, i \notin \{j, k+l\}\big]. \\
\theta_{j,l,m}(\mathbf{z}_1, \dots, \mathbf{z}_{3k}) &:= \mathbb{E}\big[g(\mathbf{X}_1, \dots, \mathbf{X}_{j-1}, \mathbf{X}, \mathbf{X}_{j+1}, \dots, \mathbf{X}_k) \\
&\quad g(\mathbf{X}_{k+1}, \dots, \mathbf{X}_{k+l-1}, \mathbf{X}, \mathbf{X}_{k+l+1}, \dots, \mathbf{X}_{2k})g(\mathbf{X}_{2k+1}, \dots, \mathbf{X}_{2k+m-1}, \mathbf{X}, \mathbf{X}_{2k+m+1}, \dots, \mathbf{X}_{3k}) \\
&\quad |\mathbf{Z} = \mathbf{z}_j; \mathbf{Z}_i = \mathbf{z}_i, \forall i = 1, \dots, 3k, i \notin \{j, k+l, 2k+m\}\big].
\end{aligned} \tag{C1}$$

Assumption 11.

- ***i*** $h_n \to 0$ and $nh_n^p \to \infty$;
- ***ii*** K is symmetric at 0, bounded and compactly supported;
- ***iii*** $\theta_{j,l}$ is continuous at $(\mathbf{z}_1, \dots, \mathbf{z}_k)$ for all $1 \leq j, l \leq k$;
- 1. θ is two times continuously differentiable in a neighborhood of $(\mathbf{z}_1, \dots, \mathbf{z}_k)$;
- ***iv*** $\theta_{j,l,m}$ is bounded in a neighborhood of $(\mathbf{z}_1, \dots, \mathbf{z}_k, \mathbf{z}_1, \dots, \mathbf{z}_k, \mathbf{z}_1, \dots, \mathbf{z}_k) \in \mathcal{Z}^{3k}$, for all $1 \leq j, l, m \leq k$;
- ***v*** $f_{\mathbf{Z}}$ is twice differentiable in neighborhoods of $\mathbf{z}_i$, $1 \leq i \leq k$.

Proposition 4 (Asymptotic normality of $\hat{\theta}$, corollary 2.4 in Stute (1991)). *Under Assumption 11, we have*

$$\sqrt{nh_n^p}\left(\hat{\theta}(\boldsymbol{z}_1, \dots, \boldsymbol{z}_k) - \theta(\boldsymbol{z}_1, \dots, \boldsymbol{z}_k)\right) \xrightarrow{D} \mathcal{N}(0, \rho^2),$$

where $\rho^2 := \sum_{j,l=1}^{k} \mathbb{1}_{\{z_j = z_l\}}\left(\theta_{j,l}(\boldsymbol{z}_1, \dots, \boldsymbol{z}_k) - \theta^2(\boldsymbol{z}_1, \dots, \boldsymbol{z}_k)\right) ||K||_2^2 / f_Z(z_j)$.

Moreover, let N be a positive integer, and $\left(\boldsymbol{z}_1^{(1)}, \dots, \boldsymbol{z}_k^{(1)}, \dots, \boldsymbol{z}_1^{(N)}, \dots, \boldsymbol{z}_k^{(N)}\right) \in \mathcal{Z}^{k\times N}$. *Then under similar regularity conditions,* $\sqrt{nh_n^p}\big(\hat{\theta}(\boldsymbol{z}_1^{(i)}, \dots, \boldsymbol{z}_k^{(i)}) - \theta(\boldsymbol{z}_1^{(i)}, \dots, \boldsymbol{z}_k^{(i)})\big)_{i=1,\dots,N} \xrightarrow{D} \mathcal{N}(0, \mathbb{H})$, *where, for* $1 \le \tilde{j}, \tilde{l} \le N$,

$$[\mathbb{H}]_{\tilde{j},\tilde{l}} := \sum_{j,l=1}^{k} \mathbb{1}_{\left\{\boldsymbol{z}_j^{(\tilde{j})} = \boldsymbol{z}_l^{(\tilde{l})}\right\}} \left(\tilde{\theta}_{j,l}\left(\boldsymbol{z}_1^{(\tilde{j})}, \dots, \boldsymbol{z}_k^{(\tilde{j})}, \boldsymbol{z}_1^{(\tilde{l})}, \dots, \boldsymbol{z}_k^{(\tilde{l})}\right) - \theta\left(\boldsymbol{z}_1^{(\tilde{j})}, \dots, \boldsymbol{z}_k^{(\tilde{j})}\right)\theta\left(\boldsymbol{z}_1^{(\tilde{l})}, \dots, \boldsymbol{z}_k^{(\tilde{l})}\right)\right) \frac{||K||_2^2}{f_Z\left(\boldsymbol{z}_j^{(\tilde{j})}\right)}.$$

Note that the second part of Proposition 4 above is a consequence of the first one. Indeed, for every $(c_1, \dots, c_N) \in \mathbb{R}^N$, we can define $\theta\left(\mathbf{z}_1^{(1)}, \dots, \mathbf{z}_k^{(1)}, \dots, \mathbf{z}_1^{(N)}, \dots, \mathbf{z}_k^{(N)}\right) := \sum_{\tilde{\zeta}=1}^{N} c_{\tilde{\zeta}} \theta(\mathbf{z}_1^{(\tilde{\zeta})}, \dots, \mathbf{z}_k^{(\tilde{\zeta})})$ and corresponding versions of g, $\hat{\theta}$ and ρ^2. Finally, the conclusion follows from the Cramér–Wold device.

APPENDIX D. FINITE DISTANCE PROOFS FOR $\hat{\theta}$ AND $\hat{\beta}$

For convenience, we recall Berk's (1970) inequality (see theorem A in (Serfling, 1980, p. 201)). Note that, if $m = 1$, this reduces to Bernstein's inequality.

Lemma 3. *Let* $k > 0$, $n \ge k$, $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ *i.i.d. random vectors with values in a measurable space* $\mathcal{X}$ *and* $g : \mathcal{X}^k \to [a, b]$ *be a real bounded function. Set* $\theta := \mathbb{E}[g(\boldsymbol{X}_{1:k})]$ *and* $\sigma^2 := Var[g(\boldsymbol{X}_{1:k})]$. *Then, for any* $t > 0$,

$$\mathbb{P}\left(\binom{n}{k}^{-1} \sum_{\sigma \in \mathfrak{I}_{k,n}^{\uparrow}} g\left(\boldsymbol{X}_{\sigma(1:k)}\right) - \theta \ge t\right) \le \exp\left(-\frac{[n/k]t^2}{2\sigma^2 + (2/3)(b-\theta)t}\right),$$

where $\mathfrak{I}_{k,n}$ *is the set of injective functions from* $\{1, \dots, k\}$ *to* $\{1, \dots, n\}$ *and* $\mathfrak{I}_{k,n}^{\uparrow}$ *is the subset of* $\mathfrak{I}_{k,n}$ *made of increasing functions.*

Note that g does not need to be symmetric for this bound to hold. Indeed, if g is not symmetric, we can nonetheless apply this lemma to the symmetrized version $\tilde{g}$ defined as $\tilde{g}(\mathbf{x}_{1:k}) := (k!)^{-1} \sum_{\sigma \in \mathfrak{I}_{k,k}} g(\mathbf{x}_{\sigma(1:k)})$, and we get the result.

D.1 Proof of Lemma 1

We will actually show the result under the following weaker assumption.

Assumption 12. $f_{\mathbf{Z}}$ is α-times continuously differentiable on $\mathcal{Z}$ and there exists a finite constant $C_{K,\alpha}$ such that, for all $\mathbf{z}_1, \dots \mathbf{z}_k$,

$$\int \left|K(\mathbf{u}_1)\cdots K(\mathbf{u}_k)\right| \sum_{m_1+\cdots+m_k=\alpha} \binom{\alpha}{m_{1:k}} \cdot \prod_{i=1}^{k} \sum_{j_1,\dots j_{m_i}=1}^{p} \left|u_{i,j_1} \dots u_{i,j_{m_i}}\right| \sup_{t\in[0,1]} \left| \frac{\partial^{m_i} f_{\mathbf{Z}}}{\partial z_{j_1} \cdots \partial z_{j_{m_i}}} (\mathbf{z}_i + t\mathbf{u}_i)\right| d\mathbf{u}_1 \dots d\mathbf{u}_k \leq C_{K,\alpha}$$

where $\binom{\alpha}{m_{1:k}} := \alpha!/\left(\prod_{i=1}^k (m_i!)\right)$ is the multinomial coefficient.

Indeed, Assumption 3 implies that

$$\begin{aligned}
&\int \left|K(\mathbf{u}_1) \dots K(\mathbf{u}_k)\right| \sum_{m_1+\cdots+m_k=\alpha} \binom{\alpha}{m_{1:k}} \prod_{i=1}^{k} \\
&\quad \sum_{j_1,\dots j_{m_i}=1}^{p} \left|u_{i,j_1} \dots u_{i,j_{m_i}}\right| \sup_{t\in[0,1]} \left| \frac{\partial^{m_i} f_{\mathbf{Z}}}{\partial z_{j_1} \dots \partial z_{j_{m_i}}} (\mathbf{z}_i + t\mathbf{u}_i)\right| d\mathbf{u}_1 \dots d\mathbf{u}_k \\
&\leq C_{f,\alpha}^k \int \left|K(\mathbf{u}_1) \dots K(\mathbf{u}_k)\right| \sum_{m_1+\cdots+m_k=\alpha} \binom{\alpha}{m_{1:k}} \prod_{i=1}^{k} \sum_{j_1,\dots j_{m_i}=1}^{p} \left|u_{i,j_1} \dots u_{i,j_{m_i}}\right| d\mathbf{u}_1 \dots d\mathbf{u}_k \\
&\leq C_{f,\alpha}^k \sum_{m_1+\cdots+m_k=\alpha} \binom{\alpha}{m_{1:k}} \prod_{i=1}^{k} \sum_{j_1,\dots j_{m_i}=1}^{p} \int \left|K(\mathbf{u}_i)\right| \left|u_{i,j_1} \dots u_{i,j_{m_i}}\right| d\mathbf{u}_1 \dots d\mathbf{u}_k \\
&\leq C_{f,\alpha}^k \sum_{m_1+\cdots+m_k=\alpha} \binom{\alpha}{m_{1:k}} \prod_{i=1}^{k} \sum_{j_1,\dots j_{m_i}=1}^{p} C_{K,\text{int}}^k \\
&\leq C_{f,\alpha}^k \sum_{m_1+\cdots+m_k=\alpha} \binom{\alpha}{m_{1:k}} \prod_{i=1}^{k} p^{m_i} C_{K,\text{int}} \\
&\leq C_{f,\alpha}^k p^\alpha C_{K,\text{int}}^k \sum_{m_1+\cdots+m_k=\alpha} \binom{\alpha}{m_{1:k}} \\
&\leq C_{f,\alpha}^k p^\alpha C_{K,\text{int}}^k k^\alpha,
\end{aligned}$$

where in the last line, we use the multinomial theorem. The proof is complete by proving the following result.

Lemma 4. *Under Assumptions 2, 4, and 12, we have for any* $t > 0$,

$$\mathbb{P}\left(\left|N_k(\boldsymbol{z}_1, \dots, \boldsymbol{z}_k) - \prod_{i=1}^{k} f_Z(\boldsymbol{z}_i)\right| \leq \frac{C_{K,\alpha}}{\alpha!} h^\alpha + t\right) \geq 1 - 2\exp\left(-\frac{[n/k]t^2}{h^{-kp}C_1 + h^{-kp}C_2 t}\right),$$

where $C_1 := 2f_{Z,\max}^k ||K||_2^{2k}$, *and* $C_2 := (4/3)C_K^k$ *and* $||K||_2^2 := \int K^2$.

We decompose the quantity to bound into a stochastic part and a bias as follows:

$$N_k(\mathbf{z}_{1:k}) - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i) = (N_k(\mathbf{z}_{1:k}) - \mathbb{E}[N_k(\mathbf{z}_{1:k})]) + \left(\mathbb{E}[N_k(\mathbf{z}_{1:k})] - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)\right).$$

We first bound the bias.

$$\begin{aligned}\left|\mathbb{E}\left[N_k(\mathbf{z}_{1:k})\right] - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)\right| &= \left|\mathbb{E}\left[\binom{n}{k}^{-1} \sum_{\sigma\in\mathfrak{I}_{k,n}} \prod_{i=1}^k K_h\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right)\right] - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)\right| \\ &= \left|\int \left(\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i + h\mathbf{u}_i) - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)\right) \prod_{i=1}^k K(\mathbf{u}_i) d\mathbf{u}_i\right| \\ &= \left|\int \left(\phi_{\mathbf{z},\mathbf{u}}(1) - \phi_{\mathbf{z},\mathbf{u}}(0)\right) \prod_{i=1}^k K(\mathbf{u}_i) d\mathbf{u}_i\right|,\end{aligned}$$

where $\phi_{\mathbf{z},\mathbf{u}}(t) := \prod_{j=1}^k f_{\mathbf{Z}}\left(\mathbf{z}_i + th\mathbf{u}_j\right)$ for $t \in [-1, 1]$. Note that this function has at least the same regularity as $f_{\mathbf{Z}}$, so it is α-differentiable, and by a Taylor–Lagrange expansion, we get

$$\left|\mathbb{E}[N_k(\mathbf{z}_{1:k})] - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)\right| = \left|\int_{\mathbb{R}^{kp}} \left(\sum_{i=1}^{\alpha-1} \frac{1}{i!}\phi^{(i)}_{\mathbf{z},\mathbf{u}}(0) + \frac{1}{\alpha!}\phi^{(\alpha)}_{\mathbf{z},\mathbf{u}}(t_{\mathbf{z},\mathbf{u}})\right) \prod_{i=1}^k K(\mathbf{u}_i) d\mathbf{u}_i\right|.$$

For $l > 0$, we have

$$\begin{aligned}\phi^{(l)}_{\mathbf{z},\mathbf{u}}(0) &= \sum_{m_1+\cdots+m_k=l} \binom{\alpha}{m_{1:k}} \prod_{i=1}^k \frac{\partial^{m_i}\left(f_{\mathbf{Z}}\left(\mathbf{z}_i + ht\mathbf{u}_i\right)\right)}{\partial t^{m_i}}(0) \\ &= \sum_{m_1+\cdots+m_k=l} \binom{\alpha}{m_{1:k}} \prod_{i=1}^k \sum_{j_1,\dots,j_{m_i}=1}^{p} h^{m_i} u_{i,j_1} \dots u_{i,j_{m_i}} \frac{\partial^{m_i} f_{\mathbf{Z}}}{\partial z_{j_1} \cdots \partial z_{j_{m_i}}}\left(\mathbf{z}_i + t_{\mathbf{z},\mathbf{u}} h\mathbf{u}_i\right),\end{aligned}$$

where $\binom{\alpha}{m_{1:k}} := \alpha!/\left(\prod_{i=1}^k (m_i!)\right)$ is the multinomial coefficient. Using Assumption 2, for every $i = 1, \dots, \alpha - 1$, we get $\int K(\mathbf{u}_1) \dots K(\mathbf{u}_k)\phi^{(i)}_{\mathbf{z},\mathbf{u}}(0) d\mathbf{u}_1 \dots d\mathbf{u}_k = 0$. Therefore, only the last term remains and we have

$$\left|\mathbb{E}[N_k(\mathbf{z}_{1:k})] - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)\right| = \left|\int \left(\frac{1}{\alpha!}\phi^{(\alpha)}_{\mathbf{z},\mathbf{u}}(t_{\mathbf{z},\mathbf{u}})\right) \prod_{i=1}^k K(\mathbf{u}_i) d\mathbf{u}_i\right| \leq \frac{C_{K,\alpha}}{\alpha!} h^{\alpha},$$

using Assumption 12.

Second, we bound the stochastic part. We have

$$N_k(\mathbf{z}_{1:k}) - \mathbb{E}[N_k(\mathbf{z}_{1:k})] = \frac{k!(n-k)!}{n!} \sum_{\sigma\in\mathfrak{I}^{\uparrow}_{k,n}} \prod_{i=1}^k K_h\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right) - \prod_{i=1}^k \mathbb{E}[K_h\left(\mathbf{Z}_i - \mathbf{z}_i\right)].$$

Then, we can apply Lemma 3 to the function g defined by $g(\tilde{\mathbf{z}}_1, \dots, \tilde{\mathbf{z}}_k) := \prod_{i=1}^k K_h(\tilde{\mathbf{z}}_i - \mathbf{z}_i)$. Here, we have $b = -a = h^{-kp}C_K^k$, and

$$\mathrm{Var}\left[g(\mathbf{Z}_1, \dots, \mathbf{Z}_k)^2\right] \leq \mathbb{E}\left[g(\mathbf{Z}_1, \dots, \mathbf{Z}_k)^2\right] = \prod_{i=1}^k \mathbb{E}\left[K_h(\mathbf{Z}_i - \mathbf{z}_i)^2\right] \leq h^{-kp} f_{\mathbf{Z},\max}^k ||K||_2^{2k}.$$

Finally, we get

$$\mathbb{P}\left(\binom{n}{k}^{-1} N_k(\mathbf{z}_{1:k}) - \mathbb{E}[N_k(\mathbf{z}_{1:k})] \geq t\right) \leq \exp\left(-\frac{[n/k]t^2}{2h^{-kp}f_{\mathbf{Z},\max}^k ||K||_2^{2k} + (4/3)h^{-kp}C_K^k t}\right),$$

■

D.2 Proof of Proposition 2

As in the previous section, we will prove a more general results, under more sophisticated but weaker assumptions.

Assumption 13. There exists a measure μ on $(\mathcal{X}, \mathcal{A})$ such that $\mathbb{P}_{\mathbf{X},\mathbf{Z}}$ is absolutely continuous with respect to $\mu \otimes \mathrm{Leb}_p$, where Leb_p is the Lebesgue measure on $\mathbb{R}^p$.

Obviously, Assumption 13 is satisfied as soon as Assumption 6 is satisfied.

Assumption 14. For every $\mathbf{x} \in \mathcal{X}$, $\mathbf{z} \mapsto f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})$ is differentiable almost everywhere up to the order α. Moreover, there exists a finite constant $C_{g,f,\alpha} > 0$, such that, for every positive integers $m_1, \dots, m_k$ such that $\sum_{i=1}^k m_i = \alpha$, for every $0 \leq j_1, \dots, j_{m_i} \leq p$,

$$\int \prod_{i=1}^k \left| \left(g(\mathbf{x}_1, \dots, \mathbf{x}_k) - \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_i = \mathbf{z}_i, \forall i = 1, \dots, k\right]\right) \right.$$
$$\left. \cdot \left(\frac{\partial^{m_i} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \cdots \partial z_{j_{m_i}}}(\mathbf{x}_i, \mathbf{z}_i + \mathbf{u}_i) - \frac{\partial^{m_i} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \cdots \partial z_{j_{m_i}}}(\mathbf{x}_i, \mathbf{z}_i) \right) \right| d\mu(\mathbf{x}_1) \dots d\mu(\mathbf{x}_k) \leq C_{g,f,\alpha} \prod_{i=1}^k \left|\mathbf{u}_i\right|_\infty,$$

for every choices of $\mathbf{x}_1, \dots, \mathbf{x}_k \in \mathcal{X}$ and $\mathbf{z}_1, \dots, \mathbf{z}_k \in \mathcal{Z}, \mathbf{u}_1, \dots, \mathbf{u}_k \in \mathbb{R}^p$ such that $\mathbf{z}_i + \mathbf{u}_i \in \mathcal{Z}$. There exists a constant $C'_{K,\alpha}$ such that $\sum_{m_1+\cdots+m_k=\alpha} \binom{n}{m_{1:k}} \int \prod_{i=1}^k K(\mathbf{u}_i) \sum_{j_1,\dots,j_{m_i}=1}^p u_{i,j_1} \dots u_{i,j_{m_i}} \prod_{i=1}^k \left|\mathbf{u}_i\right|_\infty d\mathbf{u}_1 \cdots d\mathbf{u}_k \leq C'_{K,\alpha}$.

Lemma 5. *Assume that Assumptions 6–8 are satisfied. Then Assumption 14 is satisfied too.*

Proof of Lemma 5. Using successively the fact that g is bounded, Taylor's inequality, we obtain

$$\int \prod_{i=1}^k \left| \left(g(\mathbf{x}_1, \dots, \mathbf{x}_k) - \mathbb{E}\left[g(\mathbf{X}_1, \dots, \mathbf{X}_k) \middle| \mathbf{Z}_i = \mathbf{z}_i, \forall i = 1, \dots, k\right]\right) \right.$$
$$\left. \cdot \left(\frac{\partial^{m_i} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \dots \partial z_{j_{m_i}}}(\mathbf{x}_i, \mathbf{z}_i + \mathbf{u}_i) - \frac{\partial^{m_i} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \dots \partial z_{j_{m_i}}}(\mathbf{x}_i, \mathbf{z}_i) \right) \right| d\mu(\mathbf{x}_1) \dots d\mu(\mathbf{x}_k)$$
$$\leq C_g \int \prod_{i=1}^k \left| \left(\frac{\partial^{m_i} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \dots \partial z_{j_{m_i}}}(\mathbf{x}_i, \mathbf{z}_i + \mathbf{u}_i) - \frac{\partial^{m_i} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \dots \partial z_{j_{m_i}}}(\mathbf{x}_i, \mathbf{z}_i) \right) \right| d\mu(\mathbf{x}_1) \dots d\mu(\mathbf{x}_k)$$
$$\leq C_g^k \widetilde{C}_{f,\alpha}^k \int \prod_{i=1}^k |\mathbf{u}_i|_\infty d\mu(\mathbf{x}_1) \dots d\mu(\mathbf{x}_k) = C_g^k \widetilde{C}_{f,\alpha}^k |\mathbf{u}_i|_\infty,$$

since μ is a probability measure. The second part of Assumption 14 is a consequence of the second part of Assumption 7 and the fact that $|\mathbf{u}_i|_\infty \leq |\mathbf{u}_i|_1$. ■

An easy situation is the case when g is bounded, that is, when Assumption 8 holds. When g is not bounded, a weaker result can still be proved under a "conditional Bernstein" assumption. This assumption will help us to control the tail behavior of g so that exponential concentration bounds are available.

Assumption 15 (conditional Bernstein assumption). There exists a positive constant $\tilde{B}_g$ such that for all $l \geq 1$ and $\mathbf{z}_1, \dots, \mathbf{z}_k \in \mathbb{R}^{kp}$, $\mathbb{E}\left[\left|g(\mathbf{X}_1, \dots, \mathbf{X}_k)\right|^l \,\middle|\, \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k\right] \leq \tilde{B}_g l!$.

To obtain tighter bounds, we will use the notation $B_{g,\mathbf{z}} := B_g(\mathbf{z}_1, \dots, \mathbf{z}_k)$ to denote a positive number such that $\mathbb{E}\left[\left|g(\mathbf{X}_1, \dots, \mathbf{X}_k)\right|^l \,\middle|\, \mathbf{Z}_1 = \mathbf{z}_1, \dots, \mathbf{Z}_k = \mathbf{z}_k\right] \leq B_g(\mathbf{z}_1, \dots, \mathbf{z}_k)^l l!$. Therefore, it is enough to prove the following result.

Proposition 5 (Exponential bound for the estimator $\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k)$, with fixed $\mathbf{z}_1, \dots \mathbf{z}_k \in \mathcal{Z}^k$). *Assume either Assumption 8 or the weaker Assumption 15. Under Assumptions 2, 4, 5, 12, 13, and 14, for every $t, t' > 0$ such that $C_{K,\alpha} h^\alpha / \alpha! + t < f_{\mathbf{Z},min}/2$, we have*

$$\mathbb{P}\left(\left|\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k) - \theta(\mathbf{z}_1, \dots, \mathbf{z}_k)\right| < (1 + C_3 h^\alpha + C_4 t) \times \left(C_5 h^{k+\alpha} + t'\right)\right)$$
$$\geq 1 - 2\exp\left(-\frac{[n/k] t^2 h^{kp}}{C_1 + C_2 t}\right) - 2\exp\left(-\frac{[n/k] t'^2 h^{kp}}{C_6 + C_7 t'}\right),$$

where $C_3 := 4 f_{\mathbf{Z},\max}^k f_{\mathbf{Z},\min}^{-2k} C_{K,\alpha}/\alpha!$, $C_4 := 4 f_{\mathbf{Z},\max}^k f_{\mathbf{Z},\min}^{-2k}$ and $C_5 := C_{g,f,\alpha} C'_{K,\alpha} f_{\mathbf{Z},\min}^{-k}/\alpha!$.

If Assumption 8 is satisfied, the result holds with the following values: $C_6 := 2 C_g^2 f_{\mathbf{Z},\max}^k f_{\mathbf{Z},\min}^{-2k} ||K||_2^{2k}$, $C_7 := (8/3) C_K^k C_g^k f_{\mathbf{Z},\min}^{-k}$; in the case of Assumption 15, the result holds with the following alternative values: $\tilde{C}_6 := 128\left(B_{g,\mathbf{z}} + \tilde{B}_g\right)^2 C_K^{2k-1} f_{\mathbf{Z},\min}^{-2k}$, $\tilde{C}_7 := 2\left(B_{g,\mathbf{z}} + \tilde{B}_g\right) C_K^k f_{\mathbf{Z},\min}^{-k}$.

We have the following decomposition

$$\begin{aligned}
&|\hat{\theta}(\mathbf{z}_{1:k}) - \theta(\mathbf{z}_{1:k})| \\
&= \left| N_k(\mathbf{z}_{1:k})^{-1} \frac{(n-k)!}{n!} \sum_{\sigma \in \mathfrak{I}_{k,n}} \prod_{i=1}^k K_h\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right) \left(g(\mathbf{X}_{\sigma(1:k)}) - \mathbb{E}\left[g(\mathbf{X}_{1:k}) \middle| \mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right) \right| \\
&= \frac{\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)}{N_k(\mathbf{z}_1, \dots, \mathbf{z}_k)} \cdot \left| \frac{(n-k)!}{n!} \sum_{\sigma \in \mathfrak{I}_{k,n}} \prod_{i=1}^k \frac{K_h\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right)}{f_{\mathbf{Z}}(\mathbf{z}_i)} \left(g(\mathbf{X}_{\sigma(1:k)}) - \mathbb{E}\left[g(\mathbf{X}_{1:k}) \middle| \mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right) \right| \\
&=: \frac{\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)}{N_k(\mathbf{z}_1, \dots, \mathbf{z}_k)} \cdot \left| \sum_{\sigma \in \mathfrak{I}_{k,n}} S_\sigma \right|.
\end{aligned}$$

The conclusion will follow from the next three lemmas, where we will bound separately $\prod_{i=1}^k f_{\mathbf{Z}}/N_k$, the bias term $\left|\sum_{\sigma \in \mathfrak{I}_{k,n}} \mathbb{E}[S_\sigma]\right|$ and the stochastic component $\left|\sum_{\sigma \in \mathfrak{I}_{k,n}} (S_\sigma - \mathbb{E}[S_\sigma])\right|$.

Lemma 6 (Bound for $\prod_{i=1}^k f_Z(z_i)/N_k$). *Under Assumptions 2, 4, 5, and 12, and if for some* $t>0$, $C_{K,\alpha}h^\alpha/\alpha! + t < f_{Z,\min}^k/2$, *we have*

$$\mathbb{P}\left(\left|\frac{1}{N_k(\mathbf{z}_{1:k})} - \frac{1}{\prod_{i=1}^k f_Z(\mathbf{z}_i)}\right| \leq \frac{4}{f_{\mathbf{Z},\min}^{2k}}\left(\frac{C_{K,\alpha}h^\alpha}{\alpha!} + t\right)\right) \geq 1 - 2\exp\left(-\frac{[n/k]t^2}{2h^{-kp}f_{\mathbf{Z},\max}^k ||K||_2^{2k} + (4/3)h^{-kp}C_K^k t}\right),$$

and on the same event, $N_k(\mathbf{z}_{1:k})$ *is strictly positive and*

$$\frac{\prod_{i=1}^k f_Z(\mathbf{z}_i)}{N_k(\mathbf{z}_{1:k})} \leq 1 + \frac{4f_{\mathbf{Z},\max}^k}{f_{\mathbf{Z},\min}^{2k}}\left(\frac{C_{K,\alpha}h^\alpha}{\alpha!} + t\right).$$

Proof. Using the mean value inequality for the function $x \mapsto 1/x$, we get

$$\left|\frac{1}{N_k(\mathbf{z}_{1:k})} - \frac{1}{\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)}\right| \leq \frac{1}{N_*^2}\left|N_k(\mathbf{z}_{1:k}) - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)\right|,$$

where N_* lies between $N_k(\mathbf{z}_{1:k})$ and $\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)$. By Lemma 1, we get

$$\mathbb{P}\left(\left|N_k(\mathbf{z}_{1:k}) - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)\right| \leq \frac{C_{K,\alpha}}{\alpha!}h^\alpha + t\right) \geq 1 - 2\exp\left(-\frac{[n/k]t^2}{2h^{-kp}f_{\mathbf{Z},\max}^k ||K||_2^{2k} + (4/3)h^{-kp}C_K^k t}\right).$$

On this event, $\left|N_k(\mathbf{z}_{1:k}) - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)\right| \leq (1/2)\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)$ by assumption, so that $f_{\mathbf{Z},\min}^k/2 \leq N_k(\mathbf{z}_{1:k})$. We have also $f_{\mathbf{Z},\min}^k/2 \leq \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)$. Thus, we have $f_{\mathbf{Z},\min}^k/2 \leq N_*$. Combining the previous inequalities, we finally get

$$\left|\frac{1}{N_k(\mathbf{z}_{1:k})} - \frac{1}{\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)}\right| \leq \frac{1}{N_*^2}\left|N_k(\mathbf{z}_{1:k}) - \prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)\right| \leq \frac{4}{f_{\mathbf{Z},\min}^{2k}}\left(\frac{C_{K,\alpha}h^\alpha}{\alpha!} + t\right).$$

■

Now, we provide a bound on the bias.

Lemma 7. *Under Assumptions 2 and 14, we have* $\left|\mathbb{E}[S_\sigma]\right| \leq C_{g,f,\alpha}C_{K,\alpha}h^{k\alpha}/(f_{\mathbf{Z},\min}^k \alpha!)$.

Proof. We remark that

$$\begin{aligned} 0 &= \int \left(g(\mathbf{x}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right) f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_1}(\mathbf{x}_1)\cdots f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}_k}(\mathbf{x}_k)d\mu^{\otimes k}(\mathbf{x}_{1:k}) \\ &= \int \left(g(\mathbf{x}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right) \frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{z}_1)\cdots f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_k,\mathbf{z}_k)}{\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)} d\mu^{\otimes k}(\mathbf{x}_{1:k}). \end{aligned} \tag{D1}$$

We have

$$\mathbb{E}[S_\sigma] = \mathbb{E}\left[\frac{K_h(\mathbf{Z}_{\sigma(1)} - \mathbf{z}_1)\cdots K_h(\mathbf{Z}_{\sigma(k)} - \mathbf{z}_k)}{\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)}\left(g\left(\mathbf{X}_{\sigma(1)}, \ \dots \ , \mathbf{X}_{\sigma(k)}\right) - \mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right)\right]$$

$$= \int \left(g(\mathbf{x}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right)\prod_{i=1}^k \frac{K(\mathbf{u}_i)}{f_{\mathbf{Z}}(\mathbf{z}_i)} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i + h\mathbf{u}_i) \ d\mu(\mathbf{x}_i) d\mathbf{u}_i$$

$$= \int \left(g(\mathbf{x}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right)\left(\prod_{i=1}^k f_{\mathbf{X},\mathbf{Z}}\left(\mathbf{x}_i, \mathbf{z}_i + h\mathbf{u}_i\right) - \prod_{i=1}^k f_{\mathbf{X},\mathbf{Z}}\left(\mathbf{x}_i, \mathbf{z}_i\right)\right)$$

$$\prod_{i=1}^k \frac{K(\mathbf{u}_i)}{f_{\mathbf{Z}}(\mathbf{z}_i)} \ d\mu(\mathbf{x}_i) d\mathbf{u}_i.$$

We apply now the Taylor–Lagrange formula to the function

$$\phi_{\mathbf{x}_{1:k},\mathbf{u}_{1:k}}(t) := \prod_{i=1}^k f_{\mathbf{X},\mathbf{Z}}\left(\mathbf{x}_i, \mathbf{z}_i + h\mathbf{u}_i\right),$$

and get

$$\mathbb{E}[S_\sigma]$$

$$= \int \left(g(\mathbf{x}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right)\left(\phi_{\mathbf{x}_{1:k},\mathbf{u}_{1:k}}(t)(1) - \phi_{\mathbf{x}_{1:k},\mathbf{u}_{1:k}}(t)(0)\right)\prod_{i=1}^k \frac{K(\mathbf{u}_i)}{f_{\mathbf{Z}}(\mathbf{z}_i)} \ d\mu(\mathbf{x}_i) d\mathbf{u}_i$$

$$= \int \left(g(\mathbf{x}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right)$$

$$\cdot \left(\sum_{j=1}^{\alpha-1} \frac{1}{j!} \phi_{\mathbf{x}_{1:k},\mathbf{u}_{1:k}}(t)^{(j)}(0) + \frac{1}{\alpha!} \phi_{\mathbf{x}_{1:k},\mathbf{u}_{1:k}}(t)^{(\alpha)}(t_{\mathbf{x},\mathbf{u}})\right) \prod_{i=1}^k \frac{K(\mathbf{u}_i)}{f_{\mathbf{Z}}(\mathbf{z}_i)} \ d\mu(\mathbf{x}_i) d\mathbf{u}_i$$

$$= \int \left(g(\mathbf{x}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right)$$

$$\cdot \left(\frac{1}{\alpha!} \phi_{\mathbf{x}_{1:k},\mathbf{u}_{1:k}}(t)^{(\alpha)}(t_{\mathbf{x},\mathbf{u}})\right) \prod_{i=1}^k \frac{K(\mathbf{u}_i)}{f_{\mathbf{Z}}(\mathbf{z}_i)} \ d\mu(\mathbf{x}_i) d\mathbf{u}_i$$

$$= \int \left(g(\mathbf{x}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right)$$

$$\cdot \frac{1}{\alpha!} \left(\phi_{\mathbf{x}_{1:k},\mathbf{u}_{1:k}}(t)^{(\alpha)}(t_{\mathbf{x},\mathbf{u}}) - \phi_{\mathbf{x}_{1:k},\mathbf{u}_{1:k}}(t)^{(\alpha)}(0)\right)$$

$$\prod_{i=1}^k \frac{K(\mathbf{u}_i)}{f_{\mathbf{Z}}(\mathbf{z}_i)} \ d\mu(\mathbf{x}_i) d\mathbf{u}_i.$$

For every real t, we have

$$\phi^{(\alpha)}(t) = \sum_{m_1+\cdots+m_k=\alpha} \binom{n}{m_{1:k}} \prod_{i=1}^{k} \frac{\partial^{m_i}\left(f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i + ht\mathbf{u}_i)\right)}{\partial t^{m_i}}$$

$$= \sum_{m_1+\cdots+m_k=\alpha} \binom{n}{m_{1:k}} \prod_{i=1}^{k} \sum_{j_1,\ldots j_{m_i}=1}^{p} h^{m_i} u_{i,j_1} \ldots u_{i,j_{m_i}} \frac{\partial^{m_i} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \cdots \partial z_{j_{m_i}}}(\mathbf{x}_i, \mathbf{z}_i + ht\mathbf{u}_i) \quad \text{(D2)}$$

$$= h^{\alpha} \sum_{m_1+\cdots+m_k=\alpha} \binom{n}{m_{1:k}} \prod_{i=1}^{k} \sum_{j_1,\ldots j_{m_i}=1}^{p} u_{i,j_1} \ldots u_{i,j_{m_i}} \frac{\partial^{m_i} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \cdots \partial z_{j_{m_i}}}(\mathbf{x}_i, \mathbf{z}_i + ht\mathbf{u}_i).$$

Therefore, we get

$$\mathbb{E}[S_\sigma] = \sum_{m_1+\cdots+m_k=\alpha} \binom{n}{m_{1:k}} \int \prod_{i=1}^{k} \frac{K(\mathbf{u}_i)}{\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)} \sum_{j_1,\ldots j_{m_i}=1}^{p} u_{i,j_1} \ldots u_{i,j_{m_i}}$$

$$\cdot \left(g(\mathbf{x}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k}) \middle| \mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right)$$

$$\cdot \left(\frac{\partial^{m_i} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \ldots \partial z_{j_{m_i}}}(\mathbf{x}_i, \mathbf{z}_i + ht\mathbf{u}_i) - \frac{\partial^{m_i} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{j_1} \ldots \partial z_{j_{m_i}}}(\mathbf{x}_i, \mathbf{z}_i)\right) d\mu(\mathbf{x}_1) d\mathbf{u}_1 \ldots d\mu(\mathbf{x}_k) d\mathbf{u}_k,$$

and, using Assumption 7, this yields

$$\left|\mathbb{E}[S_\sigma]\right| \leq \frac{C_{g,f,\alpha} C_{K,\alpha} h^{\alpha+k}}{f_{\mathbf{Z},\min}^k \alpha!}.$$

■

Now we bound the stochastic component. We have the following equality

$$\left|\sum_{\sigma \in \mathfrak{I}_{k,n}} (S_\sigma - \mathbb{E}[S_\sigma])\right| = \left|\frac{(n-k)!}{n!} \sum_{\sigma \in \mathfrak{I}_{k,n}} g\left((\mathbf{X}_{\sigma(1)}, \mathbf{Z}_{\sigma(1)}), \ldots, (\mathbf{X}_{\sigma(k)}, \mathbf{Z}_{\sigma(k)})\right)\right|$$

with the function $\tilde{g}$ defined by

$$\tilde{g}\left((\mathbf{X}_1, \mathbf{Z}_1), \ldots, (\mathbf{X}_k, \mathbf{Z}_k)\right)$$

$$= \frac{K_h(\mathbf{Z}_1 - \mathbf{z}_1) \cdots K_h(\mathbf{Z}_k - \mathbf{z}_k)}{\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)} \left(g(\mathbf{X}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k}) \middle| \mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right)$$

$$- \mathbb{E}\left[\frac{K_h(\mathbf{Z}_1 - \mathbf{z}_1) \cdots K_h(\mathbf{Z}_k - \mathbf{z}_k)}{\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}_i)} \left(g(\mathbf{X}_{1:k}) - \mathbb{E}\left[g(\mathbf{X}_{1:k}) \middle| \mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right)\right]$$

By construction, $\mathbb{E}\left[\tilde{g}\left((\mathbf{X}_1, \mathbf{Z}_1), \ldots, (\mathbf{X}_k, \mathbf{Z}_k)\right)\right] = 0$. If $\tilde{g}$ is bounded, we can derive an immediate bound for this stochastic component. Indeed, we would have $||\tilde{g}||_\infty \leq 4C_K^k h^{-kp} C_g^k / f_{\mathbf{Z},\min}^k$.

Moreover, we have

$$\mathrm{Var}\left[\tilde{g}\left((\mathbf{X}_1,\mathbf{Z}_1)\,,\,\dots\,,\,(\mathbf{X}_k,\mathbf{Z}_k)\right)\right] \leq \mathbb{E}\left[\frac{K_h^2(\mathbf{Z}_1-\mathbf{z}_1)\cdots K_h^2(\mathbf{Z}_k-\mathbf{z}_k)}{\prod_{i=1}^k f_{\mathbf{Z}}^2(\mathbf{z}_i)} g^2(\mathbf{X}_1,\,\dots\,,\mathbf{X}_k)\right]$$
$$\leq C_g^2 f_{\mathbf{Z},\max}^k f_{\mathbf{Z},\min}^{-2k} h^{-kp} ||K||_2^{2k}.$$

Therefore, we can apply Lemma 3, and we get

$$\mathbb{P}\left(\left|\sum_{\sigma\in\mathfrak{I}_{k,n}} (S_\sigma - \mathbb{E}[S_\sigma])\right| > t\right) \leq 2\exp\left(-\frac{[n/k]t^2}{2C_g^2 f_{\mathbf{Z},\max}^k f_{\mathbf{Z},\min}^{-2k} h^{-kp}||K||_2^{2k} + (8/3)C_K^k h^{-kp} C_g^k f_{\mathbf{Z},\min}^{-k} t}\right).$$

In the following Lemma 8, our goal will be to bound the stochastic component using only Assumption 15 on the conditional moments of g.

Lemma 8. *Under Assumptions 2, 5, and 15, for every $t > 0$, we have*

$$\mathbb{P}\left(\sum_{\sigma\in\mathfrak{I}_{k,n}} S_\sigma - \mathbb{E}[S_\sigma] > t\right) \leq \exp\left(-\frac{t^2 f_{\mathbf{Z},\min}^{2k} h^{kp}[n/k]}{128\left(B_{g,\mathbf{z}}+\tilde{B}_g\right)^2 C_K^{2k-1} + 2t\left(B_{g,\mathbf{z}}+\tilde{B}_g\right) C_K^k f_{\mathbf{Z},\min}^k}\right).$$

Proof. Using the same decomposition for U-statistics as in Hoeffding (1963), we obtain

$$\sum_{\sigma\in\mathfrak{I}_{k,n}} S_\sigma - \mathbb{E}[S_\sigma] = \frac{1}{n!}\sum_{\sigma\in\mathfrak{I}_{n,n}} \frac{1}{[n/k]}\sum_{i=1}^{[n/k]} V_{n,i,\sigma},$$

where

$$V_{n,i,\sigma} := \tilde{g}\left(\left(\mathbf{X}_{\sigma(1+(i-1)k)},\mathbf{Z}_{\sigma(2+(i-1)k)}\right),\,\dots\,,\left(\mathbf{X}_{\sigma(ik)},\mathbf{Z}_{\sigma(jk)}\right)\right).$$

For any $\lambda > 0$, we have

$$\begin{aligned}
\mathbb{P}\left(\sum_{\sigma\in\mathfrak{I}_{k,n}} S_\sigma - \mathbb{E}[S_\sigma] > t\right) &\leq e^{-\lambda t}\mathbb{E}\left[\exp\left(\lambda\sum_{\sigma\in\mathfrak{I}_{k,n}} S_\sigma - \mathbb{E}[S_\sigma]\right)\right] \\
&\leq e^{-\lambda t}\mathbb{E}\left[\exp\left(\lambda\frac{1}{n!}\sum_{\sigma\in\mathfrak{I}_{n,n}}\frac{1}{[n/k]}\sum_{i=1}^{[n/k]} V_{n,i,\sigma}\right)\right] \\
&\leq e^{-\lambda t}\frac{1}{n!}\sum_{\sigma\in\mathfrak{I}_{n,n}}\mathbb{E}\left[\exp\left(\lambda\frac{1}{[n/k]}\sum_{i=1}^{[n/k]} V_{n,i,\sigma}\right)\right] \qquad (D3) \\
&\leq e^{-\lambda t}\frac{1}{n!}\sum_{\sigma\in\mathfrak{I}_{n,n}}\prod_{i=1}^{[n/k]}\mathbb{E}\left[\exp\left(\lambda\frac{1}{[n/k]}V_{n,i,\sigma}\right)\right] \\
&\leq e^{-\lambda t}\left(\sup_{\sigma\in\mathfrak{I}_{n,n},\; i=1,\dots,[n/k]}\mathbb{E}\left[\exp\left(\lambda[n/k]^{-1}V_{n,i,\sigma}\right)\right]\right)^{[n/k]}.
\end{aligned}$$

Let $l \geq 2$. Using the inequality $(a+b+c+d)^l \leq 4^l(a^l+b^l+c^l+d^l)$, we get

$$\begin{aligned}
\mathbb{E}\left[|V_{n,i,\sigma}|^l\right] = \mathbb{E}\left[|V_{n,1,\sigma}|^l\right] &\leq 4^l\, \mathbb{E}\left[|g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)})|^l \prod_{i=1}^k \frac{|K_h|^l\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right)}{f_{\mathbf{Z}}^l(\mathbf{z}_i)}\right] \\
&+ 4^l\, \mathbb{E}\left[\left|\mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right|^l \prod_{i=1}^k \frac{|K_h|^l\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right)}{f_{\mathbf{Z}}^l(\mathbf{z}_i)}\right] \\
&+ 4^l \left|\mathbb{E}\left[g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)}) \prod_{i=1}^k \frac{K_h\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right)}{f_{\mathbf{Z}}^l(\mathbf{z}_i)}\right]\right|^l \\
&+ 4^l \left|\mathbb{E}\left[\left|\mathbb{E}\left[g(\mathbf{X}_{1:k})\middle|\mathbf{Z}_{1:k} = \mathbf{z}_{1:k}\right]\right| \prod_{i=1}^k \frac{K_h\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right)}{f_{\mathbf{Z}}^l(\mathbf{z}_i)}\right]\right|^l.
\end{aligned}$$

Using Jensen's inequality for the function $x \mapsto |x|^p$ with the second, third, and fourth terms, and the law of iterated expectations for the first and the third terms, we get

$$\begin{aligned}
\mathbb{E}\left[|V_{n,i,\sigma}|^l\right] &\leq 4^l \cdot 2\, \mathbb{E}\left[\mathbb{E}\left[|g(\mathbf{X}_{\sigma(1)}, \dots, \mathbf{X}_{\sigma(k)})|^l \middle| \mathbf{Z}_{\sigma(1)}, \dots, \mathbf{Z}_{\sigma(k)}\right] \prod_{i=1}^k \frac{|K_h|^l\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right)}{f_{\mathbf{Z}}^l(\mathbf{z}_i)}\right] \\
&+ 4^l \cdot 2\, \mathbb{E}\left[\mathbb{E}\left[\left|g(\mathbf{X}_{1:k})\right|^l \middle| \mathbf{Z}_i = \mathbf{z}_i, \forall i = 1, \dots, k\right] \prod_{i=1}^k \frac{|K_h|^l\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right)}{f_{\mathbf{Z}}^l(\mathbf{z}_i)}\right] \\
&\leq 4^l \cdot 2\, \mathbb{E}\left[\left(B_g^l(\mathbf{Z}_1, \dots, \mathbf{Z}_k) + B_g^l(\mathbf{z}_1, \dots, \mathbf{z}_k)\right)^l l! \prod_{i=1}^k \frac{|K_h|^l\left(\mathbf{Z}_{\sigma(i)} - \mathbf{z}_i\right)}{f_{\mathbf{Z}}^l(\mathbf{z}_i)}\right] \\
&\leq 4^l \cdot 2\, \left(\tilde{B}_g^l + B_g^l(\mathbf{z}_1, \dots, \mathbf{z}_k)\right) l!(h^{-kp} C_K^k f_{\mathbf{Z},\min}^{-k})^{l-1} f_{\mathbf{Z},\min}^{-k} \\
&\leq 2\left(4\left(\tilde{B}_g + B_{g,\mathbf{z}}\right) h^{-kp} C_K^k f_{\mathbf{Z},\min}^{-k}\right)^l l!\, h^{kp} C_K^{-1},
\end{aligned}$$

where $B_{g,\mathbf{z}} := B_g(\mathbf{z}_1, \dots, \mathbf{z}_k)$. Remarking that $\mathbb{E}[V_{n,i,\sigma}] = 0$ by construction of $\tilde{g}$, we obtain

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda[n/k]^{-1} V_{n,i,\sigma}\right)\right] &= 1 + \sum_{l=2}^{\infty} \frac{\mathbb{E}\left[(\lambda[n/k]^{-1} V_{n,i,\sigma})^l\right]}{l!} \\
&\leq 1 + 2C_K^{-1} h^{kp} \sum_{l=2}^{\infty} (4\lambda[n/k]^{-1}\left(B_{g,\mathbf{z}} + \tilde{B}_g\right) h^{-kp} C_K^k f_{\mathbf{Z},\min}^{-k})^l \\
&\leq 1 + 2C_K^{-1} h^{kp} \cdot \frac{\left(4\lambda[n/k]^{-1}\left(B_{g,\mathbf{z}} + \tilde{B}_g\right) h^{-kp} C_K^k f_{\mathbf{Z},\min}^{-k}\right)^2}{1 - 4\lambda[n/k]^{-1}\left(B_{g,\mathbf{z}} + \tilde{B}_g\right) h^{-kp} C_K^k f_{\mathbf{Z},\min}^{-k}} \\
&\leq \exp\left(\frac{32\lambda^2[n/k]^{-2}\left(B_{g,\mathbf{z}} + \tilde{B}_g\right)^2 h^{-kp} C_K^{2k-1} f_{\mathbf{Z},\min}^{-2k}}{1 - 4\lambda[n/k]^{-1}\left(B_{g,\mathbf{z}} + \tilde{B}_g\right) h^{-kp} C_K^k f_{\mathbf{Z},\min}^{-k}}\right),
\end{aligned}$$

where the last statement follows from the inequality $1 + x \le \exp(x)$. Combining the latter bound with Equation (D3), we get

$$\mathbb{P}\left(\sum_{\sigma\in\mathfrak{I}_{k,n}} S_\sigma - \mathbb{E}[S_\sigma] > t\right) \le \exp\left(-\lambda t + \frac{32\lambda^2\left(B_{g,\mathbf{z}}+\tilde{B}_g\right)^2 C_K^{2k-1}}{f_{\mathbf{Z},\min}^{2k} h^{kp}[n/k] - 4\lambda\left(B_{g,\mathbf{z}}+\tilde{B}_g\right)C_K^k f_{\mathbf{Z},\min}^k}\right). \tag{D4}$$

Remarking that the right-hand side term inside the exponential is of the form $-\lambda t + \frac{a\lambda^2}{b-c\lambda}$, we choose the value

$$\lambda_* = \frac{tb}{2a+tc} = \frac{t f_{\mathbf{Z},\min}^{2k} h^{kp}[n/k]}{64\left(B_{g,\mathbf{z}}+\tilde{B}_g\right)^2 C_K^{2k-1} + t\left(B_{g,\mathbf{z}}+\tilde{B}_g\right)C_K^k f_{\mathbf{Z},\min}^k}, \tag{D5}$$

such that $-\lambda_* t + \frac{a\lambda_*^2}{b-c\lambda_*} = -\frac{t^2 b}{4a+2ct} = -\frac{t}{2}\lambda_*$. Therefore, the right-hand side term of Equation (D4) can be simplified, and combining this with Equation (D5), we obtain

$$\mathbb{P}\left(\sum_{\sigma\in\mathfrak{I}_{k,n}} S_\sigma - \mathbb{E}[S_\sigma] > t\right) \le \exp\left(-\frac{t^2 f_{\mathbf{Z},\min}^{2k} h^{kp}[n/k]}{128\left(B_{g,\mathbf{z}}+\tilde{B}_g\right)^2 C_K^{2k-1} + 2t\left(B_{g,\mathbf{z}}+\tilde{B}_g\right) C_K^k f_{\mathbf{Z},\min}^k}\right).$$

■

D.3 Proof of Theorem 1

By Proposition 2, for every $t_1, t_2 > 0$ such that $C_{K,\alpha}h^\alpha/\alpha! + t < f_{\mathbf{Z},\min}/2$, we have

$$\begin{aligned}
&\mathbb{P}\left(|\hat{\theta}(\mathbf{z}_1, \dots, \mathbf{z}_k) - \theta(\mathbf{z}_1, \dots, \mathbf{z}_k)| < (1 + C_3 h^\alpha + C_4 t_1) \times \left(C_5 h^{k+\alpha} + t_2\right)\right) \\
&\ge 1 - 2\exp\left(-\frac{[n/k]t_1^2 h^{kp}}{C_1 + C_2 t_1}\right) - 2\exp\left(-\frac{[n/k]t_2^2 h^{kp}}{C_6 + C_7 t_2}\right),
\end{aligned}$$

We apply this proposition to every k-tuple $\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)$ where $\sigma \in \mathfrak{I}_{k,n'}$. Combining it with Assumption 9, we get

$$\begin{aligned}
&\mathbb{P}\left(\sup_i |\xi_{i,n}| < C_{\Lambda'}(1 + C_3 h^\alpha + C_4 t_1) \times \left(C_5 h^{k+\alpha} + t_2\right)\right) \\
&\ge 1 - 2\sum_{i=1}^{|\mathfrak{I}_{k,n'}|}\left[\exp\left(-\frac{[n/k]t_1^2 h^{kp}}{C_1 + C_2 t_1}\right) + \exp\left(-\frac{[n/k]t_2^2 h^{kp}}{C_6 + C_7 t_2}\right)\right],
\end{aligned}$$

Choosing $t_1 := f_{\mathbf{Z},\min}/4$ and using the bound (7) on h, we get

$$\begin{aligned}
&\mathbb{P}\left(\sup_i |\xi_{i,n}| < C_{\Lambda'}\left(1 + C_3\frac{f_{\mathbf{Z},\min}\alpha!}{4C_{K,\alpha}} + C_4\frac{f_{\mathbf{Z},\min}}{4}\right) \times \left(C_5 h^{k+\alpha} + t_2\right)\right) \\
&\ge 1 - 2\sum_{i=1}^{|\mathfrak{I}_{k,n'}|}\left[\exp\left(-\frac{[n/k]f_{\mathbf{Z},\min}^2 h^{kp}}{16C_1 + 4C_2 f_{\mathbf{Z},\min}}\right) + \exp\left(-\frac{[n/k]t_2^2 h^{kp}}{C_6 + C_7 t_2}\right)\right].
\end{aligned}$$

Choosing $t_2 = t/(2C_8) = t/\left(2C_\psi C_{\Lambda'}\left(1 + C_3\frac{f_{\mathbf{Z},min}\alpha!}{4C_{K,\alpha}} + C_4\frac{f_{\mathbf{Z},min}}{4}\right)\right)$, and using the bound (7) on h^α, we get

$$\mathbb{P}\left(\sup_i|\xi_{i,n}| < t/C_\psi\right) \geq 1 - 2\sum_{i=1}^{|\mathfrak{I}_{k,n'}|}\left[\exp\left(-\frac{[n/k]f^2_{\mathbf{Z},\min}h^{kp}}{16C_1+4C_2f_{\mathbf{Z},\min}}\right) + \exp\left(-\frac{[n/k]t^2h^{kp}}{4C_8^2C_6+2C_8C_7t}\right)\right].$$

This gives the values of $C_{6,\sigma}$ and $C_{7,\sigma}$. Note that, if Assumption 8 is satisfied, the result holds with $C_{6,\sigma}$ and $C_{7,\sigma}$ constant, respectively to C_6 and C_7 defined in Proposition 5. In the case of Assumption 15, the result holds with the following alternative values: $C_{6,\sigma} := 128\left(B_g(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}) + \tilde{B}_g\right)^2 C_K^{2k} f_{\mathbf{Z},\min}^{-2k}$ and $C_{7,\sigma} := 2\left(B_g(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}) + \tilde{B}_g\right) C_K^k f_{\mathbf{Z},\min}^{-k}$.

On the same event, we have $\max_{j=1,\dots,p'}\left|\frac{1}{n'}\sum_{i=1}^{n'} Z'_{i,j}\xi_{i,n}\right| \leq t$, by Assumption 9. The conclusion results from the following lemma.

Lemma 9 (From Derumigny and Fermanian, 2020, lemma 25). *Assume that* $\max_{j=1,\dots,p'}\left|\frac{1}{n'}\sum_{i=1}^{n'} Z'_{i,j}\xi_{i,n}\right| \leq t$, *for some* $t > 0$, *that the assumption* RE$(s, 3)$ *is satisfied, and that the tuning parameter is given by* $\lambda = \gamma t$, *with* $\gamma \geq 4$. *Then,* $||\mathbb{Z}'(\hat{\beta} - \beta^*)|| \leq \dfrac{4(\gamma+1)t\sqrt{s}}{\kappa(s,3)}$ *and* $|\hat{\beta} - \beta^*|_q \leq \dfrac{4^{2/q}(\gamma+1)ts^{1/q}}{\kappa^2(s,3)}$, *for every* $1 \leq q \leq 2$.

■

APPENDIX E. PROOF OF THEOREM 6

We detail the assumption which we will use to prove Theorem 6.

Assumption 16.

(i) The support of the kernel $K(\cdot)$ is included into $[-1,1]^p$. Moreover, for all n, n' and every $(i,j) \in \{1, \dots, n'\}^2$, $i \neq j$, we have $|\mathbf{z}'_i - \mathbf{z}'_j|_\infty > 2h_{n,n'}$.

(ii) (a) $n'(nh_{n,n'}^{p+4\alpha} + h_{n,n'}^{2\alpha} + h_{n,n'}^p + (nh_{n,n'}^p)^{-1}) \to 0$, (b) $\lambda_{n,n'}(n'\ n\ h_{n,n'}^p)^{1/2} \to 0$, (c) $n'\ n\ h_{n,n'}^p \to \infty$ and $n\ h_{n,n'}^{p+2\alpha-\epsilon}/\ln n' \to \infty$ for some $\epsilon \in [0, 2\alpha[$.

(iii) The distribution $\mathbb{P}_{\mathbf{z}',n'} := |\mathfrak{I}_{k,n'}|^{-1}\sum_{\sigma\in\mathfrak{I}_{k,n'}}\delta_{(\mathbf{z}'_{\sigma(1)},\dots,\mathbf{z}'_{\sigma(k)})}$ weakly converges as $n' \to \infty$, to a distribution $\mathbb{P}_{\mathbf{z}',k,\infty}$ on $\mathbb{R}^{kp}$. There exists a distribution $\mathbb{P}_{\mathbf{z}',\infty}$ on $\mathbb{R}^{kp}$, with a density $f_{\mathbf{z}',\infty}$ with respect to the p-dimensional Lebesgue measure such that $\mathbb{P}_{\mathbf{z}',k,\infty} = \mathbb{P}_{\mathbf{z}',\infty}^{\otimes k}$.

(iv) The matrix $V_1 := \int \boldsymbol{\psi}(\mathbf{z}'_1, \dots \mathbf{z}'_k)\boldsymbol{\psi}(\mathbf{z}'_1, \dots \mathbf{z}'_k)^T f_{\mathbf{z}',\infty}(\mathbf{z}'_1) \dots f_{\mathbf{z}',\infty}(\mathbf{z}'_k)d\mathbf{z}'_1 \dots d\mathbf{z}'_k$ is nonsingular.

(v) $\Lambda(\cdot)$ is two times continuously differentiable. Let $\mathcal{T}$ be the range of θ, from $\mathcal{Z}^k$ toward $\mathbb{R}$. On an open neighborhood of $\mathcal{T}$, the second derivative of $\Lambda(\cdot)$ is bounded by a constant $C_{\Lambda''}$.

(vi) Several integrals exist and are finite, including

$$\tilde{V}_1 := \int \theta\left(\mathbf{z}'_1, \dots, \mathbf{z}'_k\right)\Lambda'\left(\theta\left(\mathbf{z}'_1, \dots, \mathbf{z}'_k\right)\right)\boldsymbol{\psi}\left(\mathbf{z}'_1, \dots, \mathbf{z}'_k\right)$$
$$f_{\mathbf{z}',\infty}(\mathbf{z}_1)\cdots f_{\mathbf{z}',\infty}(\mathbf{z}_k)\, d\mathbf{z}_1\cdots d\mathbf{z}_k \text{ and}$$

$$V_2 := \int \frac{||K||_2^2}{f_{\mathbf{Z}}(\mathbf{z}'_1)} g(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) g(\mathbf{x}_1, \mathbf{y}_2, \dots, \mathbf{y}_k)$$
$$\Lambda'^2 \left(\theta(\mathbf{z}'_1, \dots, \mathbf{z}'_k)\right) \boldsymbol{\psi}\left(\mathbf{z}'_1, \dots, \mathbf{z}'_k\right) \boldsymbol{\psi}\left(\mathbf{z}'_1, \dots, \mathbf{z}'_k\right)^T$$
$$\times f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}'_1}(\mathbf{x}_1)\, d\mu(\mathbf{x}_1) d\mu(\mathbf{z}'_1) \prod_{i=2}^{k} f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}'_i}(\mathbf{y}_i) f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}'_i}(\mathbf{x}_i) f_{\mathbf{z}',\infty}(\mathbf{z}'_i)\, d\mu(\mathbf{x}_i) d\mu(\mathbf{y}_i) d\mathbf{z}'_i.$$

Define $\tilde{r}_{n,n'} := (n \times n' \times h^p_{n,n'})^{1/2}$, $\mathbf{u} := \tilde{r}_{n,n'}(\beta - \beta^*)$ and $\hat{\mathbf{u}}_{n,n'} := \tilde{r}_{n,n'}(\hat{\beta}_{n,n'} - \beta^*)$, so that $\hat{\beta}_{n,n'} = \beta^* + \hat{\mathbf{u}}_{n,n'}/\tilde{r}_{n,n'}$. We define for every $\mathbf{u} \in \mathbb{R}^{p'}$,

$$\begin{aligned} \mathbb{F}_{n,n'}(\mathbf{u}) &:= \frac{-2\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \xi_{\sigma,n} \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)^T \mathbf{u} \\ &+ \frac{1}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \left\{ \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)^T \mathbf{u} \right\}^2 + \lambda_{n,n'} \tilde{r}^2_{n,n'} \left(\left| \beta^* + \frac{\mathbf{u}}{\tilde{r}_{n,n'}} \right|_1 - |\beta^*|_1 \right), \end{aligned} \tag{E1}$$

and we obtain $\hat{\mathbf{u}}_{n,n'} = \arg\min_{\mathbf{u} \in \mathbb{R}^{p'}} \mathbb{F}_{n,n'}(\mathbf{u})$ applying Lemma 9.

Lemma 10. *Under the same assumptions as in Theorem 6,*

$$T_1 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \xi_{\sigma,n} \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right) \xrightarrow{D} \mathcal{N}(0, V_2).$$

This lemma is proved in E.1. It will help to control the first term of Equation (E1), which is simply $-2T_1^T\mathbf{u}$.

Concerning the second term of Equation (E1), using Assumption 16(iii), we have for every $\mathbf{u} \in \mathbb{R}^{p'}$

$$\begin{aligned} &\frac{1}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \left\{ \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)^T \mathbf{u} \right\}^2 \\ &\to \int \left(\boldsymbol{\psi}(\mathbf{z}'_1, \dots, \mathbf{z}'_k)^T \mathbf{u}\right)^2 f_{\mathbf{z}',\infty}(\mathbf{z}'_1) \dots f_{\mathbf{z}',\infty}(\mathbf{z}'_k)\, d\mathbf{z}'_1 \dots d\mathbf{z}'_k. \end{aligned} \tag{E2}$$

This has to be read as a convergence of a sequence of real numbers indexed by $\mathbf{u}$, because the design points $\mathbf{z}'_i$ are deterministic. We also have, for any $\mathbf{u} \in \mathbb{R}^{p'}$ and when n is large enough,

$$\left| \beta^* + \frac{\mathbf{u}}{\tilde{r}_{n,n'}} \right|_1 - \left| \beta^* \right|_1 = \sum_{i=1}^{p'} \left(\frac{|u_i|}{\tilde{r}_{n,n'}} \mathbb{1}_{\{\beta^*_i = 0\}} + \frac{u_i}{\tilde{r}_{n,n'}} \mathrm{sign}(\beta^*_i) \mathbb{1}_{\{\beta^*_i \neq 0\}} \right).$$

Therefore, by Assumption 16(ii)(b), for every $\mathbf{u} \in \mathbb{R}^{p'}$,

$$\lambda_{n,n'} \tilde{r}^2_{n,n'} \left(\left| \beta^* + \frac{\mathbf{u}}{\tilde{r}_{n,n'}} \right|_1 - \left| \beta^* \right|_1 \right) \to 0, \tag{E3}$$

when (n, n') tends to the infinity. Combining Lemma 10 and Equations (E1)–(E3), and defining the function $\mathbb{F}_{\infty,\infty}$ by

$$\mathbb{F}_{\infty,\infty}(\mathbf{u}) := 2\tilde{\mathbf{W}}^T \mathbf{u} + \int \left(\boldsymbol{\psi}(\mathbf{z}'_1, \dots, \mathbf{z}'_k)^T \mathbf{u}\right)^2 f_{\mathbf{z}',\infty}(\mathbf{z}'_1) \dots f_{\mathbf{z}',\infty}(\mathbf{z}'_k)\, d\mathbf{z}'_1 \dots d\mathbf{z}'_k,$$

where $\mathbf{u} \in \mathbb{R}^r$ and $\tilde{W} \sim \mathcal{N}(0, V_2)$, we obtain that every finite-dimensional margin of $\mathbb{F}_{n,n'}$ weakly converges to the corresponding margin of $\mathbb{F}_{\infty,\infty}$. Now, applying the convexity lemma, we get

$$\hat{\mathbf{u}}_{n,n'} \xrightarrow{D} \mathbf{u}_{\infty,\infty}, \text{ where } \mathbf{u}_{\infty,\infty} := \arg\min_{\mathbf{u}\in\mathbb{R}^r} \mathbb{F}_{\infty,\infty}(\mathbf{u}).$$

Since $\mathbb{F}_{\infty,\infty}(\mathbf{u})$ is a continuously differentiable convex function, apply the first-order condition $\nabla \mathbb{F}_{\infty,\infty}(\mathbf{u}) = 0$, which yields

$$2\tilde{\mathbf{W}} + 2\int \boldsymbol{\psi}(\mathbf{z}'_1, \dots, \mathbf{z}'_k)\boldsymbol{\psi}(\mathbf{z}'_1, \dots, \mathbf{z}'_k)^T \mathbf{u}_{\infty,\infty} f_{\mathbf{z}',\infty}(\mathbf{z}'_1) \dots f_{\mathbf{z}',\infty}(\mathbf{z}'_k)\, d\mathbf{z}'_1 \dots d\mathbf{z}'_k = 0.$$

As a consequence $\mathbf{u}_{\infty,\infty} = -V_1^{-1}\tilde{\mathbf{W}} \sim \mathcal{N}(0, \tilde{V}_{as})$, using Assumption 16(iv). We finally obtain $\tilde{r}_{n,n'}\left(\hat{\beta}_{n,n'} - \beta^*\right) \xrightarrow{D} \mathcal{N}\left(0, \tilde{V}_{as}\right)$, as claimed. ■

E.1 Proof of Lemma 10

Using a Taylor expansion yields

$$\begin{aligned}
T_1 &:= \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma\in\mathfrak{I}_{k,n'}} \xi_{\sigma,n} \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right) \\
&= \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma\in\mathfrak{I}_{k,n'}} \left(\Lambda\left(\hat{\theta}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)\right) - \Lambda\left(\theta\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)\right)\right) \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right) \\
&= T_2 + T_3,
\end{aligned}$$

where the main term is

$$\begin{aligned}
T_2 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma\in\mathfrak{I}_{k,n'}} \Lambda'\left(\theta\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)\right)\Big(\hat{\theta}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right) \\
- \theta\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)\Big) \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right),
\end{aligned}$$

and the remainder is

$$T_3 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma\in\mathfrak{I}_{k,n'}} \alpha_{3,\sigma} \cdot \left(\hat{\theta}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right) - \theta\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)\right)^2 \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right),$$

with $\forall \sigma \in \mathfrak{I}_{k,n'}$, $|\alpha_{3,\sigma}| \leq C_{\Lambda''}/2$, by Assumption 16(v).

Let us define $\overline{\boldsymbol{\psi}}_\sigma := \Lambda'\left(\theta\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)\right)\boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)$, for every $\sigma \in \mathfrak{I}_{k,n'}$. Using the definition (2), we rewrite $T_2 := T_4 + T_5$ where

$$\begin{aligned}
T_4 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma\in\mathfrak{I}_{k,n'}} \sum_{\varsigma\in\mathfrak{I}_{k,n}} \frac{\prod_{i=1}^k K_h\left(\mathbf{Z}_{\varsigma(i)} - \mathbf{z}'_{\sigma(i)}\right)}{\prod_{i=1}^k f_{\mathbf{Z}}(\mathbf{z}'_{\sigma(i)})} \\
\left(g\left(\mathbf{X}_{\varsigma(1)}, \dots, \mathbf{X}_{\varsigma(k)}\right) - \theta\left(\mathbf{z}'_{\sigma(1)}, \dots, \mathbf{z}'_{\sigma(k)}\right)\right)\overline{\boldsymbol{\psi}}_\sigma,
\end{aligned}$$

$$T_5 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} \prod_{i=1}^{k} K_h \left(\mathbf{Z}_{\varsigma(i)} - \mathbf{z}'_{\sigma(i)} \right) \left(g\left(\mathbf{X}_{\varsigma(1)}, \ldots, \mathbf{X}_{\varsigma(k)} \right) - \theta\left(\mathbf{z}'_{\sigma(1)}, \ldots, \mathbf{z}'_{\sigma(k)} \right) \right)$$
$$\times \left(\frac{1}{N_k(\mathbf{z}'_{\sigma(1)}, \ldots, \mathbf{z}'_{\sigma(k)})} - \frac{1}{\prod_{i=1}^{k} f_{\mathbf{Z}}(\mathbf{z}'_{\sigma(i)})} \right) \overline{\boldsymbol{\psi}}_{\sigma}.$$

To lighten the notations, we will define $K_{\sigma,\varsigma} := \prod_{i=1}^{k} K_h \left(\mathbf{Z}_{\varsigma(i)} - \mathbf{z}'_{\sigma(i)} \right)$, $g_{\varsigma} := g\left(\mathbf{X}_{\varsigma(1)}, \ldots, \mathbf{X}_{\varsigma(k)} \right)$, $\theta_{\sigma} := \theta\left(\mathbf{z}'_{\sigma(1)}, \ldots, \mathbf{z}'_{\sigma(k)} \right)$, $f_{\mathbf{Z}',\sigma} := \prod_{i=1}^{k} f_{\mathbf{Z}}(\mathbf{z}'_{\sigma(i)})$, and $N_{\sigma} := N_k(\mathbf{z}'_{\sigma(1)}, \ldots, \mathbf{z}'_{\sigma(k)})$, for every $\sigma \in \mathfrak{I}_{k,n'}$ and $\varsigma \in \mathfrak{I}_{k,n}$, so that

$$T_4 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} \frac{K_{\sigma,\varsigma}}{f_{\mathbf{Z}',\sigma}} \left(g_{\varsigma} - \theta_{\sigma} \right) \overline{\boldsymbol{\psi}}_{\sigma}, \tag{E4}$$

$$T_5 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} K_{\sigma,\varsigma} \left(g_{\varsigma} - \theta_{\sigma} \right) \left(\frac{1}{N_{\sigma}} - \frac{1}{f_{\mathbf{Z}',\sigma}} \right) \overline{\boldsymbol{\psi}}_{\sigma}. \tag{E5}$$

Using α-order limited expansions, we get

$$\begin{aligned}
\mathbb{E}[T_4] &= \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \int \frac{\prod_{i=1}^{k} K_h \left(\mathbf{z}_i - \mathbf{z}'_{\sigma(i)} \right)}{f_{\mathbf{Z}',\sigma}} \left(g\left(\mathbf{x}_{1:k} \right) - \theta_{\sigma} \right) \prod_{i=1}^{k} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) d\mu^{\otimes k}(\mathbf{x}_{1:k}) d\mathbf{z}_{1:k} \\
&= \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \int \frac{\prod_{i=1}^{k} K\left(\mathbf{t}_i \right)}{f_{\mathbf{Z}',\sigma}} \left(g\left(\mathbf{x}_{1:k} \right) - \theta_{\sigma} \right) \prod_{i=1}^{k} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}'_{\sigma(i)} + h\mathbf{t}_i) d\mu^{\otimes k}(\mathbf{x}_{1:k}) d\mathbf{t}_{1:k} \\
&= \frac{\tilde{r}_{n,n'} h^{k\alpha}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \int \frac{\prod_{i=1}^{k} K\left(\mathbf{t}_i \right)}{f_{\mathbf{Z}',\sigma}} \left(g\left(\mathbf{x}_{1:k} \right) - \theta_{\sigma} \right) \prod_{i=1}^{k} d_{\mathbf{Z}}^{(\alpha)} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}^*_{\sigma(i)}) d\mu^{\otimes k}(\mathbf{x}_{1:k}) d\mathbf{t}_{1:k} \\
&= O\left(\tilde{r}_{n,n'} h^{k\alpha} \right) = O\left((n \times n' \times h_{n,n'}^{p+2k\alpha})^{1/2} \right) = o(1),
\end{aligned} \tag{E6}$$

where above, $\mathbf{z}^*_i$ denote some vectors in $\mathbb{R}^p$ such that $|\mathbf{z}'_i - \mathbf{z}^*_i|_\infty \leq 1$, depending on $\mathbf{z}'_i$ and $\mathbf{x}_i$. We can therefore use the centered version of T_4, defined as

$$T_4 - \mathbb{E}[T_4] = \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} g_{\sigma,\varsigma},$$
$$g_{\sigma,\varsigma} := \frac{\overline{\boldsymbol{\psi}}_{\sigma}}{f_{\mathbf{Z}',\sigma}} \left(K_{\sigma,\varsigma} \left(g_{\varsigma} - \theta_{\sigma} \right) - \mathbb{E}\left[K_{\sigma,\varsigma} \left(g_{\varsigma} - \theta_{\sigma} \right) \right] \right).$$

Computation of the limit of the variance matrix $\mathrm{Var}[T_4]$.

We have $\mathrm{Var}[T_4] = \mathbb{E}[T_4 T_4^T] + o(1)$.

$$\mathrm{Var}[T_4] = \frac{\tilde{r}^2_{n,n'}}{|\mathfrak{I}_{k,n'}|^2 \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma,\overline{\sigma} \in \mathfrak{I}_{k,n'}} \sum_{\varsigma,\overline{\varsigma} \in \mathfrak{I}_{k,n}} \mathbb{E}[g_{\sigma,\varsigma} g^T_{\overline{\sigma},\overline{\varsigma}}] + o(1).$$

By independence, $\mathbb{E}[g_{\sigma,\varsigma} g^T_{\overline{\sigma},\overline{\varsigma}}] = 0$ as soon as $\varsigma \cap \overline{\varsigma} = \emptyset$, where we identify a permutation ς and its image $\varsigma(\{1, \dots, k\})$. Therefore, we get

$$\begin{aligned}
\mathrm{Var}[T_4] &\simeq \frac{nn' h^p_{n,n'}}{|\mathfrak{I}_{k,n'}|^2 \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma, \overline{\sigma} \in \mathfrak{I}_{k,n'}} \sum_{\substack{\varsigma, \overline{\varsigma} \in \mathfrak{I}_{k,n} \\ \varsigma \cap \overline{\varsigma} \neq \emptyset}} \mathbb{E}\left[g_{\sigma,\varsigma}\, g^T_{\overline{\sigma},\overline{\varsigma}}\right] \\
&= \frac{nn' h^p_{n,n'}}{|\mathfrak{I}_{k,n'}|^2 \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma, \overline{\sigma} \in \mathfrak{I}_{k,n'}} \sum_{\substack{\varsigma, \overline{\varsigma} \in \mathfrak{I}_{k,n} \\ \varsigma \cap \overline{\varsigma} \neq \emptyset}} g_{\sigma,\varsigma,\overline{\sigma},\overline{\varsigma}} - \tilde{g}_\sigma \tilde{g}^T_{\overline{\sigma}},
\end{aligned}$$

where $\tilde{g}_\sigma := \overline{\boldsymbol{\psi}}_\sigma \mathbb{E}\left[K_{\sigma,\varsigma}\left(g_\varsigma - \theta_\sigma\right)\right] / f_{\mathbf{Z}',\sigma}$ and

$$g_{\sigma,\varsigma,\overline{\sigma},\overline{\varsigma}} := \frac{\overline{\boldsymbol{\psi}}_\sigma \overline{\boldsymbol{\psi}}^T_{\overline{\sigma}}}{f_{\mathbf{Z}',\sigma} f_{\mathbf{Z}',\overline{\sigma}}} \mathbb{E}\left[K_{\sigma,\varsigma} K_{\overline{\sigma},\overline{\varsigma}} \left(g_\varsigma - \theta_\sigma\right)\left(g_{\overline{\varsigma}} - \theta_{\overline{\sigma}}\right)\right].$$

Assume now that $\varsigma \cap \overline{\varsigma}$ is of cardinality 1, that is, there exists only one couple $(j, \overline{j}) \in \{1, \dots, k\}^2$ such that $\varsigma(j) = \overline{\varsigma}(\overline{j})$. Then,

$$\begin{aligned}
g_{\sigma,\varsigma,\overline{\sigma},\overline{\varsigma}} &= \frac{\overline{\boldsymbol{\psi}}_\sigma \overline{\boldsymbol{\psi}}^T_{\overline{\sigma}}}{f_{\mathbf{Z}',\sigma} f_{\mathbf{Z}',\overline{\sigma}}} \int (g(\mathbf{X}_{1:k}) - \theta_\sigma)\left(g(\mathbf{x}_{k+1}, \dots, \mathbf{x}_{k+\overline{j}-1}, \mathbf{x}_j, \mathbf{x}_{k+\overline{j}+1}, \dots, \mathbf{x}_{2k}) - \theta_{\overline{\sigma}}\right) \\
&\cdot \prod_{i=1}^k K_h\left(\mathbf{z}_i - \mathbf{z}'_{\sigma(i)}\right) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) d\mu(\mathbf{x}_i) d\mathbf{z}_i \cdot K_h\left(\mathbf{z}_j - \mathbf{z}'_{\overline{\sigma}(\overline{j})}\right) \\
&\cdot \prod_{\overline{i}=1,\ \overline{i} \neq \overline{j}}^k K_h\left(\mathbf{z}_{k+i} - \mathbf{z}'_{\overline{\sigma}(\overline{i})}\right) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{k+i}, \mathbf{z}_{k+i}) d\mu(\mathbf{x}_{k+i}) d\mathbf{z}_{k+i} \\
&= \frac{\overline{\boldsymbol{\psi}}_\sigma \overline{\boldsymbol{\psi}}^T_{\overline{\sigma}}}{f_{\mathbf{Z}}(\mathbf{z}_j)} \int (g(\mathbf{X}_{1:k}) - \theta_\sigma)\left(g(\mathbf{x}_{k+1}, \dots, \mathbf{x}_{k+\overline{j}-1}, \mathbf{x}_j, \mathbf{x}_{k+\overline{j}+1}, \dots, \mathbf{x}_{2k}) - \theta_{\overline{\sigma}}\right) \\
&\cdot \prod_{i=1}^k K(\mathbf{t}_i) \frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}'_{\sigma(i)} + h\mathbf{t}_i)}{f_{\mathbf{Z}}(\mathbf{z}'_{\sigma(i)})} d\mu(\mathbf{x}_i) d\mathbf{t}_i \cdot h^{-p} K\left(\mathbf{t}_i + \frac{\mathbf{z}'_{\sigma(j)} - \mathbf{z}'_{\overline{\sigma}(\overline{j})}}{h}\right) \\
&\cdot \prod_{\overline{i}=1,\ \overline{i} \neq \overline{j}}^k K(\mathbf{t}_{k+i}) \frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{k+i}, \mathbf{z}'_{\overline{\sigma}(\overline{i})} + h\mathbf{t}_{k+i})}{f_{\mathbf{Z}}(\mathbf{z}_{k+i})} d\mu(\mathbf{x}_{k+i}) d\mathbf{t}_{k+i} \\
&\simeq \frac{\overline{\boldsymbol{\psi}}_\sigma \overline{\boldsymbol{\psi}}^T_{\overline{\sigma}}}{f_{\mathbf{Z}}(\mathbf{z}_j)} \int (g(\mathbf{X}_{1:k}) - \theta_\sigma)\left(g(\mathbf{x}_{k+1}, \dots, \mathbf{x}_{k+\overline{j}-1}, \mathbf{x}_j, \mathbf{x}_{k+\overline{j}+1}, \dots, \mathbf{x}_{2k}) - \theta_{\overline{\sigma}}\right) \\
&\cdot \prod_{i=1}^k K(\mathbf{t}_i) \frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}'_{\sigma(i)})}{f_{\mathbf{Z}}(\mathbf{z}_i)} d\mu(\mathbf{x}_i) d\mathbf{t}_i \cdot h^{-p} K\left(\mathbf{t}_i + \frac{\mathbf{z}'_{\sigma(j)} - \mathbf{z}'_{\overline{\sigma}(\overline{j})}}{h}\right) \\
&\cdot \prod_{\overline{i}=1,\ \overline{i} \neq \overline{j}}^k K(\mathbf{t}_{k+i}) \frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{k+i}, \mathbf{z}'_{\overline{\sigma}(\overline{i})})}{f_{\mathbf{Z}}(\mathbf{z}'_{\overline{\sigma}(\overline{i})})} d\mu(\mathbf{x}_{k+i}) d\mathbf{t}_{k+i}.
\end{aligned}$$

By assumption, this is zero unless $\sigma(j) = \overline{\sigma}(\bar{j})$. In this case, it can be simplified, giving

$$
g_{\sigma,\varsigma,\overline{\sigma},\overline{\varsigma}} \simeq \frac{\overline{\boldsymbol{\psi}}_{\sigma}\overline{\boldsymbol{\psi}}^{T}_{\overline{\sigma}}}{f_{\mathbf{Z}}(\mathbf{z}_j)h^p} \int K^2 \int \left(g(\mathbf{x}_{1:k}) - \theta_{\sigma}\right)\left(g(\mathbf{x}_{k:2k,\bar{j}\to j}) - \theta_{\overline{\sigma}}\right)
\cdot \prod_{i=1}^{k} f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}'_{\sigma(i)}}(\mathbf{x}_k)d\mu(\mathbf{x}_i) \prod_{\bar{i}=1,\ \bar{i}\neq\bar{j}}^{k} f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}'_{\overline{\sigma}(\bar{i})}}(\mathbf{x}_{k+i})d\mu(\mathbf{x}_{k+i}) \;=:\; h^{-p}g_{\sigma,\overline{\sigma},j,\bar{j}},
$$

where $\mathbf{x}_{k:2k,\bar{j}\to j} := (\mathbf{x}_{k+1}, \dots, \mathbf{x}_{k+\bar{j}-1}, \mathbf{x}_j, \mathbf{x}_{k+\bar{j}+1}, \dots, \mathbf{x}_{2k})$.

Note that, if $\varsigma \cap \overline{\varsigma}$ is of cardinality strictly greater than 1, some supplementary powers of h^{-p} arise thanks to the repeated kernels in ς and $\overline{\varsigma}$. As a consequence, they are of lower order and therefore negligible. Using α-order expansions as in Equation (E6), we get $\sup_{\sigma} |\tilde{g}_{\sigma}| = O(h^{k\alpha})$. Thus,

$$
\begin{aligned}
\mathrm{Var}[T_4] &\simeq O\left(nn'h_{n,n'}^{p+2k\alpha}\right) + \frac{nn'h_{n,n'}^{p}}{|\mathfrak{I}_{k,n'}|^2 \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\varsigma\in\mathfrak{I}_{k,n}} \sum_{j,\bar{j}=1}^{k} \sum_{\substack{\overline{\varsigma}\in\mathfrak{I}_{k,n} \\ \varsigma(j)=\overline{\varsigma}(\bar{j}),\ |\varsigma\cap\overline{\varsigma}|=1}} \sum_{\sigma,\overline{\sigma}\in\mathfrak{I}_{k,n'},\ \sigma(j)=\overline{\sigma}(\bar{j})} h^{-p}g_{\sigma,\overline{\sigma},j,\bar{j}} \\
&\simeq \frac{n'}{|\mathfrak{I}_{k,n'}|^2} \sum_{j,\bar{j}=1}^{k} \sum_{\sigma,\overline{\sigma}\in\mathfrak{I}_{k,n'},\ \sigma(j)=\overline{\sigma}(\bar{j})} g_{\sigma,\overline{\sigma},j,\bar{j}} \\
&\to \sum_{j,\bar{j}=1}^{k} g_{j,\bar{j},\infty} = V_2,
\end{aligned}
$$

where

$$
g_{j,\bar{j},\infty} := \int \Lambda'\left(\theta\left(\mathbf{z}'_{1:k}\right)\right)\Lambda'\left(\theta\left(\mathbf{z}'_{k:2k,\bar{j}\to j}\right)\right)\boldsymbol{\psi}\left(\mathbf{z}'_{1:k}\right)\boldsymbol{\psi}^{T}\left(\mathbf{z}'_{k:2k,\bar{j}\to j}\right)\frac{\int K^2}{f_{\mathbf{Z}}(\mathbf{z}'_j)} \int \left(g(\mathbf{x}_{1:k}) - \theta(\mathbf{z}'_{1:k})\right)
\cdot \left(g(\mathbf{x}_{k:2k,\bar{j}\to j}) - \theta(\mathbf{z}'_{k:2k,\bar{j}\to j})\right) \prod_{i=1,i\neq k+\bar{j}}^{2k} f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}'_i}(\mathbf{x}_i)f_{\mathbf{Z}',\infty}(\mathbf{z}'_i)d\mu(\mathbf{x}_i)d\mathbf{z}'_i.
$$

In Section E.2, we will prove that T_4 is asymptotically Gaussian; therefore, its asymptotic variance will be given by V_2.

Now, decompose the term T_5, defined in Equation (E5), using a Taylor expansion of the function $x \mapsto 1/(1+x)$ at 0.

$$
\frac{1}{N_{\sigma}} - \frac{1}{f_{\mathbf{Z}',\sigma}} = \frac{1}{f_{\mathbf{Z}',\sigma}}\left(\frac{1}{1+\frac{N_{\sigma}-f_{\mathbf{Z}',\sigma}}{f_{\mathbf{Z}',\sigma}}} - 1\right) = -\frac{N_{\sigma}-f_{\mathbf{Z}',\sigma}}{f_{\mathbf{Z}',\sigma}^2} + T_{7,\sigma},
$$

where

$$
T_{7,\sigma} := \frac{1}{f_{\mathbf{Z}',\sigma}}(1+\alpha_{7,\sigma})^{-3}\left(\frac{N_{\sigma}-f_{\mathbf{Z}',\sigma}}{f_{\mathbf{Z}',\sigma}}\right)^2, \text{ with } |\alpha_{7,\sigma}| \leq \left|\frac{N_{\sigma}-f_{\mathbf{Z}',\sigma}}{f_{\mathbf{Z}',\sigma}}\right|.
$$

We have therefore the decomposition $T_5 = -T_6 + T_7$, where

$$T_6 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} K_{\sigma,\varsigma} \left(g_\varsigma - \theta_\sigma\right) \frac{N_\sigma - f_{\mathbf{Z}',\sigma}}{f^2_{\mathbf{Z}',\sigma}} \overline{\boldsymbol{\psi}}_\sigma, \tag{E7}$$

$$T_7 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} K_{\sigma,\varsigma} \left(g_\varsigma - \theta_\sigma\right) T_{7,\sigma} \overline{\boldsymbol{\psi}}_\sigma. \tag{E8}$$

Summing up all the previous equation, we get

$$T_1 = (T_4 - \mathbb{E}[T_4]) - T_6 + T_7 + T_3 + o(1).$$

Afterwards, we will prove that all the remainders terms T_6, T_7, and T_3 are negligible, that is, they tend to zero in probability. These results are respectively proved in Sections E.3, E.4, and E.5. Combining all these elements with the asymptotic normality of T_4 (proved in Section E.2), we get $T_1 \xrightarrow{D} \mathcal{N}(0, V_2)$, as claimed. ■

E.2 Proof of the asymptotic normality of T_4

Using the Hájek projection of T_4, we define

$$\begin{aligned}
&T_4 - \mathbb{E}[T_4] = T_{4,1} + T_{4,2}, \text{ where} \\
&T_{4,1} := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} \sum_{i=1}^{k} \mathbb{E}[g_{\sigma,\varsigma} | \varsigma(i)], \\
&T_{4,2} := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} \left(g_{\sigma,\varsigma} - \sum_{i=1,\dots,k} \mathbb{E}[g_{\sigma,\varsigma} | \varsigma(i)] \right),
\end{aligned}$$

denoting by $|i$ the conditioning with respect to $(\mathbf{X}_i, \mathbf{Z}_i)$, for $i \in \{1, \dots, n\}$. We will show that $T_{4,1}$ is asymptotically normal, and that $T_{4,2} = o(1)$.

Using the fact that the $(\mathbf{X}_i, \mathbf{Z}_i)_i$ are i.i.d., and denoting by Id the injective function $i \mapsto i$, we have

$$\begin{aligned}
T_{4,1} &= \frac{k\tilde{r}_{n,n'}}{n|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{i=1}^{n} \mathbb{E}\left[\frac{\overline{\boldsymbol{\psi}}_\sigma}{f_{\mathbf{Z}',\sigma}} K_{\sigma,Id} \left(g_{Id} - \theta_\sigma\right) - \overline{g}_\sigma \middle| i \right] \\
&\simeq \frac{k\tilde{r}_{n,n'}}{n|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{i=1}^{n} \mathbb{E}\left[\frac{\overline{\boldsymbol{\psi}}_\sigma}{f_{\mathbf{Z}',\sigma}} K_{\sigma,Id} \left(g_{Id} - \theta_\sigma\right) \middle| i \right] =: \sum_{i=1}^{n} \alpha_{4,i,n},
\end{aligned}$$

because $\sup_\sigma |\overline{g}_\sigma| = O(h^{k\alpha})$, as proved in the previous section, hence negligible. The $\alpha_{4,i,n}$, for $1 \leq i \leq n$, form a triangular array of i.i.d. variables. To prove the asymptotic normality of $T_{4,1}$, it remains to check Lyapunov's condition, that is, we will show that $\sum_{i=1}^{n} \mathbb{E}\left[|\alpha_{4,i,n}|_\infty^3\right] \to 0$. We have

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\left[|\alpha_{4,i,n}|_\infty^3\right] = n\, \mathbb{E}\left[|\alpha_{4,1,n}|_\infty^3\right] \\
&\quad = \frac{k^3 n \tilde{r}^3_{n,n'}}{n^3 |\mathfrak{I}_{k,n'}|^3} \sum_{\sigma,\nu,\vartheta \in \mathfrak{I}_{k,n'}} \frac{\overline{\boldsymbol{\psi}}_\sigma \otimes \overline{\boldsymbol{\psi}}_\nu \otimes \overline{\boldsymbol{\psi}}_\vartheta}{f_{\mathbf{Z}',\sigma} f_{\mathbf{Z}',\nu} f_{\mathbf{Z}',\vartheta}}
\end{aligned}$$

$$
\begin{aligned}
&\mathbb{E}\big[\mathbb{E}\big[K_{\sigma,Id}\big(g_{Id}-\theta_\sigma\big)|1\big]\mathbb{E}\big[K_{\nu,Id}\big(g_{Id}-\theta_\nu\big)\big)|1\big]\mathbb{E}\big[K_{\vartheta,Id}\big(g_{Id}-\theta_\vartheta\big)\big)|1\big]\big] \\
&= \frac{k^3\tilde{r}^3_{n,n'}}{n^2|\mathfrak{I}_{k,n'}|^3}\sum_{\sigma,\nu,\vartheta\in\mathfrak{I}_{k,n'}}\frac{\overline{\boldsymbol{\psi}}_\sigma\otimes\overline{\boldsymbol{\psi}}_\nu\otimes\overline{\boldsymbol{\psi}}_\vartheta}{f_{\mathbf{Z}}(\mathbf{z}'_{\nu(1)})f_{\mathbf{Z}}(\mathbf{z}'_{\vartheta(1)})}\int K_h\big(\mathbf{z}_1-\mathbf{z}'_{\sigma(1)}\big)K_h\big(\mathbf{z}_1-\mathbf{z}'_{\nu(1)}\big)K_h\big(\mathbf{z}_1-\mathbf{z}'_{\vartheta(1)}\big) \\
&\cdot\prod_{i=2}^{k}K_h\big(\mathbf{z}_i-\mathbf{z}'_{\sigma(i)}\big)K_h\big(\mathbf{z}_{k+i}-\mathbf{z}'_{\nu(i)}\big)K_h\big(\mathbf{z}_{2k+i}-\mathbf{z}'_{\vartheta(i)}\big) \\
&\cdot\big(g(\mathbf{x}_{1:k})-\theta_\sigma)\big)\big(g(\mathbf{x}_1,\mathbf{x}_{(k+2):(2k)})-\theta_\nu\big)\big(g(\mathbf{x}_1,\mathbf{x}_{(2k+2):(3k)})-\theta_\vartheta\big) \\
&\cdot\prod_{i=1}^{k}\frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i,\mathbf{z}_i)}{f_{\mathbf{Z}}(\mathbf{z}'_{\sigma(i)})}d\mu(\mathbf{x}_i)d\mathbf{z}_i\prod_{i=2}^{k}\frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{k+i},\mathbf{z}_{k+i})}{f_{\mathbf{Z}}(\mathbf{z}'_{\nu(i)})}d\mu(\mathbf{x}_{k+i})d\mathbf{z}_{k+i}\prod_{i=2}^{k}\frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{2k+i},\mathbf{z}_{2k+i})}{f_{\mathbf{Z}}(\mathbf{z}'_{\vartheta(i)})}d\mu(\mathbf{x}_{2k+i})d\mathbf{z}_{2k+i} \\
&\simeq \frac{k^3\tilde{r}^3_{n,n'}}{n^2|\mathfrak{I}_{k,n'}|^3}\sum_{\sigma,\nu,\vartheta\in\mathfrak{I}_{k,n'}}\frac{\overline{\boldsymbol{\psi}}_\sigma\otimes\overline{\boldsymbol{\psi}}_\nu\otimes\overline{\boldsymbol{\psi}}_\vartheta}{f_{\mathbf{Z}}(\mathbf{z}'_{\nu(1)})f_{\mathbf{Z}}(\mathbf{z}'_{\vartheta(1)})} \\
&\int h^{-2p}K(\mathbf{t}_1)K\left(\mathbf{t}_1+\frac{\mathbf{z}'_{\sigma(1)}-\mathbf{z}'_{\nu(1)}}{h}\right)K\left(\mathbf{t}_1+\frac{\mathbf{z}'_{\sigma(1)}-\mathbf{z}'_{\vartheta(1)}}{h}\right) \\
&\cdot\prod_{i=2}^{k}K_h\big(\mathbf{t}_i\big)K_h\big(\mathbf{t}_{k+i}\big)K_h\big(\mathbf{t}_{2k+i}\big)\big(g(\mathbf{x}_{1:k})-\theta_\sigma)\big)\big(g(\mathbf{x}_1,\mathbf{x}_{(k+2):(2k)})-\theta_\nu\big)\big(g(\mathbf{x}_1,\mathbf{x}_{(2k+2):(3k)})-\theta_\vartheta\big) \\
&\cdot\prod_{i=1}^{k}f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}'_{\sigma(i)}}(\mathbf{x}_i)d\mu(\mathbf{x}_i)d\mathbf{z}_i\prod_{i=2}^{k}f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}'_{\nu(i)}}(\mathbf{x}_{k+i})d\mu(\mathbf{x}_{k+i})d\mathbf{z}_{k+i} \\
&\prod_{i=2}^{k}f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}'_{\vartheta(i)}}(\mathbf{x}_{2k+i},\mathbf{t}_{2k+i})d\mu(\mathbf{x}_{2k+i})d\mathbf{t}_{2k+i},
\end{aligned}
$$

where in the last equivalent, we use a change of variable from the $\mathbf{z}_i$ to the $\mathbf{t}_i$, and then the continuity of the density $f_{\mathbf{X},\mathbf{Z}}$ with respect to $\mathbf{z}$, because $h=o(1)$.

Because of our assumptions, the terms of the sum for which $\sigma(1)\neq 1$ or $\nu(1)\neq 1$ are zero. Therefore, we get

$$
\sum_{i=1}^{n}\mathbb{E}\left[|\alpha_{4,i,n}|^3_\infty\right]=\frac{\tilde{r}^3_{n,n'}h^{-2p}}{n^2|\mathfrak{I}_{k,n'}|^3}\sum_{\sigma,\nu,\vartheta\in\mathfrak{I}_{k,n'},\sigma(1)=\nu(1)=1}O(1)=O\left(\frac{(nn'h^p)^{3/2}}{n^2n'^2h^{2p}}\right)=O\left(\frac{1}{(nn'h^p)^{1/2}}\right)=o(1).
$$

We prove now that $T_{4,2}=o(1)$. Note first that, by construction, $\mathbb{E}[T_{4,2}]=0$. Computing its variance, we get

$$
\begin{aligned}
&\mathbb{E}\left[T_{4,2}T_{4,2}^T\right] \\
&=\mathbb{E}\left[\frac{\tilde{r}^2_{n,n'}}{|\mathfrak{I}_{k,n'}|^2\cdot|\mathfrak{I}_{k,n}|^2}\sum_{\sigma,\overline{\sigma}\in\mathfrak{I}_{k,n'}}\sum_{\varsigma,\overline{\varsigma}\in\mathfrak{I}_{k,n}}\left(g_{\sigma,\varsigma}-\sum_{i=1,\dots,k}\mathbb{E}\left[g_{\sigma,\varsigma}\middle|\varsigma(i)\right]\right)\left(g_{\overline{\sigma},\overline{\varsigma}}-\sum_{\overline{i}=1,\dots,k}\mathbb{E}\left[g_{\overline{\sigma},\overline{\varsigma}}\middle|\overline{\varsigma}(\overline{i})\right]\right)^T\right] \\
&=:\frac{\tilde{r}^2_{n,n'}}{|\mathfrak{I}_{k,n'}|^2\cdot|\mathfrak{I}_{k,n}|^2}\sum_{\sigma,\overline{\sigma}\in\mathfrak{I}_{k,n'}}\sum_{\varsigma,\overline{\varsigma}\in\mathfrak{I}_{k,n}}\mathbb{E}\left[\tilde{g}_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma}}\right].
\end{aligned}
\tag{E9}
$$

Because of $\mathbb{E}[g_{\sigma,\varsigma}]=0$ and by independence, the terms in the latter sum for which $\varsigma\cap\overline{\varsigma}=\emptyset$ are zero. Otherwise, there exists $j_1,j_2\in\{1,\dots,k\}$ such that $\varsigma(j_1)=\overline{\varsigma}(j_2)$. If $\varsigma\cap\overline{\varsigma}$ is of cardinal 1, meaning that there is no other identities between elements of ς and $\bar{\varsigma}$, then we will show that the corresponding term is zero as well. We place ourselves in this case, assuming that $|\varsigma \cap \bar{\varsigma}| = 1$, and we get

$$\begin{aligned}
\mathbb{E}\left[\tilde{g}_{\sigma,\bar{\sigma},\varsigma,\bar{\varsigma}}\right] &= \mathbb{E}\left[\left(g_{\sigma,\varsigma} - \sum_{i=1,\dots,k} \mathbb{E}\left[g_{\sigma,\varsigma}\middle|\varsigma(i)\right]\right)\left(g^T_{\bar{\sigma},\bar{\varsigma}} - \sum_{\bar{i}=1,\dots,k} \mathbb{E}\left[g^T_{\bar{\sigma},\bar{\varsigma}}\middle|\bar{\varsigma}(\bar{i})\right]\right)\right] \\
&= \mathbb{E}\left[\left(g_{\sigma,\varsigma} - \mathbb{E}\left[g_{\sigma,\varsigma}\middle|\varsigma(j_1)\right]\right)\left(g^T_{\bar{\sigma},\bar{\varsigma}} - \mathbb{E}\left[g^T_{\bar{\sigma},\bar{\varsigma}}\middle|\bar{\varsigma}(j_2)\right]\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\left(g_{\sigma,\varsigma} - \mathbb{E}\left[g_{\sigma,\varsigma}\middle|\varsigma(j_1)\right]\right)\left(g^T_{\bar{\sigma},\bar{\varsigma}} - \mathbb{E}\left[g^T_{\bar{\sigma},\bar{\varsigma}}\middle|\varsigma(j_1)\right]\right)\middle|\varsigma(j_1)\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[g_{\sigma,\varsigma}g^T_{\bar{\sigma},\bar{\varsigma}}\middle|\varsigma(j_1)\right]\right] - \mathbb{E}\left[\mathbb{E}\left[g_{\sigma,\varsigma}\middle|\varsigma(j_1)\right]\mathbb{E}\left[g^T_{\bar{\sigma},\bar{\varsigma}}\middle|\varsigma(j_1)\right]\right] = 0.
\end{aligned}$$

Therefore, nonzero terms in Equation (E9) correspond to the case where there exists $j_3 \neq j_1, j_4 \neq j_1$ such that $\varsigma(j_3) = \bar{\varsigma}(j_4)$. It is equivalent to $|\varsigma \cap \bar{\varsigma}| \geq 2$. We will ignore higher-order terms, that is, the ones for which $|\varsigma \cap \bar{\varsigma}| > 2$, as they yield higher powers of h^p and are therefore negligible. Finally, Equation (E9) becomes

$$\mathbb{E}\left[T_{4,2}T_{4,2}^T\right] \simeq \frac{\tilde{r}^2_{n,n'}}{|\mathfrak{I}_{k,n'}|^2 \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma,\bar{\sigma}\in\mathfrak{I}_{k,n'}} \sum_{\substack{\varsigma,\bar{\varsigma}\in\mathfrak{I}_{k,n} \\ |\varsigma\cap\bar{\varsigma}|=2}} \left(\mathbb{E}\left[g_{\sigma,\varsigma}g^T_{\bar{\sigma},\bar{\varsigma}}\right] - 2k\mathbb{E}\left[\mathbb{E}\left[g_{\sigma,\varsigma}\middle|\varsigma(i)\right]\mathbb{E}\left[g^T_{\bar{\sigma},\bar{\varsigma}}\middle|\bar{\varsigma}(\bar{i})\right]\right]\right).$$

As before, using change of variables and limited expansions, we can prove that

$$\frac{\tilde{r}^2_{n,n'}}{|\mathfrak{I}_{k,n'}|^2 \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma,\bar{\sigma}\in\mathfrak{I}_{k,n'}} \sum_{\substack{\varsigma,\bar{\varsigma}\in\mathfrak{I}_{k,n} \\ |\varsigma\cap\bar{\varsigma}|=2}} \mathbb{E}\left[g_{\sigma,\varsigma}g^T_{\bar{\sigma},\bar{\varsigma}}\right] = o(1),$$

and similarly for the other term.

E.3 Convergence of T_6 to 0

Using Equation (E7), we have $T_6 = T_{6,1} + T_{6,2}$, where

$$T_{6,1} := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma\in\mathfrak{I}_{k,n'}} \sum_{\varsigma\in\mathfrak{I}_{k,n}} K_{\sigma,\varsigma}\left(g_\varsigma - \theta_\sigma\right) \frac{N_\sigma - \mathbb{E}[N_\sigma]}{f^2_{\mathbf{Z}',\sigma}} \overline{\boldsymbol{\psi}}_\sigma, \tag{E10}$$

$$T_{6,2} := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma\in\mathfrak{I}_{k,n'}} \sum_{\varsigma\in\mathfrak{I}_{k,n}} K_{\sigma,\varsigma}\left(g_\varsigma - \theta_\sigma\right) \frac{\mathbb{E}[N_\sigma] - f_{\mathbf{Z}',\sigma}}{f^2_{\mathbf{Z}',\sigma}} \overline{\boldsymbol{\psi}}_\sigma. \tag{E11}$$

We first prove that $T_{6,1} = o(1)$. Using Equation (5), we have

$$\begin{aligned}
T_{6,1} &= \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma\in\mathfrak{I}_{k,n'}} \sum_{\varsigma\in\mathfrak{I}_{k,n}} \frac{1}{f^2_{\mathbf{Z}',\sigma}} K_{\sigma,\varsigma}\left(g_\varsigma - \theta_\sigma\right)\left(N_k(\mathbf{z}'_{\sigma(1:k)}) - \mathbb{E}[N_k(\mathbf{z}'_{\sigma(1:k)})]\right)\overline{\boldsymbol{\psi}}_\sigma \\
&= \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma\in\mathfrak{I}_{k,n'}} \sum_{\varsigma\in\mathfrak{I}_{k,n}} \frac{1}{f^2_{\mathbf{Z}',\sigma}} K_{\sigma,\varsigma}\left(g_\varsigma - \theta_\sigma\right)
\end{aligned}$$

$$\sum_{\nu\in\mathfrak{I}_{k,n}}\left(\prod_{i=1}^{k}K_h\left(\mathbf{Z}_{\nu(i)}-\mathbf{z}'_{\sigma(i)}\right)-\mathbb{E}\left[\prod_{i=1}^{k}K_h\left(\mathbf{Z}_{\nu(i)}-\mathbf{z}'_{\sigma(i)}\right)\right]\right)\overline{\boldsymbol{\psi}}_\sigma$$
$$=\frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|\cdot|\mathfrak{I}_{k,n}|}\sum_{\sigma\in\mathfrak{I}_{k,n'}}\sum_{\varsigma,\nu\in\mathfrak{I}_{k,n}}\frac{1}{f^2_{\mathbf{Z}',\sigma}}K_{\sigma,\varsigma}\left(g_\varsigma-\theta_\sigma\right)\left(K_{\sigma,\nu}-\mathbb{E}\left[K_{\sigma,\nu}\right]\right)\overline{\boldsymbol{\psi}}_\sigma.$$

The terms for which $|\varsigma\cap\nu|\geq 1$ induce some powers of $(nh^p)^{-1}$, and are therefore negligible. We remove them to obtain an equivalent random vector $\overline{T}_{6,1}$, which is centered. Therefore it is sufficient to show that its second moment tends to 0.

$$\mathbb{E}\left[\overline{T}_{6,1}\overline{T}_{6,1}^T\right]=\frac{\tilde{r}^2_{n,n'}}{|\mathfrak{I}_{k,n'}|^2\cdot|\mathfrak{I}_{k,n}|^2}\sum_{\sigma,\overline{\sigma}\in\mathfrak{I}_{k,n'}}\sum_{\substack{\varsigma,\nu\in\mathfrak{I}_{k,n}\\ \varsigma\cap\nu=\emptyset}}\sum_{\substack{\overline{\varsigma},\overline{\nu}\in\mathfrak{I}_{k,n}\\ \overline{\varsigma}\cap\overline{\nu}=\emptyset}}\frac{\overline{\boldsymbol{\psi}}_\sigma}{f^2_{\mathbf{Z}',\sigma}}\frac{\overline{\boldsymbol{\psi}}^T_{\overline{\sigma}}}{f^2_{\mathbf{Z}',\overline{\sigma}}}g_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma},\nu,\overline{\nu}},$$
$$g_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma},\nu,\overline{\nu}}:=\mathbb{E}\left[K_{\sigma,\varsigma}\left(g_\varsigma-\theta_\sigma\right)\left(K_{\sigma,\nu}-\mathbb{E}\left[K_{\sigma,\nu}\right]\right)K_{\overline{\sigma},\overline{\varsigma}}\left(g_{\overline{\varsigma}}-\theta_{\overline{\sigma}}\right)\left(K_{\overline{\sigma},\overline{\nu}}-\mathbb{E}\left[K_{\overline{\sigma},\overline{\nu}}\right]\right)\right].$$

The term $g_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma},\nu,\overline{\nu}}$ is 0 in two cases: if $\nu\cap(\varsigma\cup\overline{\varsigma}\cup\overline{\nu})$ or if $\overline{\nu}\cap(\varsigma\cup\overline{\varsigma}\cup\nu)$. This condition can be written as

$$\emptyset=\left[\nu\cap(\overline{\varsigma}\cup\overline{\nu})\right]\cup\left[\overline{\nu}\cap(\varsigma\cup\nu)\right]=(\nu\cup\overline{\nu})\cap(\overline{\varsigma}\cup\overline{\nu})\cap(\varsigma\cup\nu).$$

We deduce that nonzero terms arise only when there exists $j_1,j_2\in\{1,\dots,k\}$ such that: $\nu(j_1)=\overline{\nu}(j_2)$ or $\nu(j_1)=\overline{\varsigma}(j_2)$ or $\overline{\nu}(j_1)=\varsigma(j_2)$. Therefore, we can write $\mathbb{E}\left[\overline{T}_{6,1}\overline{T}_{6,1}^T\right]=T_{6,1,1}+T_{6,1,2}+T_{6,1,3}$, where

$$T_{6,1,1}=\frac{\tilde{r}^2_{n,n'}}{|\mathfrak{I}_{k,n'}|^2\cdot|\mathfrak{I}_{k,n}|^2}\sum_{j_1,j_2=1}^{k}\sum_{\sigma,\overline{\sigma}\in\mathfrak{I}_{k,n'}}\sum_{\substack{\varsigma,\nu\in\mathfrak{I}_{k,n}\\ \varsigma\cap\nu=\emptyset}}\sum_{\substack{\overline{\varsigma},\overline{\nu}\in\mathfrak{I}_{k,n}\\ \overline{\varsigma}\cap\overline{\nu}=\emptyset,\overline{\nu}(j_2)=\nu(j_1)}}\frac{\overline{\boldsymbol{\psi}}_\sigma}{f^2_{\mathbf{Z}',\sigma}}\frac{\overline{\boldsymbol{\psi}}^T_{\overline{\sigma}}}{f^2_{\mathbf{Z}',\overline{\sigma}}}g_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma},\nu,\overline{\nu}},$$
$$T_{6,1,2}=\frac{\tilde{r}^2_{n,n'}}{|\mathfrak{I}_{k,n'}|^2\cdot|\mathfrak{I}_{k,n}|^2}\sum_{j_1,j_2=1}^{k}\sum_{\sigma,\overline{\sigma}\in\mathfrak{I}_{k,n'}}\sum_{\substack{\varsigma,\nu\in\mathfrak{I}_{k,n}\\ \varsigma\cap\nu=\emptyset}}\sum_{\substack{\overline{\varsigma},\overline{\nu}\in\mathfrak{I}_{k,n}\\ \overline{\varsigma}\cap\overline{\nu}=\emptyset,\overline{\varsigma}(j_2)=\nu(j_1)}}\frac{\overline{\boldsymbol{\psi}}_\sigma}{f^2_{\mathbf{Z}',\sigma}}\frac{\overline{\boldsymbol{\psi}}^T_{\overline{\sigma}}}{f^2_{\mathbf{Z}',\overline{\sigma}}}g_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma},\nu,\overline{\nu}},$$
$$T_{6,1,3}=\frac{\tilde{r}^2_{n,n'}}{|\mathfrak{I}_{k,n'}|^2\cdot|\mathfrak{I}_{k,n}|^2}\sum_{j_1,j_2=1}^{k}\sum_{\sigma,\overline{\sigma}\in\mathfrak{I}_{k,n'}}\sum_{\substack{\varsigma,\nu\in\mathfrak{I}_{k,n}\\ \varsigma\cap\nu=\emptyset}}\sum_{\substack{\overline{\varsigma},\overline{\nu}\in\mathfrak{I}_{k,n}\\ \overline{\varsigma}\cap\overline{\nu}=\emptyset,\overline{\nu}(j_1)=\varsigma(j_2)}}\frac{\overline{\boldsymbol{\psi}}_\sigma}{f^2_{\mathbf{Z}',\sigma}}\frac{\overline{\boldsymbol{\psi}}^T_{\overline{\sigma}}}{f^2_{\mathbf{Z}',\overline{\sigma}}}g_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma},\nu,\overline{\nu}},$$

We will prove that $T_{6,1,1}=o(1)$. The two other terms can be treated in a similar way. Because of our assumptions, the terms for which $\overline{\sigma}(j_1)\neq\sigma(j_2)$ are zero. This divides the number of possible terms by n'. By using limited expansions as in Equation (E6), we get that $g_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma},\nu,\overline{\nu}}=O(h^{k\alpha-p})$. Therefore, we have $T_{6,1,1}=O\left(\frac{nn'h^p}{nn'}h^{k\alpha-p}\right)=O(h^{k\alpha})=o(1)$.

Concerning $T_{6,2}$, its variance matrix is given by

$$Var\left[T_{6,2}\right]=\frac{\tilde{r}^2_{n,n'}}{|\mathfrak{I}_{k,n'}|^2\cdot|\mathfrak{I}_{k,n}|^2}\sum_{\sigma,\overline{\sigma}\in\mathfrak{I}_{k,n'}}\sum_{\varsigma,\overline{\varsigma}\in\mathfrak{I}_{k,n}}\frac{\mathbb{E}[N_\sigma]-f_{\mathbf{Z}',\sigma}}{f^2_{\mathbf{Z}',\sigma}}\frac{\mathbb{E}[N_{\overline{\sigma}}]-f_{\mathbf{Z}',\overline{\sigma}}}{f^2_{\mathbf{Z}',\overline{\sigma}}}\overline{\boldsymbol{\psi}}_\sigma\overline{\boldsymbol{\psi}}_{\overline{\sigma}}\overline{g}_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma}},$$
$$\overline{g}_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma}}:=\mathbb{E}\left[K_{\sigma,\varsigma}K_{\overline{\sigma},\overline{\varsigma}}\left(g_\varsigma-\theta_\sigma\right)\left(g_{\overline{\varsigma}}-\theta_{\overline{\sigma}}\right)\right]-\mathbb{E}\left[K_{\sigma,\varsigma}\left(g_\varsigma-\theta_\sigma\right)\right]\mathbb{E}\left[K_{\overline{\sigma},\overline{\varsigma}}\left(g_{\overline{\varsigma}}-\theta_{\overline{\sigma}}\right)\right].$$

Note that $\overline{g}_{\sigma,\overline{\sigma},\varsigma,\overline{\varsigma}} = 0$ when $\varsigma \cap \overline{\varsigma} = \emptyset$. This divides the number of terms in the sum above by n, and imposes that $\sigma \cap \overline{\varsigma} \neq 0$, which divides the number of terms in the sum above by another n'. Finally, limited expansions gives a bound of $h^{k\alpha - p}$. Summing up all these elements, we obtain $Var\left[T_{6,2}\right] = O\left(\frac{\tilde{r}^2_{n,n'}}{nn'} h^{k\alpha-p}\right) = O(h^{k\alpha}) = o(1)$. Similarly, we get $\mathbb{E}\left[T_{6,2}\right] = o(1)$ by a Taylor expansion.

E.4 Convergence of T_7 to 0

We recall Equation (E8):

$$T_7 = \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} K_{\sigma,\varsigma} \left(g_\varsigma - \theta_\sigma\right) T_{7,\sigma} \overline{\boldsymbol{\psi}}_\sigma,$$

$$T_{7,\sigma} := \frac{1}{f_{\mathbf{Z}',\sigma}} (1 + \alpha_{7,\sigma})^{-3} \left(\frac{N_\sigma - f_{\mathbf{Z}',\sigma}}{f_{\mathbf{Z}',\sigma}}\right)^2, \text{ with } |\alpha_{7,\sigma}| \leq \left|\frac{N_\sigma - f_{\mathbf{Z}',\sigma}}{f_{\mathbf{Z}',\sigma}}\right|.$$

By Lemma 1 applied to $\mathbf{z}_1 = \mathbf{z}'_{\sigma(1)}, \ldots, \mathbf{z}_{n'} = \mathbf{z}'_{\sigma(n')}$, for $\sigma \in \mathfrak{I}_{k,n'}$, we get

$$\mathbb{P}\left(\sup_{\sigma \in \mathfrak{I}_{k,n'}} \left|N_\sigma - f_{\mathbf{Z}',\sigma}\right| \leq \frac{C_{K,\alpha}}{\alpha} h^\alpha + t\right) \geq 1 - 2\exp\left(-\frac{[n/k]t^2}{h^{-kp}C_1 + h^{-kp}C_2 t}\right),$$

for any $t > 0$. Therefore, $\sup_{\sigma \in \mathfrak{I}_{k,n'}} |T_{7,\sigma}| = O_{\mathbb{P}}(h^{2\alpha})$ by choosing $t = h^{\alpha/k}$. Then,

$$|T_7| \leq \sup_{\sigma \in \mathfrak{I}_{k,n'}} |T_{7,\sigma}| \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} |K_{\sigma,\varsigma}| \cdot \left|g_\varsigma - \theta_\sigma\right| \cdot |\overline{\boldsymbol{\psi}}_\sigma|.$$

The expectation of the double sum is $O(h^\alpha)$, by α-order limited expansions. By Markov's inequality, we deduce

$$T_7 = O_{\mathbb{P}}\left(\tilde{r}_{n,n'} \sup_{\sigma \in \mathfrak{I}_{k,n'}} |T_{7,\sigma}| h^\alpha\right) = O_{\mathbb{P}}(\tilde{r}_{n,n'} h^{3\alpha}) = O_{\mathbb{P}}\left((nn' h^{p+3\alpha})^{1/2}\right),$$

therefore $T_7 = o_{\mathbb{P}}(1)$.

E.5 Convergence of T_3 to 0

We have

$$T_3 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \alpha_{3,\sigma} \cdot \left(\hat{\theta}\left(\mathbf{z}'_{\sigma(1)}, \ldots, \mathbf{z}'_{\sigma(k)}\right) - \theta\left(\mathbf{z}'_{\sigma(1)}, \ldots, \mathbf{z}'_{\sigma(k)}\right)\right)^2 \boldsymbol{\psi}\left(\mathbf{z}'_{\sigma(1)}, \ldots, \mathbf{z}'_{\sigma(k)}\right),$$

with $\forall \sigma \in \mathfrak{I}_{k,n'}, |\alpha_{3,\sigma}| \leq C_{\Lambda''}/2$. Therefore

$$T_3 \lesssim \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \left(\hat{\theta}\left(\mathbf{z}'_{\sigma(1)}, \ldots, \mathbf{z}'_{\sigma(k)}\right) - \theta\left(\mathbf{z}'_{\sigma(1)}, \ldots, \mathbf{z}'_{\sigma(k)}\right)\right)^2$$

$$\lesssim \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}|} \left(\frac{1}{|\mathfrak{I}_{k,n}|} \sum_{\varsigma \in \mathfrak{I}_{k,n}} \frac{K_{\sigma,\varsigma}}{f_{\mathbf{Z}',\sigma}} \left(g_\varsigma - \theta_\sigma\right) + K_{\sigma,\varsigma}\left(g_\varsigma - \theta_\sigma\right)\left(\frac{1}{N_\sigma} - \frac{1}{f_{\mathbf{Z}',\sigma}}\right)\right)^2 = T_8 + T_9 + T_{10},$$

where

$$T_8 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma, \overline{\varsigma} \in \mathfrak{I}_{k,n}} \frac{K_{\sigma,\varsigma} K_{\sigma,\overline{\varsigma}}}{f^2_{\mathbf{Z}',\sigma}} \left(g_\varsigma - \theta_\sigma\right)\left(g_{\overline{\varsigma}} - \theta_\sigma\right),$$

$$T_9 := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma, \overline{\varsigma} \in \mathfrak{I}_{k,n}} \frac{K_{\sigma,\varsigma} K_{\sigma,\overline{\varsigma}}}{f_{\mathbf{Z}',\sigma}} \left(g_\varsigma - \theta_\sigma\right)\left(g_{\overline{\varsigma}} - \theta_\sigma\right)\left(\frac{1}{N_\sigma} - \frac{1}{f_{\mathbf{Z}',\sigma}}\right),$$

$$T_{10} := \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma, \overline{\varsigma} \in \mathfrak{I}_{k,n}} K_{\sigma,\varsigma} K_{\sigma,\overline{\varsigma}} \left(g_\varsigma - \theta_\sigma\right)\left(g_{\overline{\varsigma}} - \theta_\sigma\right)\left(\frac{1}{N_\sigma} - \frac{1}{f_{\mathbf{Z}',\sigma}}\right)^2.$$

We show that $T_8 = o(1)$. The two other terms can be treated in a similar way.

$$\begin{aligned}
\mathbb{E}\left[|T_8|\right] &= \mathbb{E}\left[\frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma, \overline{\varsigma} \in \mathfrak{I}_{k,n}} \frac{|K_{\sigma,\varsigma} K_{\sigma,\overline{\varsigma}}|}{f^2_{\mathbf{Z}',\sigma}} \left|g_\varsigma - \theta_\sigma\right| \cdot \left|g_{\overline{\varsigma}} - \theta_\sigma\right|\right] \\
&= \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma, \overline{\varsigma} \in \mathfrak{I}_{k,n}} \int \frac{\prod_{i=1}^k \left| K_h\left(\mathbf{z}_{\varsigma(i)} - \mathbf{z}'_{\sigma(i)}\right) K_h\left(\mathbf{z}_{\overline{\varsigma}(i)} - \mathbf{z}'_{\sigma(i)}\right)\right|}{f^2_{\mathbf{Z}',\sigma}} \\
&\quad \cdot \left|g\left(\mathbf{x}_{\varsigma(1:k)}\right) - \theta_\sigma\right| \left|g\left(\mathbf{x}_{\overline{\varsigma}(1:k)}\right) - \theta_\sigma\right| \prod_{i \,\in\, \varsigma(1:k) \,\cup\, \overline{\varsigma}(1:k)} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) d\mu(\mathbf{x}_i) d\mathbf{z}_i.
\end{aligned}$$

Note that terms for which $\varsigma \neq \overline{\varsigma} \in \mathfrak{I}_{k,n'}$ are zero, because the $\mathbf{z}'_i$ are distinct and because of our Assumption 16(i). Therefore, we get

$$\begin{aligned}
\mathbb{E}\left[|T_8|\right] &= \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|^2} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \sum_{\varsigma \in \mathfrak{I}_{k,n}} \int \frac{\prod_{i=1}^k K_h\left(\mathbf{z}_{\varsigma(i)} - \mathbf{z}'_{\sigma(i)}\right)^2}{f^2_{\mathbf{Z}',\sigma}} \left(g\left(\mathbf{x}_{\varsigma(1:k)}\right) - \theta_\sigma\right)^2 \\
&\quad \prod_{i \,\in\, \varsigma(1:k)} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) d\mu(\mathbf{x}_i) d\mathbf{z}_i \\
&= \frac{\tilde{r}_{n,n'}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \int \frac{\prod_{i=1}^k K_h\left(\mathbf{z}_i - \mathbf{z}'_{\sigma(i)}\right)^2}{f^2_{\mathbf{Z}',\sigma}} \left(g\left(\mathbf{x}_{\varsigma(1:k)}\right) - \theta_\sigma\right)^2 \prod_{i=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}_i) d\mu(\mathbf{x}_i) d\mathbf{z}_i \\
&= \frac{\tilde{r}_{n,n'} h^{-kp}}{|\mathfrak{I}_{k,n'}| \cdot |\mathfrak{I}_{k,n}|} \sum_{\sigma \in \mathfrak{I}_{k,n'}} \int \frac{\prod_{i=1}^k K(\mathbf{t}_i)^2}{f^2_{\mathbf{Z}',\sigma}} \left(g\left(\mathbf{x}_{\varsigma(1:k)}\right) - \theta_\sigma\right)^2 \prod_{i=1}^k f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_i, \mathbf{z}'_{\sigma(i)} + h\mathbf{t}_i) d\mu(\mathbf{x}_i) d\mathbf{z}_i \\
&= O\left(\frac{\tilde{r}_{n,n'} h^{-kp}}{|\mathfrak{I}_{k,n}|}\right) = O\left(\left(\frac{n \times n' \times h^{(1-k)p}}{|\mathfrak{I}_{k,n}|^2}\right)^{1/2}\right) = o(1).
\end{aligned}$$

■