About Kendall's regression

Alexis Derumigny[*], Jean-David Fermanian[†]

February 22, 2018

Abstract

Conditional Kendall's tau is a measure of dependence between two random variables, conditionally on some covariates. We study nonparametric estimators of such quantities using kernel smoothing techniques. Then, we assume a regression-type relationship between conditional Kendall's tau and covariates, in a parametric setting with possibly a large number of regressors. This model may be sparse, and the underlying parameter is estimated through a penalized criterion. The theoretical properties of all these estimators are stated. We prove non-asymptotic bounds with explicit constants that hold with high probability. We derive their consistency, their asymptotic law and some oracle properties. Some simulations and applications to real data conclude the paper.

Keywords: conditional dependence measures, kernel smoothing, regression-type models

MCS: 62H20, 62F12, 62J12.

1 The framework

Let $(X_1, X_2, \mathbf{Z}) \in \mathbb{R}^{2+p}$ be a random vector, $p \geq 1$. Our goal is to model the dependence between the first two scalar variables X_1 and X_2 given the vector of covariates $\mathbf{Z}$. For $j = 1, 2$ and $\mathbf{z} \in \mathbb{R}^p$,

[*]ENSAE, 5, avenue Henry Le Chatelier, 91764 Palaiseau cedex, France. alexis.derumigny@ensae.fr

[†]ENSAE, 5, avenue Henry Le Chatelier, 91764 Palaiseau cedex, France. jean-david.fermanian@ensae.fr. This research has been supported by the Labex Ecodec.

denote by $F_{j|\mathbf{Z}}(\cdot|\mathbf{Z}=\mathbf{z})$ the conditional cdf of X_j given $\mathbf{Z}=\mathbf{z}$, and by $F^{(-1)}_{j|\mathbf{Z}}(\cdot|\mathbf{Z}=\mathbf{z})$ its generalized inverse. Moreover, let $F_{1,2|\mathbf{Z}}(\cdot,\cdot|\mathbf{Z}=\mathbf{z})$ be the cdf of the random vector $\mathbf{X}:=(X_1,X_2)$ given $\mathbf{Z}=\mathbf{z}$ that will be continuous for every $\mathbf{z}$. Following Patton [12, 13], define the conditional copula of X_1 and X_2 given $\mathbf{Z}=\mathbf{z}$ as

$$C_{1,2|\mathbf{Z}}(u_1,u_2|\mathbf{Z}=\mathbf{z}) := F_{1,2|\mathbf{Z}}\Big(F^{-1}_{1|\mathbf{Z}}(u_1|\mathbf{Z}=\mathbf{z}), F^{-1}_{2|\mathbf{Z}}(u_2|\mathbf{Z}=\mathbf{z})\Big|\mathbf{Z}=\mathbf{z}\Big),$$

for any $(u_1,u_2,\mathbf{z})\in[0,1]^2\times\mathbb{R}^p$, by an extension of Sklar's Theorem. Conditional copulas and their inference have been studied in the recent literature: see Fermanian and Wegkamp [7], Gijbels et al. [9], etc. One of the challenges is often to check whether these copulas depend or not on the conditioning variable (the so-called "simplifying assumption"), a key assumption for vine modeling in particular: see Derumigny and Fermanian [4], Portier and Segers [14], in particular.

In the field of dependence modeling, it is standard to work with dependence measures (i.e. scalars) instead of copulas (i.e. functions). Such summaries of information are very popular and are functionals of the underlying copulas: Kendall's tau, Spearman's rho, Blomqvist's coefficient... See Nelsen [11, Section 5.1.1] for an introduction. In this paper, we focus on dependence measures given some covariates. Of particular interest is the conditional Kendall's tau between X_1 and X_2 given $\mathbf{Z}=\mathbf{z}$, that is defined by

$$\begin{aligned}
&\tau_{1,2|\mathbf{Z}=\mathbf{z}} := 4\int_{[0,1]^2} C_{1,2|\mathbf{Z}}(u_1,u_2|\mathbf{Z}=\mathbf{z})\, C_{1,2|\mathbf{Z}}(du_1,du_2|\mathbf{Z}=\mathbf{z}) - 1\\
&= 4\,\mathbb{P}\big(X_{1,1}>X_{2,1}, X_{1,2}>X_{2,2}\big|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}\big) - 1\\
&= \mathbb{P}\big((X_{1,1}-X_{2,1})(X_{1,2}-X_{2,2})>0\big|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}\big) - \mathbb{P}\big((X_{1,1}-X_{2,1})(X_{1,2}-X_{2,2})<0\big|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}\big),
\end{aligned}$$

for each $\mathbf{z}\in\mathbb{R}^p$, where $(X_{i,1},X_{i,2},\mathbf{Z}_i), i=1,2$ are two independent versions of $(X_1,X_2,\mathbf{Z})\in\mathbb{R}^{2+p}$. We will denote $\tau_{1,2|\mathbf{Z}=\mathbf{z}} = \mathbb{E}[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}]$, where g^* is the function that returns 1 if the pairs in its arguments are concordant, -1 if such pairs are discordant, and 0 else.

Note that, as conditional copulas themselves, conditional Kendall's taus are invariant wrt increasing transformations of the conditional margins X_1 and X_2, given $\mathbf{Z}$. Conditional Kendall's tau are of interest because they measure the evolution of the dependence between X_1 and X_2 when the covariates $\mathbf{Z}$ change. [1]

Several parametric families of copulas have a simple one-to-one mapping between their parameter and the associated Kendall's tau: Gaussian, Student with a fixed degree of freedom, Clayton, Gumbel and Frank copulas, etc. See Nelsen [11, Section 5.1.1] for a presentation of Kendall's tau in such usual unconditional cases. Gijbels et al. [9] introduced non-parametric estimator of conditional Kendall's taus. Apparently, their statistical properties have not been studied "per se" in the literature. For instance, the asymptotic normality of such estimators is seen as a consequence of the weak convergence of the empirical conditional copula process (see Veraverbeke et al. [18]), but its asymptotic law is not specified. The uniform consistency of conditional Kendall's tau has been stated in the case of single index copulas by Fermanian and Lopez [6] in passing. In our paper, we directly focus on finite-distance and asymptotic properties of conditional Kendall's tau,

Moreover, we will assume a linear-type model for conditional Kendall's taus, without any parametric assumption on the underlying copulas. The associated parameter can be obtained through penalized least squares. We will yield finite-distance, asymptotic and oracle properties of such an estimated parameter.

To be specific, assume that the following model holds:

$$\Lambda\left(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\right) = \boldsymbol{\psi}(\mathbf{z})^T \beta^*, \tag{1}$$

for every $\mathbf{z} \in \mathcal{Z}$, a compact subset in $\mathbb{R}^p$. Here, $\beta^* \in \mathbb{R}^{p'}$ denotes the "true" unknown parameter. Λ is a known link function from $[-1,1]$ to $\mathbb{R}$, that is increasing and continuously differentiable. The function $\boldsymbol{\psi}(\cdot) := \left(\psi_1(\cdot), \dots, \psi_{p'}(\cdot)\right)^T$ from $\mathbb{R}^p$ to $\mathbb{R}^{p'}$ is known and corresponds to deterministic transforms of the covariates $\mathbf{z}$. Typically, $p' \gg p$, allowing rich and flexible nonlinear transforms of $\mathbf{z}$. In order to simplify notations, the mention of the conditioning event $\mathbf{Z} \in \mathcal{Z}$ will be omitted. We will set $C_{\mathbf{Z}} := \sup\left\{|\mathbf{z}|_\infty : \mathbf{z} \in \mathcal{Z}\right\}$, denoting by $|\cdot|_q$ the l_q-norm when $1 \le q \le \infty$.

Equation (1) looks like a GLM, and it has three important characteristics:

[1]Our $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ has not to be confused with so-called "conditional Kendall's tau" in the case of truncated data, as in Tsai [17].

1. *The explained variable* $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ *is not observed.* Therefore, a direct estimation of the parameter β^* (for example, by the ordinary least squares, or by the Lasso) is unfeasible. Nevertheless, we will replace $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ by a nonparametric estimate $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$, and use it as an approximation of the explained variable.

2. *It is a "noiseless regression model"*: unlike most regression models, there is no random variable disturbing the variable of interest. In particular, we can make the trivial observation that the quantity $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ is deterministic knowing $\mathbf{Z} = \mathbf{z}$. If we knew the true conditional Kendall's taus, the estimation of β^* in the model (1) would be reduced to a problem of numerical analysis, i.e. the decomposition of a given function into some linear combination of known functions.

3. *The conditioning event is unusual*: usual regression models consider $\mathbb{E}[g(\mathbf{X})|\mathbf{Z} = \mathbf{z}]$ as a function of the conditioning variable $\mathbf{z}$. Here, at the opposite, the expectation is made conditionally on $\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}$. This unusual conditioning event will necessitate some peculiar theoretical treatments.

The functions ψ_i allow to take into account potential non-linearities and even discontinuities of conditional Kendall's taus' with respect to their conditioning variables. We will assume the identifiability of β^*, which is equivalent to the linear independence of the $\psi_i(\cdot)$, by the following proposition.

Proposition 1. *The parameter* β^* *is identifiable in Model (1) if and only if the functions* $(\psi_1(\cdot), \dots, \psi_{p'}(\cdot))$ *are linearly independent* $\mathbb{P}_{\mathbf{Z}}$*-almost everywhere in the sense that, for any vector* $\mathbf{t} = (t_1, \dots, t_{p'})$, $\mathbb{P}_{\mathbf{Z}}\big(\boldsymbol{\psi}(\mathbf{Z})^T\mathbf{t} = 0\big) = 1$ *implies* $\mathbf{t} = 0$.

Proof : Assume that there exists $\mathbf{t} \neq 0$ such that $\mathbb{P}_{\mathbf{Z}}\big(\boldsymbol{\psi}(\mathbf{Z})^T\mathbf{t} = 0\big) = 1$. Then $\Lambda\left(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\right) = \boldsymbol{\psi}(\mathbf{z})^T\beta^* = \boldsymbol{\psi}(\mathbf{z})^T(\beta^* + \mathbf{t})$ for almost every $\mathbf{z}$. As a consequence, β^* and $\beta^* + \mathbf{t}$ induce the same conditional Kendall's taus and the parameter β^* is not identifiable.

Conversely, assume that the model is identifiable. Let $\mathbf{t} \in \mathbb{R}^{p'}$ such that $\mathbb{P}_{\mathbf{Z}}\big(\boldsymbol{\psi}(\mathbf{Z})^T\mathbf{t} = 0\big) = 1$. Then $\Lambda\left(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\right) = \boldsymbol{\psi}(\mathbf{z})^T\beta^* = \boldsymbol{\psi}(\mathbf{z})^T(\beta^* + \mathbf{t})$ holds almost surely. Due to the identifiability, $\beta^* = \beta^* + \mathbf{t}$, that is $\mathbf{t} = 0$. $\square$

A desirable empirical feature of model (1) would be the possibility of obtaining very high/low levels of dependence between X_1 and X_2, for some $\mathbf{Z}$ values, i.e. $\Lambda^{(-1)}(\boldsymbol{\psi}(\mathbf{z})^T\beta^*)$ should be close (or even equal) to 1 or -1 for some $\mathbf{z}$. This can be the case even if $\mathcal{Z}$ is compact, that is required for theoretical reasons. Indeed, Λ is a mapping from $[-1,1]$ into an interval $[\Lambda_{\min},\Lambda_{\max}]$. If $\boldsymbol{\psi}(\mathbf{z})^T\beta^* \geq \Lambda_{\max}$ (resp. $\boldsymbol{\psi}(\mathbf{z})^T\beta^* \leq \Lambda_{\min}$), then we set $\tau_{1,2|\mathbf{z}} = 1$ (resp. $\tau_{1,2|\mathbf{z}} = (-1)$). [2]

Our goal is to estimate the true parameter β^* in the high-dimensional case, i.e. p' is large and possibly larger than the number of observations. We will assume that β^* is sparse, in the sense that $|\mathcal{S}| = |\beta^*|_0 \leq s$, for some $s \in \{1,\dots,p'\}$, where $|\cdot|_0$ is the number of non-zero components of a vector of $\mathbb{R}^{p'}$ and $\mathcal{S}$ is the set of non-zero components of β^*.

Having in mind the model specification (1), we will estimate the parameter β^* using estimated conditional Kendall's tau $\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ as approximations of the "true" conditional Kendall's tau $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$. An iid n-sample $(X_{1,i,}, X_{2,i}, \mathbf{Z}_i)_{i=1,\dots,n}$ will yield such quantities $\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}}$. Afterwards and independently, we will fix n' values $\mathbf{Z}'_i \in \mathcal{Z}$ (fixed design setting). As the number of parameters p' is possibly larger than n and n', we will use a penalized estimator, with the classical l_1 penalty:

$$\hat\beta := \arg\min_{\beta\in\mathbb{R}^{p'}} \left[\frac{1}{n'} \sum_{i=1}^{n'} \left(\Lambda(\hat\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}) - \boldsymbol{\psi}(\mathbf{Z}'_i)^T\beta\right)^2 + \lambda|\beta|_1 \right], \tag{2}$$

where λ is a positive tuning parameter (that may depend on n and n').

In Section 2, we provide the theoretical results concerning the nonparametric estimation of $\tau_{1,2|\mathbf{Z}}$ and the inference of β^* through the latter penalized linear procedure. In particular, we will study the cases when n' is fixed and $n \to \infty$, and when both indices tend to the infinity. Section 3 illustrates the numerical performances of such procedures. Most proofs have been postponed into appendices.

---

[2] It would be tempting to invoke usual transforms as $\Lambda_1(\tau) = \log\left(\frac{1+\tau}{1-\tau}\right)$ (the Fisher transform) or $\Lambda_2(\tau) = \log(-\log((1-\tau)/2))$. Even if such functions are unbounded, they can be invoked, but restricted on some compact subsets of $(-1,1)$. In other words, we need to assume the possible conditional Kendall's tau cannot be larger then $1-\epsilon$, or smaller than $-1+\epsilon$, for some (small) $\epsilon > 0$.

2 Theoretical results

2.1 Estimation of the conditional Kendall's tau

Before studying $\hat{\beta}$, it is necessary to state some results about our estimators of conditional Kendall's tau. Using an i.i.d. sample $(X_{i,1}, X_{i,2}, \mathbf{Z}_i)$, $i = 1, \dots, n$, we estimate the conditional cumulative distribution function of X_j given $\mathbf{Z} = \mathbf{z}$, for $j = 1, 2$ and $\mathbf{z} \in \mathcal{Z}$. Denote by $\hat{F}_{j|\mathbf{Z}}(\cdot|\mathbf{Z} = \mathbf{z})$ this estimator. Then, an estimator of the conditional copula of X_1 and X_2 given $\mathbf{Z} = \mathbf{z}$ is

$$\hat{C}_{1,2|\mathbf{Z}}(u_1, u_2|\mathbf{Z} = \mathbf{z}) := \sum_{i=1}^{n} w_{i,n}(\mathbf{z}) \mathbb{1}\big\{\hat{F}_{1|\mathbf{Z}}(X_{i,1}|\mathbf{Z} = \mathbf{z}) < u_1 \, , \, \hat{F}_{2|\mathbf{Z}}(X_{i,2}|\mathbf{Z} = \mathbf{z}) < u_2\big\}, \tag{3}$$

where $\mathbb{1}$ is the indicator function, $w_{i,n}$ is a sequence of weights given by

$$w_{i,n}(\mathbf{z}) = \frac{K_h(\mathbf{Z}_i - \mathbf{z})}{\sum_{j=1}^{n} K_h(\mathbf{Z}_j - \mathbf{z})}, \tag{4}$$

with $K_h(\cdot) := h^{-p} K(\cdot/h)$ for some kernel K on $\mathbb{R}^p$, and $h = h(n)$ denotes a usual bandwidth sequence that tends to zero when $n \to \infty$. Note that these weights are well-defined if and only if $\sum_{j=1}^{n} K_h(\mathbf{Z}_j - \mathbf{z}) > 0$. Otherwise, simply set $w_{i,n}(\mathbf{z}) = 0$ for every i. The estimator (3) of the conditional copula can be used to define an estimator of the conditional Kendall's tau

$$\begin{aligned}
\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}} &:= 4 \int \hat{C}_{1,2|\mathbf{Z}}(u_1, u_2|\mathbf{Z} = \mathbf{z}) \, \hat{C}_{1,2|\mathbf{Z}}(du_1, du_2|\mathbf{Z} = \mathbf{z}) - 1 \\
&= 4 \sum_{i=1}^{n} w_{i,n}(\mathbf{z}) \hat{C}_{1,2|\mathbf{Z}}\big(\hat{F}_{1|\mathbf{Z}}(X_{i,1}|\mathbf{Z} = \mathbf{z}), \hat{F}_{2|\mathbf{Z}}(X_{i,2}|\mathbf{Z} = \mathbf{z}) \big| \mathbf{Z} = \mathbf{z}\big) - 1 \\
&= 4 \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,n}(\mathbf{z}) w_{j,n}(\mathbf{z}) \mathbb{1}\big\{X_{i,1} < X_{j,1}, X_{i,2} < X_{j,2}\big\} - 1.
\end{aligned} \tag{5}$$

Note that the latter estimator does not depend on $\hat{F}_{1|\mathbf{Z}}$ and $\hat{F}_{2|\mathbf{Z}}$. In other words, we only have to choose the weights $w_{i,n}$ to obtain an estimator of the conditional Kendall's tau. This is coherent with the fact that the conditional Kendall's taus are invariant with respect to conditional marginal distributions. Moreover, note that, in (5), the inequalities are strict, which implies that there are no terms corresponding to the cases $i = j$. This is inline with the definition of (conditional) Kendall's tau itself through coherent/discordant pairs of observations, and it will make the theoretical developments easier.

Let $\hat{f}_{\mathbf{Z}}(\mathbf{z}) := n^{-1}\sum_{j=1}^n K_h(\mathbf{Z}_j - \mathbf{z})$ be an estimator of the marginal density $f_{\mathbf{Z}}$ of the conditioning variable $\mathbf{Z}$. We notice that the estimator $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$ is well-behaved only whenever $\hat{f}_{\mathbf{Z}}(\mathbf{z}) > 0$. Denote the joint density of $(\mathbf{X}, \mathbf{Z})$ by $f_{\mathbf{X},\mathbf{Z}}$. In our study, we need some conditions of regularity.

Assumption 2.1. *The kernel K is bounded, and set $\|K\|_\infty =: C_K$. It is symmetrical and satisfies $\int K = 1$, $\int |K| < \infty$. This kernel is of order α for some integer $\alpha > 1$: for all $j = 1, \dots, \alpha - 1$ and every indices $i_1, \dots, i_j$ in $\{1, \dots, p\}$, $\int_{\mathbb{R}^p} K(\mathbf{u}) u_{i_1} \dots u_{i_j} \, d\mathbf{u} = 0,$.*

Assumption 2.2. *$f_{\mathbf{Z}}$ is α-times continuously differentiable and there exists a constant $C_{K,\alpha} > 0$ s.t., for all $\mathbf{z} \in \mathcal{Z}$,*

$$\int |K|(\mathbf{u}) \sum_{i_1,\dots,i_\alpha=1}^{p} |u_{i_1} \dots u_{i_\alpha}| \sup_{t\in[0,1]} \Big| \frac{\partial^\alpha f_{\mathbf{Z}}}{\partial z_{i_1} \dots \partial z_{i_\alpha}} (\mathbf{z} + t\mathbf{u}) \Big| \, d\mathbf{u} \le C_{K,\alpha}.$$

Assumption 2.3. *There exists a constant $f_{\mathbf{Z},min} > 0$ such that for every $\mathbf{z} \in \mathcal{Z}$, $f_{\mathbf{Z}}(\mathbf{z}) \ge f_{\mathbf{Z},min}$.*

Assumption 2.4. *$f_{\mathbf{Z}}(\cdot) \le f_{\mathbf{Z},max}$ for some finite constant $f_{\mathbf{Z},max}$.*

Proposition 2. *Under Assumptions 2.1-2.4 and if $C_{K,\alpha} h^\alpha/\alpha! < f_{\mathbf{Z},min}$, the estimator $\hat{f}_{\mathbf{Z}}(\mathbf{z})$ is strictly positive with a probability larger than $1 - 2\exp\Big(-nh^p \big(f_{\mathbf{Z},min} - C_{K,\alpha} h^\alpha/\alpha! \big)^2 \,/\, \big(2 f_{\mathbf{Z},max} \int K^2 + (2/3) C_K (f_{\mathbf{Z},min} - C_{K,\alpha} h^\alpha/\alpha!) \big) \Big)$.*

The latter proposition is proved in Section A.1. It will guarantee that $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$ is well-behaved with a probability close to one.

Assumption 2.5. *For every $\mathbf{x} \in \mathbb{R}^2$, $\mathbf{z} \mapsto f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})$ is differentiable almost everywhere up to the order α. For every $0 \le k \le \alpha$ and every $1 \le i_1, \dots, i_\alpha \le p$, let*

$$\mathcal{H}_{k,\vec{\iota}}(\mathbf{u}, \mathbf{v}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{z}) := \sup_{t\in[0,1]} \left| \frac{\partial^k f_{\mathbf{X},\mathbf{Z}}}{\partial z_{i_1} \dots \partial z_{i_k}} \Big(\mathbf{x}_1, \mathbf{z} + t\mathbf{u} \Big) \frac{\partial^{\alpha-k} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{i_{k+1}} \dots \partial z_{i_\alpha}} \Big(\mathbf{x}_2, \mathbf{z} + t\mathbf{v} \Big) \right|,$$

denoting $\vec{\iota} = (i_1, \dots, i_\alpha)$*. Assume that* $\mathcal{H}_{k,\vec{\iota}}(\mathbf{u}, \mathbf{v}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{z})$ *is integrable and there exists a finite constant* $C_{\mathbf{XZ},\alpha} > 0$*, such that, for every* $\mathbf{z} \in \mathcal{Z}$*,*

$$\int_{\mathbb{R}^{4+2p}} |K|(\mathbf{u})|K|(\mathbf{v}) \sum_{k=0}^{\alpha} \binom{\alpha}{k} \sum_{i_1,\dots,i_\alpha=1}^{p} \mathcal{H}_{k,\vec{\iota}}(\mathbf{u}, \mathbf{v}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{z}) |u_{i_1} \dots u_{i_k} v_{i_{k+1}} \dots v_{i_\alpha}| \, d\mathbf{u} \, d\mathbf{v} \, d\mathbf{x}_1 \, d\mathbf{x}_2 \leq C_{\mathbf{XZ},\alpha}.$$

Proposition 3 (Exponential bound for the estimated conditional Kendall's tau)**.** *Under Assumptions 2.1-2.5, for every* $t > 0$ *such that* $C_{K,\alpha} h^\alpha / \alpha! + t < f_{\mathbf{Z},min}/2$ *and every* $t' > 0$*, we have*

$$\mathbb{P}\Bigg(|\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}} - \tau_{1,2|\mathbf{Z}=\mathbf{z}}| > 4\Big(1 + \frac{16 f^2_{\mathbf{Z},max}}{f^3_{\mathbf{Z},min}} \Big(\frac{C_{K,\alpha} h^\alpha}{\alpha!} + t \Big)\Big) \times \Big(\frac{C_{\mathbf{XZ},\alpha} h^\alpha}{f^2_{\mathbf{Z}}(\mathbf{z}) \alpha!} + t' \Big) \Bigg)$$
$$\leq 2 \exp\Big(- \frac{n h^p t^2}{2 f_{\mathbf{Z},max} \int K^2 + (2/3) C_K t} \Big) + 2 \exp\Big(- \frac{(n-1) h^{2p} t'^2 f^4_{\mathbf{Z},min}}{4 f^2_{\mathbf{Z},max} (\int K^2)^2 + (8/3) C_K^2 f^2_{\mathbf{Z},min} t'} \Big).$$

Remark 4. *In the latter inequality, the constant* $f_{\mathbf{Z},min}$ *can be replaced by* $f_{\mathbf{Z}}(\mathbf{z})$*. Moreover, when* K *is compactly supported,* $f_{\mathbf{Z},max}$ *can be replaced by* $\sup_{\tilde{\mathbf{z}} \in \mathcal{V}(\mathbf{z},\epsilon)} f_{\mathbf{Z}}(\tilde{\mathbf{z}})$*, denoting by* $\mathcal{V}(\mathbf{z}, \epsilon)$ *a closed ball of center* $\mathbf{z}$ *and any radius* $\epsilon > 0$*.*

This technical proposition will be key hereafter. It is proved in Section A.2. As a corollary, we have proven the weak consistency of $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$ for every $\mathbf{z} \in \mathcal{Z}$, under the assumptions of Proposition 3, if $nh^{2p} \to \infty$ (set $t = 1$ and $t' = \epsilon > 0$). The next proposition states the same result under weaker conditions.

Proposition 5 (Consistency)**.** *Assume 2.1, and that*

- $f_{\mathbf{Z}}$ *and* $\mathbf{z} \mapsto \tau_{1,2|\mathbf{Z}=\mathbf{z}}$ *are continuous on* $\mathcal{Z}$*,*
- $\lim K(\mathbf{t})|\mathbf{t}|^p = 0$ *when* $|\mathbf{t}| \to \infty$*,*
- $nh_n^p \to \infty$*.*

Then $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$ *tends to* $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ *in probability, when* $n \to \infty$*.*

This property is proved in Section A.3. To derive the asymptotic law of this estimator, we will assume:

Assumption 2.6. *(i) $nh_n^p \to \infty$ and $nh_n^{p+2\alpha} \to 0$; (ii) $K(\cdot)$ is compactly supported.*

Proposition 6 (Asymptotic normality)**.** *Assume 2.1, 2.5, 2.6, and that*

- *$f_{\mathbf{Z}}$ and $\mathbf{z} \mapsto f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})$ are continuous on $\mathcal{Z}$, for every $\mathbf{x}$;*
- *the $\mathbf{Z}'_i$ are distinct.*

Then, $(nh_n^p)^{1/2} \big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\big)_{i=1,\dots,n'} \xrightarrow{D} \mathcal{N}(0, \mathbb{H})$ as $n \to \infty$, where $\mathbb{H}$ is a $n' \times n'$ real matrix defined by

$$[\mathbb{H}]_{i,j} = \frac{4 \int K^2 \mathbb{1}_{\{\mathbf{Z}'_i = \mathbf{Z}'_j\}}}{f_{\mathbf{Z}}(\mathbf{Z}'_i)} \big\{ \mathbb{E}[g^*(\mathbf{X}_1, \mathbf{X}) g^*(\mathbf{X}_2, \mathbf{X}) | \mathbf{Z} = \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}'_i] - \tau^2_{1,2|\mathbf{Z}=\mathbf{Z}'_i} \big\},$$

for every $1 \leq i, j \leq n'$, where $(\mathbf{X}, \mathbf{Z})$, $(\mathbf{X}_1, \mathbf{Z}_1)$, $(\mathbf{X}_2, \mathbf{Z}_2)$ are independent copies.

This proposition is proved in Subsection A.4. We recall that $g^*(\mathbf{X}_1, \mathbf{X}_2) := \mathbb{1}_{\{(X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) > 0\}} - \mathbb{1}_{\{(X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) < 0\}}$, denoting the bivariate vectors $\mathbf{X}_k := (X_{k,1}, X_{k,2})$, $k = 1, 2$. Note that the latter function is symmetrical: $g^*(\mathbf{x}, \mathbf{y}) = g^*(\mathbf{y}, \mathbf{x})$ for every bivariate vectors $\mathbf{x}$ and $\mathbf{y}$.

2.2 Estimation of β^* by penalized linear regression on estimated conditional Kendall's taus

In the previous section, we have provided and study some estimators of conditional Kendall's taus. Having in mind the model specification (1), we now estimate the parameter β^* using the estimated conditional Kendall's tau $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$ as an approximation of the "true" $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$. As the number of parameters p' is possibly larger than n, we will use the penalized estimator given by (2), invoking the usual l_1 penalty. The estimation algorithm is summed up in the following Algorithm 1. Its time complexity is $O\big(n'n(n+p)\big)$, plus the cost of the convex optimization program (2).

Remark 7. *Instead of a fixed design setting $(\mathbf{Z}'_i)_{i=1,\dots,n'}$ in the optimization program, it would be possible to consider a random design: simply draw n' realizations of $\mathbf{Z}$, independently of the n-sample that has been used for the estimation of the conditional Kendall's taus. The differences between fixed and random designs are mainly a matter of presentation and the reader could easily*

Algorithm 1: Estimation algorithm for $\hat{\beta}$

for $i \leftarrow 1$ **to** n' **do**
  for $j \leftarrow 1$ **to** n **do**
    Compute the kernel $K_{i,j} \leftarrow K_h(\mathbf{Z}_j - \mathbf{Z}'_i)$.
  end
  for $j \leftarrow 1$ **to** n **do**
    Compute the weight $w_{j,n}(\mathbf{Z}'_i) \leftarrow K_{i,j} / \sum_{j=1}^n K_{i,j}$ (Equation (4)).
  end
  Compute the conditional Kendall's tau $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i}$ using Equation (5).
end
Solve the convex optimization program (2) and return $\hat{\beta}$.

rewrite our results in a random design setting. We have preferred the former one to study the finite distance properties and asymptotics when n' is fixed (Subsection 2.3). When n and n' will tend to the infinity (Subsection 2.4), both designs are encompassed de facto because we assume the weak convergence of the empirical distribution associated to the sample $(\mathbf{Z}'_i)_{i=1,\dots,n'}$, when $n' \to \infty$.

Our first goal is to prove finite-distance bounds in probability for the estimator $\hat{\beta}$. Let us introduce some notations. Let $\mathbb{Z}'$ be the matrix of size $n' \times p'$ whose lines are $\boldsymbol{\psi}(\mathbf{Z}'_i)^T$, $i = 1, \dots, n'$, and let $\mathbf{Y} \in \mathbb{R}^{n'}$ be the column vector whose components are $Y_i = \Lambda(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i})$, $i = 1, \dots, n'$. For a vector $\mathbf{v} \in \mathbb{R}^{p'}$, denote by $||\mathbf{v}||_{n'} := |\mathbf{v}|_2^2 / n'$ its empirical norm. We can then rewrite the criterion (2) as

$$\hat{\beta} := \arg \min_{\beta \in \mathbb{R}^{p'}} \Big[||\mathbf{Y} - \mathbb{Z}'\beta||_{n'}^2 + \lambda |\beta|_1 \Big],$$

where $\mathbf{Y}$ and $\mathbb{Z}'$ may be considered as "observed", so that the practical problem is reduced to a standard Lasso estimation procedure. Define some "residuals" by

$$\xi_{i,n} := \Lambda(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i}) - \boldsymbol{\psi}(\mathbf{Z}'_i)^T \beta^* = \Lambda(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i}) - \Lambda(\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}), \; i = 1, \dots, n'. \tag{6}$$

Note that these $\xi_{i,n}$ are not "true residuals" in the sense that they do not depend on the estimator $\hat{\beta}$, but on the true parameter β^*. We also emphasized the dependence on n in the notation $\xi_{i,n}$, which is a consequence of the estimated conditional Kendall's tau.

To get non-asymptotic bounds on $\hat{\beta}$, we will assume the *Restricted Eigenvalue* (RE) condition, introduced by Bickel et al. [2]. 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*

$$\kappa(s, c_0) := \min_{\substack{J_0 \subset \{1,\dots,p'\} \\ Card(J_0) \le s}} \quad \min_{\substack{\delta \neq 0 \\ |\delta_{J_0^C}|_1 \le c_0 |\delta_{J_0}|_1}} \frac{|\mathbb{Z}'\delta|_2}{\sqrt{n'}|\delta|_2} > 0.$$

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. [1, Section 8.1] for references and a discussion. Moreover, we will need the following assumption.

Assumption 2.7. *The functions* Λ', $\boldsymbol{\psi}$ *are bounded by some constants, respectively denoted by* $C_{\Lambda'}$ *and* $C_{\boldsymbol{\psi}}$.

Theorem 8 (Fixed design case)**.** *Suppose that Assumptions 2.1-2.5 and 2.7 hold and that the design matrix* $\mathbb{Z}'$ *satisfies the* $RE(s,3)$ *condition. Choose the tuning parameter as* $\lambda = \gamma t$, *with* $\gamma \ge 4$ *and* $t > 0$, *and assume that we choose* h *small enough such that*

$$h^\alpha \le \frac{f_{\mathbf{Z},min}\alpha!}{2\,C_{K,\alpha}}, \text{ and} \tag{7}$$

$$h^\alpha \le \frac{f^4_{\mathbf{Z},min}\alpha!}{2\,C_\psi C_{\Lambda'}(f^2_{\mathbf{Z},min} + 16 f^2_{\mathbf{Z},max})C_{\mathbf{XZ},\alpha}}\, t. \tag{8}$$

Then, we have

$$\mathbb{P}\left(||\mathbb{Z}'(\hat\beta - \beta^*)||_{n'} \le \frac{4(\gamma+1)t\sqrt{s}}{\kappa(s,3)} \text{ and } |\hat\beta - \beta^*|_q \le \frac{4^{2/q}(\gamma+1)ts^{1/q}}{\kappa^2(s,3)}, \text{ for every } 1 \le q \le 2 \right)$$
$$\ge 1 - 2n' \exp\Big(-nh^p C_1 \Big) - 2n' \exp\left(-\frac{(n-1)h^{2p}t^2}{C_2 + C_3 \cdot t} \right), \tag{9}$$

where $C_1 := f^2_{\mathbf{Z},min}/\big(8 f_{\mathbf{Z},max}\int K^2 + (4/3)C_K f_{\mathbf{Z},min}\big)$, $C_2 := 4\{8C_\psi C_{\Lambda'}(f^2_{\mathbf{Z},min} + 16 f^2_{\mathbf{Z},max})\int K^2\}^2/f^6_{\mathbf{Z},min}$, *and* $C_3 := (64/3)C_\psi C_{\Lambda'} C_K^2 (f^2_{\mathbf{Z},min} + 16 f^2_{\mathbf{Z},max})/f^4_{\mathbf{Z},min}$.

This theorem, proved in Section B.2, yields some bounds that holds in probability for the prediction error $||\mathbb{Z}'(\hat\beta - \beta^*)||_{n'}$ and for the estimation error $|\hat\beta - \beta^*|_q$, $1 \le q \le 2$, under the specification (1). Note that the influence of n' and p' is hidden through the Restricted Eigenvalue number $\kappa(s,3)$. The result depends on three parameters γ, t and h. Apparently, the choice of γ seems to be easy, as a larger γ deteriorates the upper bounds. Nonetheless, it is a bit misleading because $\hat{\beta}$ implicitly depends on λ and then on γ (for a fixed t). Nonetheless, choosing $\gamma = 4$ is a reasonable "by default" choice. Moreover, a lower t provides a smaller upper bound, but at the same time the probability of this event is lowered. This induces a trade-off between the probability of the desired event and the size of the bound, as we want the smallest possible bound with the highest probability. Moreover, we cannot choose a too small t, because of the lower bound (8): t is limited by a value proportional to h^α. The latter h cannot be chosen as too small, otherwise the probability in Equation (9) will decrease. To be short: *low values of h and t yield a sharper upper bound with a lower probability, and the opposite.* Therefore, there is a compromise to be found, depending of the kind of result we are interested in.

Clearly, we would like to exhibit the sharpest upper bounds in (9), with the "highest probabilities". Let us look for parameters of the form $t \propto n^{-a}$ and $h \propto n^{-b}$, with $a, b > 0$. The assumptions of Theorem 8 imply $b\alpha \geq a$ (to satisfy (8)) and $1 - 2a - 2pb > 0$ (so that the right-hand side of (9) tends to 1 as $n \to \infty$, i.e. $nh^p \to \infty$ and $nt^2h^{2p} \to \infty$). For fixed α and p, what are the "optimal" choices a and b under the constraints $b\alpha \geq a$ and $1 - 2a - 2pb > 0$? The latter domain is the interior of a triangle in the plane $(a,b) \in \mathbb{R}_+^2$, whose vertices are $O := (0,0)$, $A := (0, 1/(2p))$ and $B := (\alpha/(2p+2\alpha), 1/(2p+2\alpha))$, plus the segment $]0, B[$. All points in such a domain would provide admissible couples (a,b) and then admissible tuning parameters (t,h). In particular, choosing the neighborhood of B, i.e. $a = \alpha(1-\epsilon)/(2p+2\alpha)$ and $b = 1/(2p+2\alpha)$ for some (small) $\epsilon > 0$, will be nice because the upper bounds will be minimized.

Corollary 9. *For $0 < \epsilon < 1$, choosing the parameters $\lambda = 4t$, $t = (n-1)^{-\alpha(1-\epsilon)/(2\alpha+2p)}$ and*

$$h = c_h (n-1)^{-1/(2\alpha+2p)}, \; c_h := \left(\frac{f_{\mathbf{Z},min}^4 \alpha!}{2\, C_\psi C_{\Lambda'} (f_{\mathbf{Z},min}^2 + 16 f_{\mathbf{Z},max}^2) C_{\mathbf{XZ},\alpha}} \right)^{\frac{1}{\alpha}},$$

for any n sufficiently large, we have

$$\mathbb{P}\Bigg(||\mathbb{Z}'(\hat{\beta}-\beta^*)||_{n'} \leq \frac{20\sqrt{s}}{\kappa(s,3)(n-1)^{\alpha(1-\epsilon)/(2\alpha+2p)}} \;\; \textit{and}$$
$$|\hat{\beta}-\beta^*|_q \leq \frac{5\cdot 4^{2/q}s^{1/q}}{\kappa^2(s,3)(n-1)^{\alpha(1-\epsilon)/(2\alpha+2p)}}, \; \textit{for every } 1\leq q\leq 2 \Bigg)$$
$$\geq 1-2n'\exp\Big(-C_1c_h^p(n-1)^{(2\alpha+p)/(2\alpha+2p)}\Big)-2n'\exp\Big(-\frac{c_h^{2p}(n-1)^{2\alpha\epsilon/(2p+2\alpha)}}{C_2+C_3(n-1)^{-\alpha(1-\epsilon)/(2\alpha+2p)}}\Big).$$

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

In this part, n' is still supposed to be fixed and we state the consistency and the asymptotic normality asymptotic of $\hat{\beta}$ as $n\to\infty$. As above, we adopt a fixed design: the $\mathbf{Z}'_i$ are arbitrarily fixed or, equivalently, our reasonings are made conditionally on the second sample.

For $n,n'>0$, denote by $\hat{\beta}_{n,n'}$ the estimator (2) with $h=h_n$ and $\lambda=\lambda_{n,n'}$. The following lemma provides another representation of this estimator $\hat{\beta}_{n,n'}$ that will be useful for deriving its asymptotic properties. It is proved in Section C.1.

Lemma 10. *We have $\hat{\beta}_{n,n'}=\arg\min_{\beta\in\mathbb{R}^{p'}}\mathbb{G}_{n,n'}(\beta)$, where*

$$\mathbb{G}_{n,n'}(\beta) := \frac{2}{n'}\sum_{i=1}^{n'}\xi_{i,n}\boldsymbol{\psi}(\mathbf{Z}'_i)^T(\beta^*-\beta)+\frac{1}{n'}\sum_{i=1}^{n'}\big\{\boldsymbol{\psi}(\mathbf{Z}'_i)^T(\beta^*-\beta)\big\}^2+\lambda_{n,n'}|\beta|_1. \tag{10}$$

We will also invoke a so-called *convexity argument*, summarized in Kato [10] e.g.: "Let g_n and g_∞ be random convex functions taking minimum values at x_n and x_∞, respectively. If all finite dimensional distributions of g_n converge weakly to those of g_∞ and x_∞ is the unique minimum point of g_∞ with probability one, then x_n converges weakly to x_∞."

Theorem 11 (Consistency of $\hat{\beta}$)**.** *Under the assumptions of Proposition 5, if n' is fixed and $\lambda=\lambda_{n,n'}\to\lambda_0$, then, given $\mathbf{Z}'_1,\dots,\mathbf{Z}'_{n'}$ and as n tends to the infinity,*

$$\hat{\beta}_{n,n'}\xrightarrow{\mathbb{P}}\beta^{**} := \inf_{\beta}\mathbb{G}_{\infty,n'}(\beta),\;\; \mathbb{G}_{\infty,n'}(\beta) := \frac{1}{n'}\sum_{i=1}^{n'}\big(\boldsymbol{\psi}(\mathbf{Z}'_i)^T(\beta^*-\beta)\big)^2+\lambda_0|\beta|_1.$$

In particular, if $\lambda_0 = 0$ *and* $< \boldsymbol{\psi}(\mathbf{Z}'_1), \dots, \boldsymbol{\psi}(\mathbf{Z}'_{n'}) >= \mathbb{R}^{p'}$, *then* $\hat{\beta}_{n,n'} \xrightarrow{\mathbb{P}} \beta^*$.

Proof : By Proposition 5, the first term in the r.h.s. of (10) converges to 0 as $n \to \infty$. The third term in the r.h.s. of (10) converges to $\lambda_0 |\beta_1|$ by assumption. We have just proven that $\mathbb{G}_{n,n'} \to \mathbb{G}_{\infty,n'}$ pointwise as $n \to \infty$.

We can now apply the convexity argument, because $\mathbb{G}_{n,n'}$ and $\mathbb{G}_{\infty,n'}$ are convex functions. As a consequence, $\arg\min_\beta \mathbb{G}_{n,n'}(\beta) \to \arg\min_\beta \mathbb{G}_{\infty,n'}(\beta)$ in law. Since we have adopted a fixed design setting, β^{**} is non random, given $(\mathbf{Z}'_1, \dots, \mathbf{Z}'_{n'})$. The convergence in law towards a deterministic quantity implies convergence in probability, which concludes the proof.

Moreover, when $\lambda_0 = 0$, β^* is the minimum of $\mathbb{G}_{\infty,n'}$ because the vectors $\boldsymbol{\psi}(\mathbf{Z}'_i)$, $i = 1, \dots, p'$ generate the space $\mathbb{R}^{p'}$. Therefore, this implies the consistency of $\hat{\beta}_{n,n'}$. $\square$

To evaluate the limiting behavior of $\hat{\beta}_{n,n'}$, we need the joint asymptotic normality of $(\xi_{1,n}, \dots, \xi_{n',n})$, when $n \to \infty$ and given $\mathbf{Z}'_1, \dots, \mathbf{Z}'_{n'}$. By applying the Delta-method to the function $\Lambda(\,\cdot\,)$ component-wise, this is given in the following corollary of Proposition 6.

Corollary 12. *Under the assumptions of Proposition 6 and given* $(\mathbf{Z}'_1, \dots, \mathbf{Z}'_{n'})$, *we have*

$$(nh_n^p)^{1/2} \left[\xi_{1,n}, \dots, \xi_{n',n}\right]^T \xrightarrow{D} \mathcal{N}\left(0, \tilde{\mathbb{H}}\right),$$

where $\tilde{\mathbb{H}}$ *is a* $n' \times n'$ *real matrix, defined for every integers* $1 \leq i, j \leq n'$ *by*

$$[\tilde{\mathbb{H}}]_{i,j} := \frac{4 \int K^2 \mathbb{1}_{\{\mathbf{Z}'_i = \mathbf{Z}'_j\}}}{f_{\mathbf{Z}}(\mathbf{Z}'_i)} \times \Big(\Lambda'\big(\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\big)\Big)^2 \times \Big\{\mathbb{E}[g^*(\mathbf{X}_1, \mathbf{X}) g^*(\mathbf{X}_2, \mathbf{X}) | \mathbf{Z} = \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}'_i] - \tau^2_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\Big\}.$$

Theorem 13 (Asymptotic law of the estimator)**.** *Under the assumptions of Proposition 6, and if* $\lambda_{n,n'}(nh^p_{n,n'})^{1/2} \to \ell$ *when* $n \to \infty$, *we have*

$$(nh^p_{n,n'})^{1/2}(\hat{\beta}_{n,n'} - \beta^*) \xrightarrow{D} \mathbf{u}^* := \arg\min_{\mathbf{u} \in \mathbb{R}^{p'}} \mathbb{F}_{\infty,n'}(\mathbf{u}),$$

given $\mathbf{Z}'_1, \ldots, \mathbf{Z}'_{n'}$ *and as* $n \to \infty$, *where*

$$\mathbb{F}_{\infty,n'}(\mathbf{u}) := \frac{2}{n'} \sum_{i=1}^{n'} \sum_{j=1}^{p'} W_i \psi_j(\mathbf{Z}'_i) u_j + \frac{1}{n'} \sum_{i=1}^{n'} \big(\boldsymbol{\psi}(\mathbf{Z}'_i)^T \mathbf{u}\big)^2 + \ell \sum_{i=1}^{p'} \big(|u_i| \mathbb{1}_{\{\beta^*_i = 0\}} + u_i \ \mathrm{sign}(\beta^*_i) \mathbb{1}_{\{\beta^*_i \neq 0\}} \big),$$

with $\mathbf{W} = (W_1, \ldots, W_{n'}) \sim \mathcal{N}\left(0, \tilde{\mathbb{H}}\right)$.

This theorem is proved in Section C.2. When $\ell = 0$, we can say more about the limiting law in general. Indeed, in such a case, $\mathbf{u}^* = \arg\min_{\mathbf{u} \in \mathbb{R}^{p'}} \mathbb{F}_{\infty,n'}(\mathbf{u})$ is the solution of the first order conditions $\nabla \mathbb{F}_{\infty,n'}(\mathbf{u}) = 0$, that are written as

$$\sum_{i=1}^{n'} W_i \boldsymbol{\psi}(\mathbf{Z}'_i) + \sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{Z}'_i)(\boldsymbol{\psi}(\mathbf{Z}'_i)^T \mathbf{u}) = 0.$$

Therefore,

$$\mathbf{u}^* = -\Big(\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{Z}'_i) \boldsymbol{\psi}(\mathbf{Z}'_i)^T \Big)^{-1} \sum_{i=1}^{n'} W_i \boldsymbol{\psi}(\mathbf{Z}'_i),$$

when $\Sigma_{n'} := \sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{Z}'_i) \boldsymbol{\psi}(\mathbf{Z}'_i)^T$ is invertible. In this case, the limiting law of $(nh^p_{n,n'})^{1/2}(\hat{\beta}_{n,n'} - \beta^*)$ is Gaussian, and its asymptotic covariance is $V_{as} := \Sigma_{n'}^{-1} \sum_{i,j=1}^{n'} [\tilde{\mathbb{H}}]_{i,j} \boldsymbol{\psi}(\mathbf{Z}'_i) \boldsymbol{\psi}(\mathbf{Z}'_j)^T \Sigma_{n'}^{-1}$.

Let remember that $\mathcal{S} := \{j : \beta^*_j \neq 0\}$ and assume that $|\mathcal{S}| = s < p$ so that the true model depends on a subset of predictors. In the same spirit as Fan and Li [5], we say that an estimator $\hat{\beta}$ satisfies the oracle property if

- $v_n(\hat{\beta}_{\mathcal{S}} - \beta^*_{\mathcal{S}})$ converges in law towards a continuous random vector, for some conveniently chosen rate of convergence (v_n), and
- we identify the nonzero components of the true parameter β^* with probability one when the sample size n is large, i.e. the probability of the event $\{j : \hat{\beta}_j \neq 0\} = \mathcal{S}$ tends to one.

As above, let us fix n' and n will tend to the infinity. Then, denote $\{j : \hat{\beta}_j \neq 0\}$ by $\mathcal{S}_n$, that will depends on n' implicitly. It is well-known that the usual Lasso estimator does not fulfill the oracle property (Zou [19]). Here, this is still the case. The following proposition is proved in Section C.3.

Proposition 14. *Under the assumptions of Theorem 13,* $\limsup_n \mathbb{P}\left(\mathcal{S}_n = \mathcal{S}\right) = c < 1$.

A usual way of obtaining the oracle property is to modify our estimator in an "adaptive" way. Following Zou [19], consider a first step 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. Now, let us consider the same optimization program as in (2) but with a random tuning parameter given by $\lambda_{n,n'} := \mu_{n,n'}/|\tilde{\beta}|^\delta$, for some constant $\delta > 0$ and some positive deterministic sequence $(\mu_{n,n'})$. The corresponding adaptive estimator (solution of the modified Equation (2)) will be denoted by $\check{\beta}_{n,n'}$, or simply $\check{\beta}$. Hereafter, we still set $\mathcal{S}_n = \{j : \check{\beta}_j \neq 0\}$. The following theorem is proved in Section C.4.

Theorem 15 (Asymptotic law of the adaptive estimator of β)**.** *Under the assumptions of Proposition 6, if $\mu_{n,n'}(nh^p_{n,n'})^{1/2} \to \ell \geq 0$ and $\mu_{n,n'}(nh^p_{n,n'})^{1/2}\nu_n^\delta \to \infty$ when $n \to \infty$, we have*

$$(nh^p_{n,n'})^{1/2}(\check{\beta}_{n,n'} - \beta^*)_{\mathcal{S}} \xrightarrow{D} \mathbf{u}^{**}_{\mathcal{S}} := \arg\min_{\mathbf{u}_{\mathcal{S}} \in \mathbb{R}^s} \check{\mathbb{F}}_{\infty,n'}(\mathbf{u}_{\mathcal{S}}), \text{ where}$$

$$\check{\mathbb{F}}_{\infty,n'}(\mathbf{u}_{\mathcal{S}}) := \frac{2}{n'}\sum_{i=1}^{n'}\sum_{j\in\mathcal{S}} W_i\psi_j(\mathbf{Z}'_i)u_j + \frac{1}{n'}\sum_{i=1}^{n'}\Big(\sum_{j\in\mathcal{S}}\psi_j(\mathbf{Z}'_i)u_j\Big)^2 + \ell\sum_{i\in\mathcal{S}}\frac{u_i}{|\beta^*_i|^\delta}\,\mathrm{sign}(\beta^*_i),$$

with $\mathbf{W} = (W_1, \dots, W_{n'}) \sim \mathcal{N}\left(0, \tilde{\mathbb{H}}\right)$. Moreover, when $\ell = 0$, the oracle property is fulfilled: when n tends to the infinity, $\mathbb{P}\left(\mathcal{S}_n = \mathcal{S}\right) \to 1$.

2.4 Asymptotic properties of $\hat{\beta}$ when n and n' jointly tend to $+\infty$

Now, we consider a framework in which both n and n' are going to the infinity, while the dimensions p and p' stay fixed. To be specific, n and n' will not be allowed to independently go to the infinity. In particular, for a given n, the other size $n'(n)$ (simply denoted as n') will be constrained, as detailed in the assumptions below. In this section, we still work conditionally on $\mathbf{Z}'_1, \dots, \mathbf{Z}'_{n'}, \dots$. The latter vectors are considered as "fixed", inducing a deterministic sequence. Alternatively, we could consider randomly drawn $\mathbf{Z}'_i$ from a given law. The latter case can easily been stated from the results below but its specific statement is left to the reader.

Theorem 16 (Consistency jointly in (n, n'))**.** *Assume that Assumptions 2.1-2.5 and 2.7 are sat-*

isfied. Assume that $\sum_{i=1}^{n'} \psi(\mathbf{Z}'_i)\psi(\mathbf{Z}'_i)^T/n'$ *converges to a matrix* $M_{\psi,\mathbf{Z}'}$*, as* $n' \to \infty$*. Assume that* $\lambda_{n,n'} \to \lambda_0$ *and* $n' \exp(-Anh^{2p}) \to 0$ *for every* $A > 0$*, when* $(n, n') \to \infty$*. Then*

$$\hat{\beta}_{n,n'} \xrightarrow{\mathbb{P}} \arg\min_{\beta \in \mathbb{R}^{p'}} \mathbb{G}_{\infty,\infty}(\beta), \text{ as } (n, n') \to \infty,$$

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^**.*

This theorem is proved in Section D.1. Note that, since the sequence $(\mathbf{Z}'_i)$ is deterministic, we just assume the usual convergence of $\sum_{i=1}^{n'} \psi(\mathbf{Z}'_i)\psi(\mathbf{Z}'_i)^T/n'$ in $\mathbb{R}^{p'^2}$. Moreover, if the "second subset" $(\mathbf{Z}'_i)_{i=1,\dots,n'}$ were a random sample that were 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 $\hat{\beta}_{n,n'}$.

Assumption 2.8. *(i) The support of the kernel* $K(\cdot)$ *is included into* $[-1, 1]^p$*. Moreover, for all* n, n' *and every* $(i_1, i_2) \in \{1, \dots, n'\}^2$*,* $i_1 \neq i_2$*, we have* $|\mathbf{Z}'_{i_1} - \mathbf{Z}'_{i_2}|_\infty > 2h_{n,n'}$*.*

(ii) We have the following convergences : (a) $n'(nh_{n,n'}^{p+4\alpha} + h_{n,n'}^{2\alpha} + (nh_{n,n'}^p)^{-1}) \to 0$*, and (b)* $\lambda_{n,n'}(n'\, n\, h_{n,n'}^p)^{1/2} \to 0$*.*

(iii) The distribution $\mathbb{P}_{\mathbf{Z}',n'} := \frac{1}{n'}\sum_{i=1}^{n'} \delta_{\mathbf{Z}'_i}$ *weakly converges as* $n' \to \infty$*, to a distribution* $\mathbb{P}_{\mathbf{Z}',\infty}$ *on* $\mathbb{R}^p$*, with a density* $f_{\mathbf{Z}',\infty}$ *with respect to the p-dimensional Lebesgue measure.*

(iv) The matrix $V_1 := \int \boldsymbol{\psi}(\mathbf{z}')\boldsymbol{\psi}(\mathbf{z}')^T f_{\mathbf{Z}',\infty}(\mathbf{z}')d\mathbf{z}'$ *is non-singular.*

(v) Λ *is two times continuously differentiable and its second derivative is bounded by a constant* $C_{\Lambda''}$*.*

Part (i) of this assumption forbids the design points $(\mathbf{Z}'_i)_{i\geq 1}$ from being too close to each other too fast, with respect to the rate of convergence of the sequence of bandwidths $(h_{n,n'})$ to 0. This can be guaranteed by choosing an appropriate design. For example, if $p = 1$ and $\mathcal{Z} = [0, 1]$, we can choose the dyadic sequence $1/2, 1/4, 3/4, 1/8, 3/8, 5/8, 7/8, \dots$

Part (ii) can be ensured by first choosing first a slowly growing sequence $n'(n)$, and then by choosing h that would tend to 0 fast enough. Note that a compromise has to be found concerning these two rates. The sequence $\lambda_{n,n'}$ should be chosen at last, so that (b) is satisfied.

The design points $\mathbf{Z}'_i$ are deterministic, similarly to all results in the present paper. Equivalently, all results can be seen as given conditionally to the sample $(\mathbf{Z}'_i)_{i\geq 1}$. For every n', we can still use the non-random measure $\mathbb{P}_{\mathbf{Z}',n'}$ and we impose in Part (iii) that it converges to a measure with a density w.r.t. the Lebesgue measure. Intuitively, this means we do not want to observe design points that would be repeated infinitely often (this would result in a Dirac component in the limit of $\mathbb{P}_{\mathbf{Z}',n'}$). Note that we do not impose any condition on the density $f_{\mathbf{Z}',\infty}$, but we have already imposed that the true density $f_{\mathbf{Z}}$ of the variable $\mathbf{Z}$ is bounded (see Assumptions 2.3 and 2.4). An optimal choice of the density $f_{\mathbf{Z}',\infty}$ is not an easy task. Indeed, even if we knew exactly the true density $f_{\mathbf{Z}}$ (which is rather unlikely in applications), there is no obvious reasons why this choice would better than others. If we want a small asymptotic variance $\tilde{V}_{as}$, the distribution of the design should concentrate the $\mathbf{Z}'_i$ in the regions where $\Lambda'\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}'}\big)^2$ is small and where $\boldsymbol{\psi}(\mathbf{z}')\boldsymbol{\psi}(\mathbf{z}')^T$ is big.

Part (iv) of the assumption is usual, and ensure that the design is somehow "asymptotically full rank". Note that this matrix V_1 will also appear in the asymptotic variance, similarly to the result we have obtained in the case of the ordinary least squares.

Part (v) will allow us to control the remainder term in a Taylor expansion of Λ. Notice that this assumption was not necessary in the previous section, that is, with the simple asymptotic framework, where we used the Delta-method on the vector $(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i})_{i=1,\dots,n'}$. But when the number of terms n' tends to infinity, we have to use the second derivative to control a remainder term (see the details about this remainder term T_3, Equation (22)).

Theorem 17 (Asymptotic law of the estimator jointly in (n,n'))**.** *Under Assumptions 2.1-2.8, we have* $(nn'h^p_{n,n'})^{1/2}(\hat{\beta}_{n,n'} - \beta^*) \xrightarrow{D} \mathcal{N}(0,\tilde{V}_{as})$*, where* $\tilde{V}_{as} := V_1^{-1}V_2V_1^{-1}$ *and*

$$V_2 := \int K(\mathbf{t})^2 d\mathbf{t} \int g^*(\mathbf{x}_1,\mathbf{x}_3)g^*(\mathbf{x}_2,\mathbf{x}_3)\Lambda'\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}'}\big)^2\boldsymbol{\psi}(\mathbf{z}')\boldsymbol{\psi}(\mathbf{z}')^T,$$
$$\times f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{z}')f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{Z}=\mathbf{z}')f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_3|\mathbf{Z}=\mathbf{z}')\frac{f_{\mathbf{Z}',\infty}(\mathbf{z}')}{f_{\mathbf{Z}}(\mathbf{z}')}\,d\mathbf{x}_1 d\mathbf{x}_2 d\mathbf{x}_3 d\mathbf{z}',$$

and V_1 is the matrix defined in Assumption 2.8(iv).

This theorem is proved in Section D.2.

3 Numerical applications

3.1 Simulations

Now, we evaluate the numerical performance of our estimates through a small simulation study. In this subsection, we have chosen $n = 3000$, $n' = 100$ and $p = 1$. The univariate covariate Z follows a uniform distribution between 0 and 1. The marginals $X_1|Z = z$ and $X_2|Z = z$ follow some Gaussian distributions $\mathcal{N}(z, 1)$. The conditional copula of $(X_1, X_2)|Z = z$ belongs to a one-parameter family (here the Gaussian copula family). Therefore, it will be parameterized by its (conditional) Kendall's tau $\tau_{1,2|Z=z}$, and is denoted by $C_{\tau_{1,2|Z=z}}$. Obviously, $\tau_{1,2|Z=z}$ is computed using Model (1). The dependence w.r.t. $Z = z$ is specified by $\tau_{1,2|Z=z} := 3z(1-z) = 3/4 - 3/4(z - 1/2)^2$.

We will choose Λ as the identity function and the Z'_i as a uniform grid on $[0.01, 0.99]$. The values 0 and 1 are excluded to avoid boundaries numerical problems. As for regressors, we will consider $p' = 12$ functions of Z, namely $\psi_1(z) = 1$, $\psi_{i+1}(z) = 2^{-i}(z - 0.5)^i$ for $i = 1, \dots, 5$, $\psi_{5+2i}(z) = \cos(2\pi z/i)$ and $\psi_{6+2i}(z) = \sin(2\pi z/i)$ for $i = 1, 2$, $\psi_{11}(z) = \mathbb{1}\{z \leq 0.4\}$, $\psi_{12}(z) = \mathbb{1}\{z \leq 0.6\}$. They cover a mix of polynomial, trigonometric and step-functions. Then, the true parameter is $\beta^* = (3/4, 0, -3/4, \mathbf{0}_9)$, where $\mathbf{0}_9$ is the null vector of size 9.

Our reference value of the tuning parameter h is given by the usual rule-of-thumb, i.e. $h = \hat{\sigma}(Z)n^{-1/5}$, where $\hat{\sigma}$ is the estimated standard deviation of Z. Moreover, we designed a cross validation procedure (see Algorithm 2) whose output is a data-driven choice for the tuning parameter $\hat{\lambda}^{cv}$. Finally, we perform the convex optimization of the Lasso criterion using the R package `glmnet` [8].

In our simulations, we observed that the estimation of $\hat{\beta}$ are not very satisfying if the family of function ψ_i is far too large. Indeed, our model will "learn the noise" produced by the kernel estimation, and there will be "overfitting" in the sense that the function $\Lambda^{(-1)}\big(\boldsymbol{\psi}(\cdot)^T\hat{\beta}\big)$ will be very

Algorithm 2: Cross-validation algorithm for choosing λ.

Divide the dataset $\mathcal{D} = (X_{i,1}, X_{i,2}, \mathbf{Z}_i)_{i=1,\dots,n}$ into N disjoint blocks $\mathcal{D}_1, \dots, \mathcal{D}_N$;
foreach λ **do**
  for $k \leftarrow 1$ **to** N **do**
    Estimate the conditional Kendall's taus $\hat{\tau}^{(k)}_{1,2|\mathbf{Z}=\mathbf{Z}'_i}, i = 1, \dots, n'$ on the dataset $\mathcal{D}_k$;
    Estimate $\beta^{(-k)}$ using Algorithm 1 on the dataset $\mathcal{D} \backslash \mathcal{D}_k$ with the tuning parameter λ ;
    Compute $Err_k(\lambda) := \sum_{i=1,\dots,n'} \left(\hat{\tau}^{(k)}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \boldsymbol{\psi}(\mathbf{Z}'_i)^T \beta^{(-k)} \right)^2$;
  end
end
Return $\hat{\lambda}^{cv} := \arg\min_\lambda \sum_k Err_k(\lambda)$.

close to $\hat{\tau}_{1,2|\mathbf{Z}=\cdot}$, but not to the target $\tau_{1,2|\mathbf{Z}=\cdot}$. Therefore, we have to find a compromise between misspecification (to choose a family of ψ_i that is not rich enough), and over-fitting (to choose a family of ψ_i that is too rich). Note that in our simulations, we have replaced $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$ by the asymptotically equivalent symmetrical estimator

$$\hat{\tau}^*_{1,2|\mathbf{Z}=\mathbf{z}} := \sum_{i,j=1}^n w_{i,n}(\mathbf{z}) w_{j,n}(\mathbf{z}) \Big(\mathbb{1}\big\{ (X_{i,1} - X_{j,1}).(X_{i,2} - X_{j,2}) > 0 \big\} - \mathbb{1}\big\{ (X_{i,1} - X_{j,1}).(X_{i,2} - X_{j,2}) < 0 \big\} \Big),$$

that provides more stable results in the limiting cases.

We have led 100 simulations for couples of tuning parameters (λ, h), where $\lambda \propto \hat{\lambda}^{cv}$, and $h \propto \hat{\sigma}(Z) n^{-1/5}$. The results in term of empirical bias and standard deviation of $\hat{\beta}$ are displayed in Figure 1. Empirically, we find the smallest h tend to perform better than the largest ones. The influence of the tuning parameter λ (around reasonable values) is less clear. Finally, we selected $h = 0.25 \hat{\sigma}(Z) n^{-1/5}$ and $\lambda = 2\hat{\lambda}^{cv}$. With the latter choice, the coefficient by coefficient results are provided in Table 1. The empirical results are relatively satisfying, despite a small amount of over-fitting. In particular, the estimation procedure is able to identify the non-zero coefficients almost systematically.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$	$\hat{\beta}_9$	$\hat{\beta}_{10}$	$\hat{\beta}_{11}$	$\hat{\beta}_{12}$
Value	0.75	0	0.75	0	0	0	0	0	0	0	0	0
Bias	-0.13	3.6e-05	0.26	0.0033	-0.045	-0.0051	-0.011	-2e-04	-3.2e-05	0.073	-0.0013	0.00021
Std. dev.	0.15	0.00041	0.18	0.035	0.078	0.041	0.022	0.0051	0.00037	0.15	0.007	0.0041
Prob.	1	0.015	0.96	0.015	0.4	0.069	0.36	0.076	0.0076	0.33	0.038	0.023

Table 1: Estimated bias, standard deviation and probability of being non-null for each estimated component of β ($h = 0.25\, \hat{\sigma}(Z) n^{-1/5}$ and $\lambda = 2\hat{\lambda}^{cv}$).

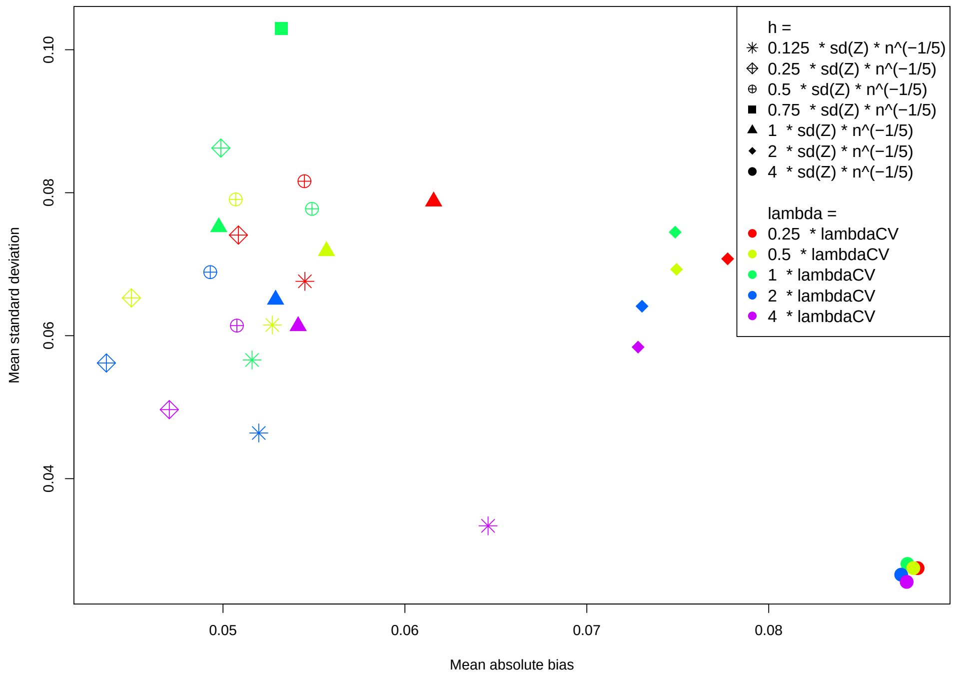

Figure 1: Mean absolute bias $\sum_{i=1}^{12} |\mathbb{E}[\hat{\beta}_i] - \beta_i^*|/12$ and mean standard deviation $\sum_{i=1}^{12} \sigma(\hat{\beta}_i)/12$, for different data-driven choices of the tuning parameters h and λ.

To give a complete picture, for one particular simulated sample, we show the results of the estimation procedure, as displayed in Figures 2, 3 and 4.

3.2 Real data application

Now, we apply the model given by (1) to a real dataset. From the website of the World Factbook of the Central Intelligence Agency, we have collected data of male and female life expectancy and GDP per capita for $n = 206$ countries in the world. We seek to analyze the dependence between male and female life expectancies conditionally on the GDP per capita, i.e. given the explanatory variable $Z = \log 10(GDP/capita)$. This dataset and these variables are similar as those in the first example given in Gijbels et al. [9].

We use $n' = 100$, $h = 2\sigma(Z)n^{-1/5}$ and the same family of functions ψ_i as above (once composed with a linear transform to be defined on $[\min(Z), \max(Z)]$). As expected, the levels of conditional dependence between male and female expectancies are strongly dependent overall. When some

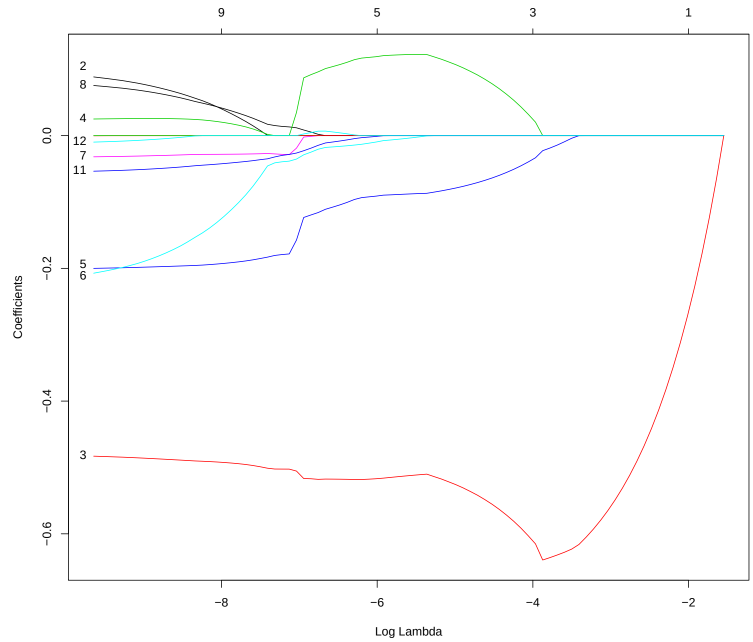

Figure 2: Evolution of the estimated non-zero coefficients as a function of the regularization parameter λ. Note that the coefficients $\hat{\beta}_9$ and $\hat{\beta}_{10}$ coefficients are always zero (and are not displayed).

countries are by epidemics or wars, both genders suffer. And the opposite when an efficient national Health system is put in place. Nonetheless, we observe a decrease of the conditional Kendall's tau when Z is larger, up to $Z \simeq 4.5$. This is inline with the analysis of Gijbels et al. [9]. Globally, when countries become richer, more developed and safe, men and women less and less depend on their environment (and on its risks, potentially). Interestingly, when Z become even larger (the richest countries in the world), conditional dependencies between male and female life expectancies increase again, because men and women behave similarly in terms of way of life. They can benefit from the same levels of security and health.

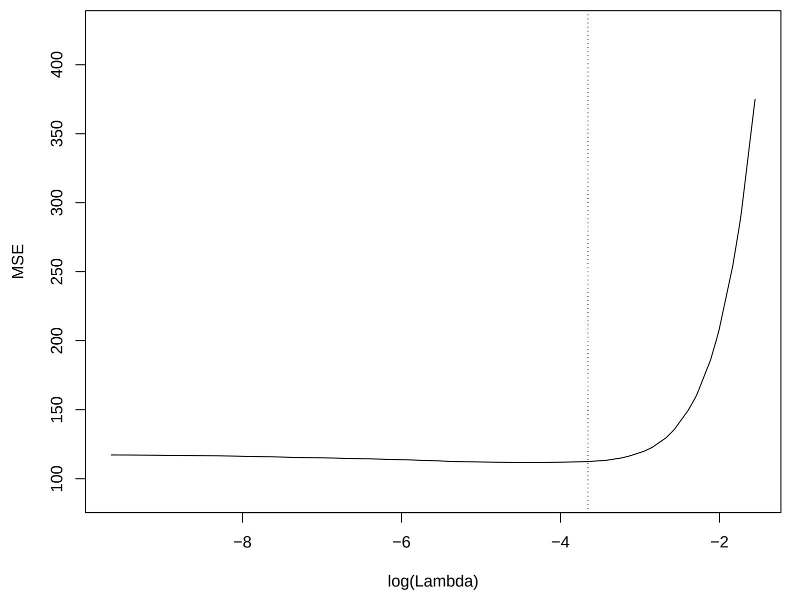

Figure 3: 5-fold cross-validation mean-squared error as a function of the regularization parameter λ. The dotted line correspond to the value $2 \cdot \hat{\lambda}^{cv}$ where $\hat{\lambda}^{cv}$ is selected by Algorithm 2.

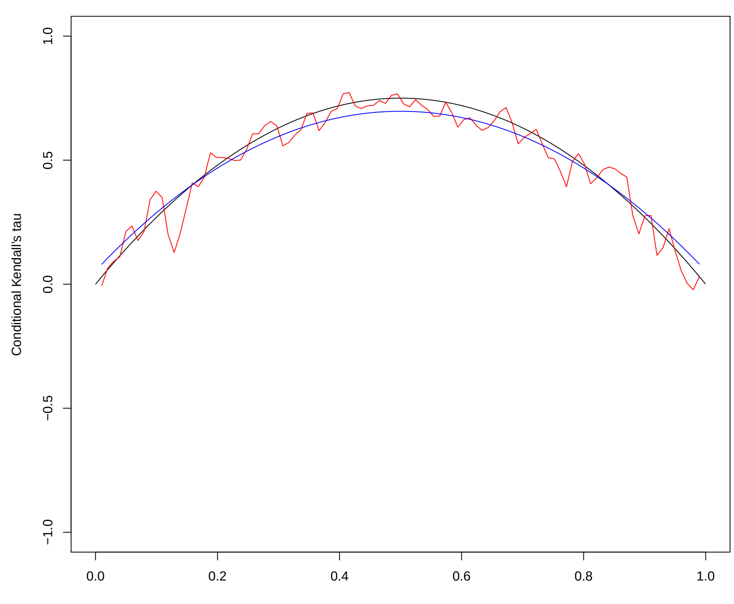

Figure 4: True conditional Kendall's tau $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ (black curve), estimated conditional Kendall's tau $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$ (red curve), and prediction $\Lambda^{(-1)}\big(\boldsymbol{\psi}(\mathbf{z})^T \hat{\beta}\big)$ (blue curve) as a function of $\mathbf{z}$. For the blue curve, the regularization parameter is $2\hat{\lambda}^{cv}$ where $\hat{\lambda}^{cv}$ is selected by Algorithm 2.

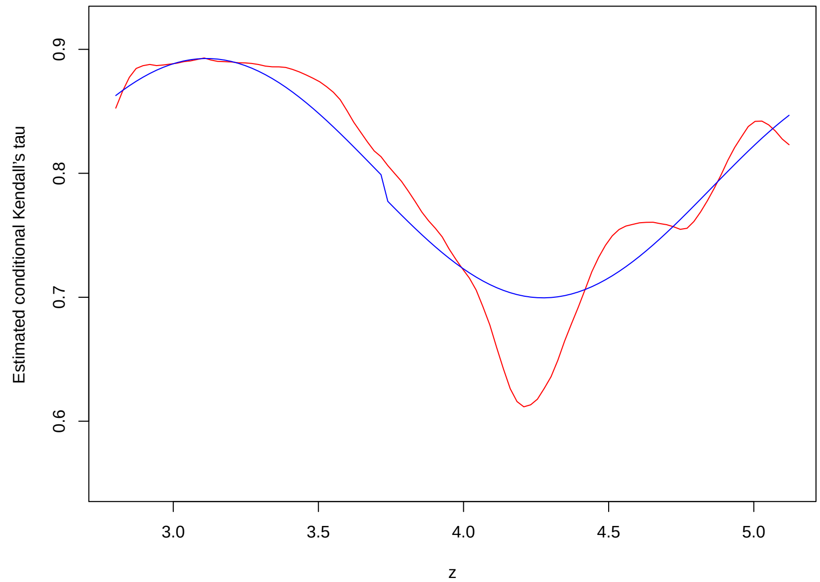

Figure 5: Estimated conditional Kendall's tau $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$ (red curve), and prediction $\Lambda^{(-1)}\big(\boldsymbol{\psi}(\mathbf{z})^T\hat{\beta}\big)$ (blue curve) as a function of $\mathbf{z}$ for the application on real data, where the estimated non-zero coefficients are $\hat{\beta}_1 = 0.78$, $\hat{\beta}_7 = -0.043$, $\hat{\beta}_8 = 0.069$ and $\hat{\beta}_{11} = 0.020$. The regularization parameter is $\hat{\lambda}^{cv}/2$ where $\hat{\lambda}^{cv}$ is selected by Algorithm 2.

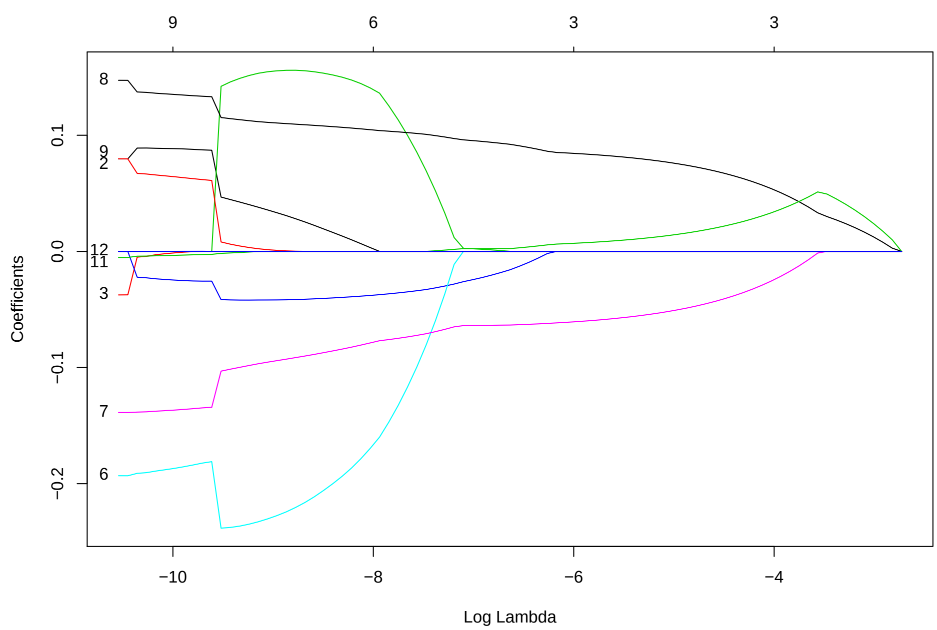

Figure 6: Evolution of the estimated non-zero coefficients as a function of the regularization parameter λ for the application on real data. Note that all the other ψ_i coefficients are zero (except the intercept).

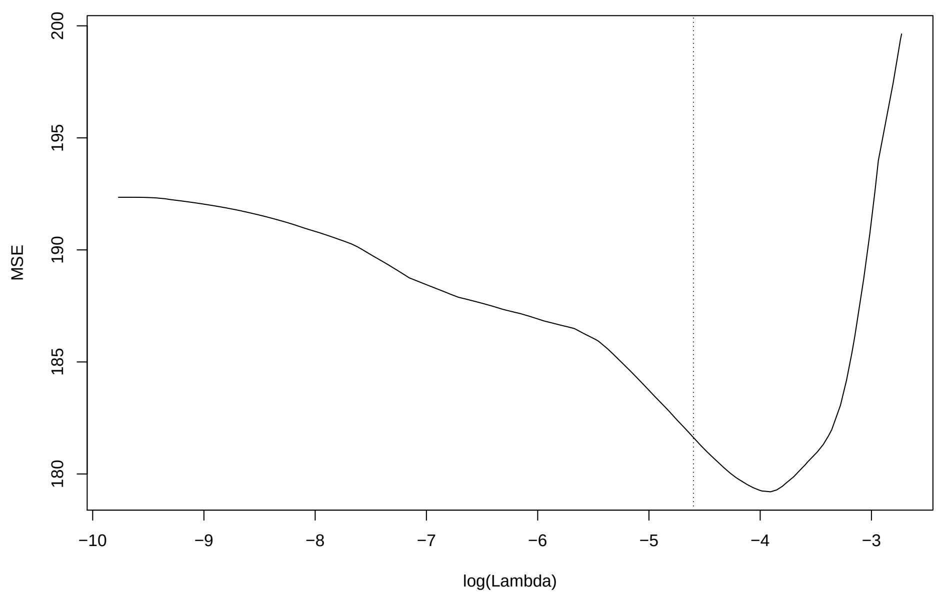

Figure 7: 5-fold cross-validation mean-squared error as a function of the regularization parameter λ for the application on real data. The dotted line correspond to the value $\hat{\lambda}^{cv}/2$ where $\hat{\lambda}^{cv}$ is selected by Algorithm 2.

References

[1] Pierre C. Bellec, Guillaume Lecué, and Alexandre B. Tsybakov. Slope meets lasso: improved oracle bounds and optimality. *ArXiv preprint, arXiv:1605.08651*, 2016.

[2] Peter J Bickel, Ya'acov Ritov, Alexandre B Tsybakov, et al. Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, 37(4):1705–1732, 2009.

[3] Denis Bosq and Jean-Pierre Lecoutre. *Théorie de l'estimation fonctionnelle*. Economica, 1987.

[4] Alexis Derumigny and Jean-David Fermanian. About tests of the "simplifying" assumption for conditional copulas. *Dependence Modeling*, 5(1):154–197, 2017.

[5] Jianqing Fan and Runze Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

[6] Jean-David Fermanian and Olivier Lopez. Single-index copulae. *Journal of Multivariate Analysis*, 165:27–55, 2018.

[7] Jean-David Fermanian and Marten H. Wegkamp. Time-dependent copulas. *Journal of Multivariate Analysis*, 110:19–29, 2012.

[8] Jerome Friedman, Trevor Hastie, Rob Tibshirani, and Noah Simon. glmnet: Lasso and elastic-net regularized generalized linear models. R package version 2.0–2, 2017.

[9] Irène Gijbels, Noël Veraverbeke, and Marek Omelka. Conditional copulas, association measures and their applications. *Computational Statistics & Data Analysis*, 55(5):1919–1932, 2011.

[10] Kengo Kato. Asymptotics for argmin processes: Convexity arguments. *Journal of Multivariate Analysis*, 100(8):1816–1829, 2009.

[11] Roger B. Nelsen. *An introduction to copulas*. Springer Science & Business Media, 2007.

[12] Andrew J. Patton. Estimation of multivariate models for time series of possibly different lengths. *Journal of applied econometrics*, 21(2):147–173, 2006.

[13] Andrew J. Patton. Modelling asymmetric exchange rate dependence. *International economic review*, 47(2):527–556, 2006.

[14] François Portier and Johan Segers. On the weak convergence of the empirical conditional copula under a simplifying assumption. *ArXiv preprint, arXiv:1511.06544*, 2015.

[15] Robert J. Serfling. *Approximation theorems of mathematical statistics.* John Wiley & Sons, 1980.

[16] Winfried Stute. Conditional U-statistics. *The Annals of Probability*, 19(2):812–825, 1991.

[17] Wei-Yann Tsai. Testing the assumption of independence of truncation time and failure time. *Biometrika*, 77(1):169–177, 1990.

[18] Noël Veraverbeke, Marek Omelka, and Irène Gijbels. Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics*, 38(4):766–780, 2011.

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

A Proofs for $\hat{\tau}$

For convenience, remind Berk's (1970) inequality (see Theorem A in Serfling [15, p.201]). Note that, if $m = 1$, this reduces to Bernstein's inequality.

Lemma 18. *Let $m > 0$, $\mathbf{X}_1, \dots, \mathbf{X}_m$ some random vectors with values in a measurable space $\mathcal{X}$ and $g : \mathcal{X}^m \to [a, b]$ be a symmetric real bounded function. Set $\theta := \mathbb{E}[g(\mathbf{X}_1, \dots, \mathbf{X}_m)]$ and $\sigma^2 := Var[g(\mathbf{X}_1, \dots, \mathbf{X}_m)]$. Then, for any $t > 0$ and $n \geq m$,*

$$\mathbb{P}\left(\binom{n}{m}^{-1} \sum_c g(\mathbf{X}_{i_1}, \dots, \mathbf{X}_{i_m}) - \theta \geq t \right) \leq \exp\left(- \frac{[n/m]t^2}{2\sigma^2 + (2/3)(b - \theta)t} \right),$$

where $\sum_c$ denotes summation over all subgroups of m distinct integers $(i_1, \dots, i_m)$ of $\{1, \dots n\}$.

A.1 Proof of Proposition 2

Lemma 19. *Under Assumptions 2.1, 2.2 and 2.4, we have for any $t > 0$,*

$$\mathbb{P}\bigg(\big|\hat{f}_{\mathbf{Z}}(\mathbf{z}) - f_{\mathbf{Z}}(\mathbf{z})\big| \geq \frac{C_{K,\alpha}h^\alpha}{\alpha!} + t\bigg) \leq 2\exp\bigg(-\frac{nh^p t^2}{2f_{\mathbf{Z},max}\int K^2 + (2/3)C_K t}\bigg).$$

This Lemma is proved below. 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 a probability larger than $1 - 2\exp\big(-nh^p t^2/(2f_{\mathbf{Z},max}\int K^2 + (2/3)C_K t)\big)$. We should therefore choose the largest t as possible, which yields Proposition 2.

It remains to prove Lemma 19. Use the usual decomposition between a stochastic component and a bias: $\hat{f}_{\mathbf{Z}}(\mathbf{z}) - f_{\mathbf{Z}}(\mathbf{z}) = \big(\hat{f}_{\mathbf{Z}}(\mathbf{z}) - \mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{z})]\big) + \big(\mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{z})] - f_{\mathbf{Z}}(\mathbf{z})\big)$.

We first upper-bound the bias.

$$\begin{aligned}
\mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{z})] - f_{\mathbf{Z}}(\mathbf{z}) &= \mathbb{E}\big[n^{-1}\sum_{i=1}^n K_h(\mathbf{Z}_i - \mathbf{z})\big] - f_{\mathbf{Z}}(\mathbf{z}) \\
&= \int_{\mathbb{R}^p} K_h(\mathbf{y} - \mathbf{z})\big(f_{\mathbf{Z}}(\mathbf{y}) - f_{\mathbf{Z}}(\mathbf{z})\big)d\mathbf{y} \\
&= \int_{\mathbb{R}^p} K(\mathbf{u})\Big(f_{\mathbf{Z}}\big(\mathbf{z} + h\mathbf{u}\big) - f_{\mathbf{Z}}(\mathbf{z})\Big)d\mathbf{u}.
\end{aligned}$$

Set $\phi_{\mathbf{z},\mathbf{u}}(t) := f_{\mathbf{Z}}\big(\mathbf{z} + th\mathbf{u}\big)$ for $t \in [0, 1]$. This function has at least the same regularity as $f_{\mathbf{Z}}$, so it is α-differentiable. By a Taylor-Lagrange expansion, we get

$$\begin{aligned}
\int_{\mathbb{R}^p} K(\mathbf{u})\Big(f_{\mathbf{Z}}\big(\mathbf{z} + h\mathbf{u}\big) - f_{\mathbf{Z}}(\mathbf{z})\Big)d\mathbf{u} &= \int_{\mathbb{R}^p} K(\mathbf{u})\Big(\phi_{\mathbf{z},\mathbf{u}}(1) - \phi_{\mathbf{z},\mathbf{u}}(0)\Big)d\mathbf{u} \\
&= \int_{\mathbb{R}^p} K(\mathbf{u})\bigg(\sum_{i=1}^{\alpha-1}\frac{1}{i!}\phi^{(i)}_{\mathbf{z},\mathbf{u}}(0) + \frac{1}{\alpha!}\phi^{(\alpha)}_{\mathbf{z},\mathbf{u}}(t_{\mathbf{z},\mathbf{u}})\bigg)d\mathbf{u},
\end{aligned}$$

for some real number $t_{\mathbf{z},\mathbf{u}} \in (0, 1)$. By Assumption 2.1, for every $i = 1, \dots, \alpha-1$, $\int_{\mathbb{R}^p} K(\mathbf{u})\phi^{(i)}_{\mathbf{z},\mathbf{u}}(0)\, d\mathbf{u} = 0$. Therefore,

$$\Big|\mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{z})] - f_{\mathbf{Z}}(\mathbf{z})\Big| = \bigg|\int_{\mathbb{R}^p} K(\mathbf{u})\frac{1}{\alpha!}\phi^{(\alpha)}_{\mathbf{z},\mathbf{u}}(t_{\mathbf{z},\mathbf{u}})d\mathbf{u}\bigg|$$

$$= \frac{1}{\alpha!}\Bigg| \int_{\mathbb{R}^p} K(\mathbf{u}) \sum_{i_1,\dots,i_\alpha=1}^{p} h^\alpha u_{i_1}\dots u_{i_\alpha} \frac{\partial^\alpha f_{\mathbf{Z}}}{\partial z_{i_1}\dots\partial z_{i_\alpha}} \Big(\mathbf{z} + t_{\mathbf{z},\mathbf{u}} h\mathbf{u}\Big) d\mathbf{u} \Bigg| \leq \frac{C_{K,\alpha}}{\alpha!} h^\alpha.$$

Second, the stochastic component may be written as

$$\hat{f}_{\mathbf{Z}}(\mathbf{z}) - \mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{z})] = n^{-1}\sum_{i=1}^{n} K_h(\mathbf{Z}_i - \mathbf{z}) - \mathbb{E}\Big[n^{-1}\sum_{i=1}^{n} K_h(\mathbf{Z}_i - \mathbf{z})\Big] = n^{-1}\sum_{i=1}^{n} S_i,$$

where $S_i := K_h(\mathbf{Z}_i - \mathbf{z}) - \mathbb{E}[K_h(\mathbf{Z}_i - \mathbf{z})]$. Apply Lemma 18 with $m = 1$ and $g(\mathbf{Z}_i) = S_i$. Here, we have $b = -a = h^{-p}C_K$, $\theta = \mathbb{E}[g(\mathbf{Z}_1)] = 0$ and $\big|Var[g(\mathbf{Z}_1)]\big| \leq h^{-p} f_{\mathbf{Z},max} \int K^2$, and we get

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} K_h(\mathbf{Z}_i - \mathbf{z}) - \mathbb{E}[K_h(\mathbf{Z}_i - \mathbf{z})] \geq t\right) \leq \exp\left(-\frac{nt^2}{2h^{-p} f_{\mathbf{Z},max}\int K^2 + (2/3)h^{-p}C_K t}\right). \square$$

A.2 Proof of Proposition 3

Consider the decomposition

$$\begin{aligned}
\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}} - \tau_{1,2|\mathbf{Z}=\mathbf{z}} &= 4\sum_{1\leq i,j\leq n} w_{i,n}(\mathbf{z}) w_{j,n}(\mathbf{z}) \mathbb{1}\big\{\mathbf{X}_i < \mathbf{X}_j\big\} - 4\mathbb{P}\big(\mathbf{X}_1 < \mathbf{X}_2 \big| \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\big) \\
&= 4\sum_{1\leq i,j\leq n} \frac{K_h(\mathbf{Z}_i - \mathbf{z})K_h(\mathbf{Z}_j - \mathbf{z})}{n^2 \hat{f}^2_{\mathbf{Z}}(\mathbf{z})} \Big(\mathbb{1}\big\{\mathbf{X}_i < \mathbf{X}_j\big\} - \mathbb{P}\big(\mathbf{X}_1 < \mathbf{X}_2 \big| \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\big)\Big) \\
&= 4\frac{f^2_{\mathbf{Z}}(\mathbf{z})}{\hat{f}^2_{\mathbf{Z}}(\mathbf{z})} \times \sum_{1\leq i,j\leq n} \frac{K_h(\mathbf{Z}_i - \mathbf{z})K_h(\mathbf{Z}_j - \mathbf{z})}{n^2 f^2_{\mathbf{Z}}(\mathbf{z})} \Big(\mathbb{1}\big\{\mathbf{X}_i < \mathbf{X}_j\big\} - \mathbb{P}\big(\mathbf{X}_1 < \mathbf{X}_2 \big| \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\big)\Big) \\
&=: 4\frac{f^2_{\mathbf{Z}}(\mathbf{z})}{\hat{f}^2_{\mathbf{Z}}(\mathbf{z})} \times \left(\sum_{1\leq i,j\leq n} S_{i,j}\right).
\end{aligned}$$

The conclusion will follow from the next two lemmas. Then, $f^2_{\mathbf{Z}}(\mathbf{z})/\hat{f}^2_{\mathbf{Z}}(\mathbf{z})$ and $\big|\sum_{1\leq i,j\leq n} S_{i,j}\big|$ will be bounded separately. $\square$

Lemma 20. *Under Assumptions 2.1-2.4 and if $C_{K,\alpha}h^\alpha/\alpha! + t < f_{\mathbf{Z},min}/2$ for some $t > 0$, then*

$$\mathbb{P}\bigg(\frac{f^2_{\mathbf{Z}}(\mathbf{z})}{\hat{f}^2_{\mathbf{Z}}(\mathbf{z})} \geq 1 + \frac{16 f^2_{\mathbf{Z},max}}{f^3_{\mathbf{Z},min}}\Big(\frac{C_{K,\alpha}h^\alpha}{\alpha!} + t\Big)\bigg) \leq 2\exp\left(-\frac{nh^p t^2}{2f_{\mathbf{Z},max}\int K^2 + (2/3)C_K t}\right),$$

and $\hat{f}_{\mathbf{Z}}(\mathbf{z})$ is strictly positive on these events.

Proof : Applying the mean value inequality to the function $x \mapsto 1/x^2$, we get

$$\left|\frac{1}{\hat{f}^2_{\mathbf{Z}}(\mathbf{z})} - \frac{1}{f^2_{\mathbf{Z}}(\mathbf{z})}\right| \leq \frac{2}{f^{*3}_{\mathbf{Z}}}\big|\hat{f}_{\mathbf{Z}}(\mathbf{z}) - f_{\mathbf{Z}}(\mathbf{z})\big|,$$

where $f^*_{\mathbf{Z}}$ lies between $\hat{f}_{\mathbf{Z}}(\mathbf{z})$ and $f_{\mathbf{Z}}(\mathbf{z})$. By Lemma 19, we get

$$\mathbb{P}\Big(\big|\hat{f}_{\mathbf{Z}}(\mathbf{z}) - f_{\mathbf{Z}}(\mathbf{z})\big| \leq \frac{C_{K,\alpha}h^\alpha}{\alpha!} + t\Big) \geq 1 - 2\exp\Big(-\frac{nh^pt^2}{2f_{\mathbf{Z},max}\int K^2 + (2/3)C_K t}\Big).$$

Therefore, on this event, $\big|\hat{f}_{\mathbf{Z}}(\mathbf{z}) - f_{\mathbf{Z}}(\mathbf{z})\big| \leq f_{\mathbf{Z},min}/2$, so that $f_{\mathbf{Z},min}/2 \leq \hat{f}_{\mathbf{Z}}(\mathbf{z})$. We have also $f_{\mathbf{Z},min}/2 \leq f_{\mathbf{Z}}(\mathbf{z})$ and then $f_{\mathbf{Z},min}/2 \leq f^*_{\mathbf{Z}}$. Combining the previous inequalities, we finally get

$$\left|\frac{1}{\hat{f}^2_{\mathbf{Z}}(\mathbf{z})} - \frac{1}{f^2_{\mathbf{Z}}(\mathbf{z})}\right| \leq \frac{16}{f^3_{\mathbf{Z},min}}\big|\hat{f}_{\mathbf{Z}}(\mathbf{z}) - f_{\mathbf{Z}}(\mathbf{z})\big| \leq \frac{16}{f^3_{\mathbf{Z},min}}\Big(\frac{C_{K,\alpha}h^\alpha}{\alpha!} + t\Big),$$

and we deduce the result. □

Lemma 21. *Under Assumptions 2.1-2.5, we have*

$$\mathbb{P}\Big(\Big|\sum_{1\leq i,j\leq n} S_{i,j}\Big| > \frac{C_{\mathbf{XZ},\alpha}h^\alpha}{n^2 f^2_{\mathbf{Z}}(\mathbf{z})\alpha!} + t\Big) \leq 2\exp\Big(-\frac{(n-1)h^{2p}t^2}{16C_K^2 f^2_{\mathbf{Z},min}(\int K^2)^2 + (8/3)C_K^2 f^2_{\mathbf{Z},min} t}\Big).$$

Proof : Note that

$$\sum_{1\leq i,j\leq n} S_{i,j} = \sum_{1\leq i\neq j\leq n}\big(S_{i,j} - \mathbb{E}[S_{i,j}]\big) + n(n-1)\mathbb{E}[S_{1,2}] + \sum_{i=1}^n S_{i,i}.$$

The "diagonal term" $\sum_{i=1}^n S_{i,i} = -\mathbb{P}\big(\mathbf{X}_1 < \mathbf{X}_2 \big| \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\big)\sum_{i=1}^n K_h^2(\mathbf{Z}_i - \mathbf{z})/(n^2\hat{f}^2_{\mathbf{Z}}(\mathbf{z}))$ is negative and negligible. It will be denoted by $-\Delta_n$.

Now, let us deal with the main term, that is decomposed as a stochastic component and a bias component. First, let us deal with the bias. Simple calculations provide, if $i \neq j$,

$$\mathbb{E}[S_{i,j}] = \mathbb{E}\Big[\frac{K_h(\mathbf{Z}_i - \mathbf{z})K_h(\mathbf{Z}_j - \mathbf{z})}{n^2 f^2_{\mathbf{Z}}(\mathbf{z})}\Big(\mathbb{1}\{\mathbf{X}_i < \mathbf{X}_j\} - \mathbb{P}\big(\mathbf{X}_i < \mathbf{X}_j \big| \mathbf{Z}_i = \mathbf{Z}_j = \mathbf{z}\big)\Big)\Big]$$

$$
\begin{aligned}
&= \int_{\mathbb{R}^{2p+2}} \frac{K_h(\mathbf{z}_1 - \mathbf{z})K_h(\mathbf{z}_2 - \mathbf{z})}{n^2 f^2_{\mathbf{Z}}(\mathbf{z})} \Big(\mathbb{1}\big\{\mathbf{x}_1 < \mathbf{x}_2\big\} - \mathrm{I\!P}\big(\mathbf{X}_i < \mathbf{X}_j \big| \mathbf{Z}_i = \mathbf{Z}_j = \mathbf{z}\big)\Big) \\
&\qquad\qquad\qquad\qquad \times f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1, \mathbf{z}_1)\, f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2, \mathbf{z}_2)\, d\mathbf{x}_1\, d\mathbf{z}_1\, d\mathbf{x}_2\, d\mathbf{z}_2 \\
&= \int_{\mathbb{R}^{2p+2}} \frac{K(\mathbf{u})K(\mathbf{v})}{n^2 f^2_{\mathbf{Z}}(\mathbf{z})} \Big(\mathbb{1}\big\{\mathbf{x}_1 < \mathbf{x}_2\big\} - \mathrm{I\!P}\big(\mathbf{X}_i < \mathbf{X}_j \big| \mathbf{Z}_i = \mathbf{Z}_j = \mathbf{z}\big)\Big) \\
&\qquad\qquad\qquad\qquad \times f_{\mathbf{X},\mathbf{Z}}\Big(\mathbf{x}_1, \mathbf{z} + h\mathbf{u}\Big)\, f_{\mathbf{X},\mathbf{Z}}\Big(\mathbf{x}_2, \mathbf{z} + h\mathbf{v}\Big)\, d\mathbf{x}_1\, d\mathbf{u}\, d\mathbf{x}_2\, d\mathbf{v} \\
&= \int_{\mathbb{R}^{2p+2}} \frac{K(\mathbf{u})K(\mathbf{v})}{n^2 f^2_{\mathbf{Z}}(\mathbf{z})} \Big(\mathbb{1}\big\{\mathbf{x}_1 < \mathbf{x}_2\big\} - \mathrm{I\!P}\big(\mathbf{X}_i < \mathbf{X}_j \big| \mathbf{Z}_i = \mathbf{Z}_j = \mathbf{z}\big)\Big) \\
&\qquad \Bigg(f_{\mathbf{X},\mathbf{Z}}\Big(\mathbf{x}_1, \mathbf{z} + h\mathbf{u}\Big)\, f_{\mathbf{X},\mathbf{Z}}\Big(\mathbf{x}_2, \mathbf{z} + h\mathbf{v}\Big) - f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1, \mathbf{z})\, f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2, \mathbf{z}) \Bigg)\, d\mathbf{x}_1\, d\mathbf{u}\, d\mathbf{x}_2\, d\mathbf{v},
\end{aligned}
$$

because, for every $\mathbf{z}$,

$$
\begin{aligned}
0 &= \int_{\mathbb{R}^4} \Big(\mathbb{1}\big\{\mathbf{x}_1 < \mathbf{x}_2\big\} - \mathrm{I\!P}\big(\mathbf{X}_1 < \mathbf{X}_2 \big| \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\big)\Big) f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}(\mathbf{x}_1) f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}(\mathbf{x}_2) d\mathbf{x}_1 d\mathbf{x}_2 \\
&= \int_{\mathbb{R}^4} \Big(\mathbb{1}\big\{\mathbf{x}_1 < \mathbf{x}_2\big\} - \mathrm{I\!P}\big(\mathbf{X}_1 < \mathbf{X}_2 \big| \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\big)\Big) \frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1, \mathbf{z}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2, \mathbf{z})}{f^2_{\mathbf{Z}}(\mathbf{z})} d\mathbf{x}_1 d\mathbf{x}_2. \qquad (11)
\end{aligned}
$$

We apply the Taylor-Lagrange formula to the function

$$
\phi_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{u},\mathbf{v}}(t) := f_{\mathbf{X},\mathbf{Z}}\Big(\mathbf{x}_1, \mathbf{z} + th\mathbf{u}\Big)\, f_{\mathbf{X},\mathbf{Z}}\Big(\mathbf{x}_2, \mathbf{z} + th\mathbf{v}\Big).
$$

With obvious notations, this yields

$$
\begin{aligned}
\mathrm{I\!E}[S_{i,j}] &= \int \frac{K(\mathbf{u})K(\mathbf{v})}{n^2 f^2_{\mathbf{Z}}(\mathbf{z})} \Big(\mathbb{1}\big\{\mathbf{x}_1 < \mathbf{x}_2\big\} - \mathrm{I\!P}\big(\mathbf{X}_i < \mathbf{X}_j \big| \mathbf{Z}_i = \mathbf{Z}_j = \mathbf{z}\big)\Big) \\
&\qquad \cdot \Big(\phi_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{u},\mathbf{v}}(1) - \phi_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{u},\mathbf{v}}(0) \Big)\, d\mathbf{x}_1\, d\mathbf{u}\, d\mathbf{x}_2\, d\mathbf{v} \\
&= \int \frac{K(\mathbf{u})K(\mathbf{v})}{n^2 f^2_{\mathbf{Z}}(\mathbf{z})} \Big(\mathbb{1}\big\{\mathbf{x}_1 < \mathbf{x}_2\big\} - \mathrm{I\!P}\big(\mathbf{X}_i < \mathbf{X}_j \big| \mathbf{Z}_i = \mathbf{Z}_j = \mathbf{z}\big)\Big) \\
&\qquad \cdot \Bigg(\sum_{k=1}^{\alpha-1} \frac{1}{k!} \phi^{(k)}_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{u},\mathbf{v}}(0) + \frac{1}{\alpha!} \phi^{(\alpha)}_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{u},\mathbf{v}}(t_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{u},\mathbf{v}}) \Bigg)\, d\mathbf{x}_1\, d\mathbf{u}\, d\mathbf{x}_2\, d\mathbf{v} \\
&= \int \frac{K(\mathbf{u})K(\mathbf{v})}{n^2 f^2_{\mathbf{Z}}(\mathbf{z})} \Big(\mathbb{1}\big\{\mathbf{x}_1 < \mathbf{x}_2\big\} - \mathrm{I\!P}\big(\mathbf{X}_i < \mathbf{X}_j \big| \mathbf{Z}_i = \mathbf{Z}_j = \mathbf{z}\big)\Big) \Bigg(\frac{1}{\alpha!} \phi^{(\alpha)}_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{u},\mathbf{v}}(t_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{u},\mathbf{v}}) \Bigg)\, d\mathbf{x}_1\, d\mathbf{u}\, d\mathbf{x}_2\, d\mathbf{v}.
\end{aligned}
$$

Since $\phi^{(\alpha)}_{\mathbf{x}_1,\mathbf{x}_2,\mathbf{u},\mathbf{v}}(t)$ is equal to

$$\sum_{k=0}^{\alpha}\binom{\alpha}{k}\sum_{i_1,\dots,i_\alpha=1}^{p} h^\alpha u_{i_1}\dots u_{i_k}v_{i_{k+1}}\dots v_{i_\alpha}\frac{\partial^k f_{\mathbf{X},\mathbf{Z}}}{\partial z_{i_1}\dots\partial z_{i_k}}\Big(\mathbf{x}_1,\mathbf{z}+th\mathbf{u}\Big)\cdot\frac{\partial^{\alpha-k} f_{\mathbf{X},\mathbf{Z}}}{\partial z_{i_{k+1}}\dots\partial z_{i_\alpha}}\Big(\mathbf{x}_2,\mathbf{z}+th\mathbf{v}\Big),$$

using Assumption 2.5, we get

$$\big|\mathbb{E}[S_{1,2}]\big| \leq C_{\mathbf{XZ},\alpha}h^\alpha/(n^2 f^2_{\mathbf{Z}}(\mathbf{z})\alpha!). \tag{12}$$

Second, the stochastic component will be bounded from above. Indeed,

$$\sum_{1\leq i\neq j\leq n}(S_{i,j}-\mathbb{E}[S_{i,j}]) = \frac{1}{n^2}\sum_{1\leq i\neq j\leq n} g\big((\mathbf{X}_i,\mathbf{Z}_i)\,,\,(\mathbf{X}_j,\mathbf{Z}_j)\big),$$

with the function g defined by

$$\begin{aligned}
g\big((\mathbf{X}_i,\mathbf{Z}_i),(\mathbf{X}_j,\mathbf{Z}_j)\big) &:= \frac{K_h(\mathbf{Z}_i-\mathbf{z})K_h(\mathbf{Z}_j-\mathbf{z})}{f^2_{\mathbf{Z}}(\mathbf{z})}\Big(\mathbb{1}\big\{\mathbf{X}_i<\mathbf{X}_j\big\}-\mathbb{P}\big(\mathbf{X}_i<\mathbf{X}_j\big|\mathbf{Z}_i=\mathbf{Z}_j=\mathbf{z}\big)\Big)\\
&-\mathbb{E}\Big[K_h(\mathbf{Z}_i-\mathbf{z})K_h(\mathbf{Z}_j-\mathbf{z})f^{-2}_{\mathbf{Z}}(\mathbf{z})\Big(\mathbb{1}\big\{\mathbf{X}_i<\mathbf{X}_j\big\}-\mathbb{P}\big(\mathbf{X}_i<\mathbf{X}_j\big|\mathbf{Z}_i=\mathbf{Z}_j=\mathbf{z}\big)\Big)\Big].
\end{aligned}$$

The symmetrized version of g is $\tilde g_{i,j}=\Big(g\big((\mathbf{X}_i,\mathbf{Z}_i)\,,\,(\mathbf{X}_j,\mathbf{Z}_j)\big)+g\big((\mathbf{X}_j,\mathbf{Z}_j)\,,\,(\mathbf{X}_i,\mathbf{Z}_i)\big)\Big)/2$. We can now apply Lemma 18 to the sum of the $\tilde g_{i,j}$. With its notations, $\theta=\mathbb{E}\big[\tilde g_{i,j}\big]=0$. Moreover,

$$\begin{aligned}
&\Big|Var\Big[g\big((\mathbf{X}_i,\mathbf{Z}_i),(\mathbf{X}_j,\mathbf{Z}_j)\big)\Big]\Big|\\
&\leq \int\frac{K^2_h(\mathbf{z}_1-\mathbf{z})K^2_h(\mathbf{z}_2-\mathbf{z})}{f^4_{\mathbf{Z}}(\mathbf{z})}\Big(\mathbb{1}\big\{\mathbf{x}_1<\mathbf{x}_2\big\}-\mathbb{P}\big(\mathbf{X}_i<\mathbf{X}_j\big|\mathbf{Z}_i=\mathbf{Z}_j=\mathbf{z}\big)\Big)^2\\
&\times\ f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{z}_1)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{z}_2)\,d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{z}_1\,d\mathbf{z}_2\\
&\leq \int\frac{K^2(\mathbf{t}_1)K^2(\mathbf{t}_2)}{h^{2p}f^4_{\mathbf{Z}}(\mathbf{z})}f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{z}-h\mathbf{t}_1)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{z}-h\mathbf{t}_2)\,d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{t}_1\,d\mathbf{t}_2\\
&\leq h^{-2p}f^2_{\mathbf{Z},max}\Big(\int K^2\Big)^2/f^4_{\mathbf{Z},min},
\end{aligned}$$

and the same upper bound applies for $\tilde g_{i,j}$ (invoke Cauchy-Schwartz inequality). Here, we choose

$b=-a=2C_K^2 h^{-2p} f_{\mathbf{Z},min}^{-2}$. This yields

$$\mathbb{P}\Big(\frac{2}{n(n-1)}\sum_{1\leq i\neq j\leq n}\tilde{g}_{i,j}>t\Big)\leq \exp\bigg(-\frac{[n/2]t^2 f_{\mathbf{Z},min}^4}{2h^{-2p}f_{\mathbf{Z},max}^2(\int K^2)^2+(4/3)C_K^2 h^{-2p}f_{\mathbf{Z},min}^2 t}\bigg).$$

Then, for every $t>0$, we obtain

$$\begin{aligned}\mathbb{P}\Big(\sum_{1\leq i\neq j\leq n}\big(S_{i,j}-\mathbb{E}[S_{i,j}]-\Delta_n\big)\geq t\Big)&\leq \mathbb{P}\Big(\frac{1}{n^2}\sum_{1\leq i\neq j\leq n} g\big((\mathbf{X}_i,\mathbf{Z}_i)\,,\,(\mathbf{X}_j,\mathbf{Z}_j)\big)\geq t+\Delta_n\Big)\\ &\leq \mathbb{P}\Big(\frac{(n-1)}{n}\cdot\frac{2}{n(n-1)}\sum_{1\leq i\neq j\leq n}\tilde{g}_{i,j}\geq t+\Delta_n\Big)\\ &\leq \exp\bigg(-\frac{[n/2]t^2 f_{\mathbf{Z},min}^4}{2h^{-2p}f_{\mathbf{Z},max}^2(\int K^2)^2+(4/3)C_K^2 h^{-2p}f_{\mathbf{Z},min}^2 t}\bigg),\ \text{and}\end{aligned}$$

$$\begin{aligned}&\mathbb{P}\Big(\big|\sum_{1\leq i\neq j\leq n}\big(S_{i,j}-\mathbb{E}[S_{i,j}]\big)-\Delta_n\big|\geq t\Big)\\ &\qquad\leq 2\exp\bigg(-\frac{(n-1)t^2 f_{\mathbf{Z},min}^4}{4h^{-2p}f_{\mathbf{Z},max}^2(\int K^2)^2+(8/3)C_K^2 h^{-2p}f_{\mathbf{Z},min}^2 t}\bigg).\end{aligned}$$

The latter inequality and (12) conclude the proof. $\square$

A.3 Proof of Proposition 5

Let us note that $\tau_{1,2|\mathbf{Z}=\mathbf{z}}=\mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)\big|\mathbf{Z}_1=\mathbf{z},\mathbf{Z}_2=\mathbf{z}\big]$, and that the estimator given by (5) with the weights (4) can be rewritten as $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}:=U_n(g^*)\,/\,\{U_n(1)+\epsilon_n\}$ where

$$U_n(g):=\frac{1}{n(n-1)\mathbb{E}[K_h(\mathbf{z}-\mathbf{Z})]^2}\sum_{1\leq i\neq j\leq n} g(\mathbf{X}_i,\mathbf{X}_j)K_h(\mathbf{z}-\mathbf{Z}_i)K_h(\mathbf{z}-\mathbf{Z}_j),$$

for any measurable bounded function g, with the residual diagonal term

$$\epsilon_n=\frac{1}{n(n-1)\mathbb{E}[K_h(\mathbf{z}-\mathbf{Z})]^2}\sum_{i=1}^{n}K_h^2(\mathbf{z}-\mathbf{Z}_i).$$

By Bochner's lemma (see Bosq and Lecoutre [3]), ϵ_n is $O_P((nh^p)^{-1})$ and is negligible compared to $U_n(1)$. Note that

$$U_n(g^*) := \frac{1}{n(n-1)\mathbb{E}[K_h(\mathbf{z}-\mathbf{Z})]^2} \sum_{1\leq i\neq j\leq n} g^*(\mathbf{X}_i,\mathbf{X}_j)K_h(\mathbf{z}-\mathbf{Z}_i)K_h(\mathbf{z}-\mathbf{Z}_j) := \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} g^*_{i,j},$$

with obvious notations. In our case, its expectation is

$$\begin{aligned}
\mathbb{E}\big[U_n(g^*)\big] &:= \frac{1}{\mathbb{E}[K_h(\mathbf{z}-\mathbf{Z})]^2}\mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)K_h(\mathbf{z}-\mathbf{Z}_1)K_h(\mathbf{z}-\mathbf{Z}_2)\big] \\
&= \int g^*(\mathbf{x}_1,\mathbf{x}_2)K(\mathbf{t}_1)K(\mathbf{t}_2)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{z}+h\mathbf{t}_1)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{z}+h\mathbf{t}_2)d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{t}_1\,d\mathbf{t}_2 \\
&\to \frac{1}{f^2_{\mathbf{Z}}(\mathbf{z})}\int g^*(\mathbf{x}_1,\mathbf{x}_2)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{z})f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{z})d\mathbf{x}_1 d\mathbf{x}_2 = \mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)\big|\mathbf{Z}_1=\mathbf{z},\mathbf{Z}_2=\mathbf{z}\big],
\end{aligned}$$

applying Bochner's lemma to $\mathbf{z}\mapsto \int g^*(\mathbf{x}_1,\mathbf{x}_2)f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}(\mathbf{x}_1)f_{\mathbf{X}|\mathbf{Z}=\mathbf{z}}(\mathbf{x}_2)\,d\mathbf{x}_1\,d\mathbf{x}_2 = \tau_{1,2|\mathbf{Z}=\mathbf{z}}$, that is a continuous function.

Set $\theta_n := \mathbb{E}[g^*_{i,j}]$. Since g^* is symmetrical, the Hájek projection $\hat{U}_n(g^*)$ of $U_n(g^*)$ satisfies

$$\hat{U}_n(g^*) := \frac{2}{n}\sum_{j=1}^n \mathbb{E}[g^*_{0,j}|\mathbf{X}_j,\mathbf{Z}_j] - \theta_n =: \frac{2}{n}\sum_{j=1}^n g_n(\mathbf{X}_j,\mathbf{Z}_j) - \theta_n,$$

where $g_n(\mathbf{x},\tilde{\mathbf{z}}) := K_h(\mathbf{z}-\tilde{\mathbf{z}})\mathbb{E}\big[g^*(\mathbf{X},\mathbf{x})K_h(\mathbf{z}-\mathbf{Z})\big]\,/\,\mathbb{E}[K_h(\mathbf{z}-\mathbf{Z})]^2$. Note that $\mathbb{E}[\hat{U}_n(g^*)] = \theta_n$, $i\neq j$. Since $Var(\hat{U}_n(g^*)) = 4Var(\mathbb{E}[g^*_{i,j}|\mathbf{X}_j,\mathbf{Z}_j])/n = O((nh^p)^{-1})$, then $\hat{U}_n(g^*) = \theta_n + o_P(1) = \tau_{1,2|\mathbf{Z}_\mathbf{z}} + o_P(1)$. Moreover,

$$\begin{aligned}
U_n(g^*) - \hat{U}_n(g^*) &:= \frac{1}{n(n-1)}\sum_{1\leq i\neq j\leq n}\Big(g_{i,j} - \mathbb{E}[g^*_{i,j}|\mathbf{X}_j,\mathbf{Z}_j] - \mathbb{E}[g^*_{i,j}|\mathbf{X}_i,\mathbf{Z}_i] + \theta_n\Big) \\
&=: \frac{1}{n(n-1)}\sum_{1\leq i\neq j\leq n}\overline{g}_{i,j}.
\end{aligned}$$

By usual U-statistics calculations, it can be easily checked that

$$Var(U_n(g^*) - \hat{U}_n(g^*)) = \frac{1}{n^2(n-1)^2}\sum_{1\leq i_1\neq j_1\leq n}\sum_{1\leq i_2\neq j_2\leq n}\mathbb{E}[\overline{g}_{i_1,j_1}\overline{g}_{i_2,j_2}] = O\big(\frac{1}{n^2h^{2p}}\big). \quad (13)$$

Indeed, when all indices (i_1, i_2, j_1, j_2) are different, or when there is a single identity among them, $\mathbb{E}[\overline{g}_{i_1,j_1}\overline{g}_{i_2,j_2}] = 0$. The first nonzero terms arise when there are two identities, i.e. $i_1 = i_2$ and $j_1 = j_2$ (or $i_1 = j_2$ and $j_1 = i_2$). In the latter case, an upper bound as $O((nh^p)^{-2})$ when $f_{\mathbf{Z}}$ is continuous at $\mathbf{z}$, by usual change of variable techniques and Bochner's lemma. Then, $U_n(g^*) = \hat{U}_n(g^*) + o_P(1) = \tau_{1,2|\mathbf{Z_z}} + o_P(1)$. Note that $U_n(1) + \epsilon_n$ tends to one in probability (Bochner's lemma). As a consequence, $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}} = U_n(g^*) \,/\, (U_n(1) + \epsilon_n)$ tends to $\tau_{1,2|\mathbf{Z}=\mathbf{z}}/1$ by the continuous mapping theorem. $\square$

A.4 Proof of Proposition 6

We now study the joint behavior of $(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i})_{i=1,\dots,n'}$. We will follow Stute [16], applying the same ideas with a multivariate conditioning variable $\mathbf{z}$ and studying the joint distribution of U-statistics at several conditioning points. As in the proof of Proposition 5, the estimator (5) with the weights (4) can be rewritten as $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} := U_{n,i}(g^*) \,/\, (U_{n,i}(1) + \epsilon_{n,i})$, where

$$U_{n,i}(g) := \frac{1}{n(n-1)\mathbb{E}[K_h(\mathbf{Z}'_i - \mathbf{Z})]^2} \sum_{j_1,j_2=1, j_1\neq j_2}^{n} g(\mathbf{X}_{j_1}, \mathbf{X}_{j_2}) K_h(\mathbf{Z}'_i - \mathbf{Z}_{j_1}) K_h(\mathbf{Z}'_i - \mathbf{Z}_{j_2}),$$

for any bounded measurable function $g : \mathbb{R}^4 \to \mathbb{R}$. Moreover, $\sup_{i=1,\dots,n'} |\epsilon_{n,i}| = O_P(n^{-1}h^{-p})$. By a limited expansion of $f_{\mathbf{X},\mathbf{Z}}$ w.r.t. its second argument, and under Assumption 2.5, we easily check that $\mathbb{E}\big[U_{n,i}(g)\big] = \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i} + r_{n,i}$, where $|r_{n,i}| \leq C_0 h_n^{\alpha} / f^2_{\mathbf{Z}}(\mathbf{Z}'_i)$, for some constant C_0 that is independent of i.

Now, we prove the joint asymptotic normality of $\big(U_{n,i}(g)\big)_{i=1,\dots,n'}$. The Hájek projection $\hat{U}_{n,i}(g)$ of $U_{n,i}(g)$ satisfies $\hat{U}_{n,i}(g) := 2\sum_{j=1}^{n} g_{n,i}\big(\mathbf{X}_j, \mathbf{Z}_j\big)/n - \theta_n$, where $\theta_n := \mathbb{E}\big[U_{n,i}(g)\big]$ and

$$g_{n,i}(\mathbf{x}, \mathbf{z}) := K_h(\mathbf{Z}'_i - \mathbf{z}) \mathbb{E}\big[g(\mathbf{X}, \mathbf{x}) K_h(\mathbf{Z}'_i - \mathbf{Z})\big] \,/\, \mathbb{E}[K_h(\mathbf{Z}'_i - \mathbf{Z})]^2.$$

Lemma 22. *Under the assumptions of Proposition 6, for any measurable bounded function g,*

$$(nh^p)^{1/2}\Big(\hat{U}_{n,i}(g) - \mathbb{E}\big[U_{n,i}(g)\big]\Big)_{i=1,\dots,n'} \xrightarrow{D} \mathcal{N}(0, M_\infty(g)), \;\; \textit{as } n \to \infty,$$

where, for $1 \leq i, j \leq n'$,

$$[M_\infty(g)]_{i,j} := \frac{4\int K^2 \mathbb{1}_{\{\mathbf{Z}'_i = \mathbf{Z}'_j\}}}{f_{\mathbf{Z}}(\mathbf{Z}'_i)} \int g\big(\mathbf{x}_1, \mathbf{x}\big) g\big(\mathbf{x}_2, \mathbf{x}\big) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}'_i}(\mathbf{x}) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}'_i}(\mathbf{x}_1) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}'_i}(\mathbf{x}_2) d\mathbf{x}\, d\mathbf{x}_1\, d\mathbf{x}_2.$$

This lemma is proved in Section A.5.

Similarly as in the proof of Lemma 2.2 in Stute [16], for every $i = 1, \dots, n'$ and every bounded symmetrical measurable function g, we have $(nh^p)^{1/2} Var\big[\hat{U}_{n,i}(g) - U_{n,i}(g)\big] = o(1)$, which implies

$$(nh^p)^{1/2} \Big(U_{n,i}(g) - \mathbb{E}\big[U_{n,i}(g)\big] \Big)_{i=1,\dots,n'} \xrightarrow{D} \mathcal{N}(0, M_\infty(g)), \text{ as } n \to \infty. \tag{14}$$

Considering two measurable and bounded functions g_1 and g_2, we have $U_{n,i}(c_1 g_1 + c_2 g_2) = c_1 U_{n,i}(g_1) + c_2 U_{n,i}(g_2)$ for every real numbers c_1, c_2. By the Cramér-Wold device, we get

$$(nh^p)^{1/2} \bigg(\Big(U_{n,i}(g_1) - \mathbb{E}\big[U_{n,i}(g_1)\big] \Big)_{i=1,\dots,n'}, \Big(U_{n,i}(g_2) - \mathbb{E}\big[U_{n,i}(g_2)\big] \Big)_{i=1,\dots,n'} \bigg)$$
$$\xrightarrow{D} \mathcal{N}\left(0, \begin{bmatrix} M_\infty(g_1) & M_\infty(g_1, g_2) \\ M_\infty(g_1, g_2) & M_\infty(g_2) \end{bmatrix} \right),$$

as $n \to \infty$, where

$$[M_\infty(g_1, g_2)]_{i,j} := \frac{4\int K^2 \mathbb{1}_{\{\mathbf{Z}'_i = \mathbf{Z}'_j\}}}{f_{\mathbf{Z}}(\mathbf{Z}'_i)} \int g_1\big(\mathbf{x}_1, \mathbf{x}\big) g_2\big(\mathbf{x}_2, \mathbf{x}\big) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}'_i}(\mathbf{x}) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}'_i}(\mathbf{x}_1) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}'_i}(\mathbf{x}_2) d\mathbf{x}\, d\mathbf{x}_1\, d\mathbf{x}_2.$$

Set $\tilde{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} := U_{n,i}(g^*) \,/\, U_{n,i}(1)$. Since

$$(nh_n^p)^{1/2} \big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tilde{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} \big) = O_P\big((nh_n^p)^{1/2} \epsilon_{n,i} \big) = o_P(1),$$

it is sufficient to establish the asymptotic law of $(nh_n^p)^{1/2} \big(\tilde{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i} \big)$. Since $\mathbb{E}[U_{n,i}(1)] =$

$1 + o((nh^p)^{-1/2})$ and $\mathbb{E}[U_{n,i}(g^*)] = \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i} + o((nh_n^p)^{-1/2})$, we get

$$(nh^p)^{1/2}\bigg(\Big(U_{n,i}(g^*) - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\Big)_{i=1,\dots,n'}, \Big(U_{n,i}(1) - 1\Big)_{i=1,\dots,n'}\bigg)$$
$$\xrightarrow{D} \mathcal{N}\left(0, \begin{bmatrix} M_\infty(g^*) & M_\infty(g^*,1) \\ M_\infty(g^*,1) & M_\infty(1) \end{bmatrix}\right), \text{ as } n \to \infty.$$

Now apply the Delta-method with the function $\rho(\mathbf{x}, \mathbf{y}) := \mathbf{x}/\mathbf{y}$ where $\mathbf{x}$ and $\mathbf{y}$ are real-valued vectors of size n' and the division has to be understood component-wise. The Jacobian of ρ is given by the $n' \times 2n'$ matrix $J_\rho(\mathbf{x}, \mathbf{y}) = \Big[Diag\big(y_1^{-1}, \dots y_{n'}^{-1}\big)\, ,\, Diag\big(- x_1 y_1^{-2}, \dots - x_{n'} y_{n'}^{-2}\big)\Big]$, where for any vector v of size n', $Diag(v)$ is the diagonal matrix whose diagonal elements are the v_i, for $i = 1, \dots, n'$. We deduce $(nh^p)^{1/2}\left(\tilde\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\right)_{i=1,\dots,n'} \xrightarrow{D} \mathcal{N}(0, \mathbb{H})$, as $n \to \infty$, setting

$$\mathbb{H} := J_\rho(\vec\tau, \mathbf{e}) \begin{bmatrix} M_\infty(g^*) & M_\infty(g^*,1) \\ M_\infty(g^*,1) & M_\infty(1) \end{bmatrix} J_\rho(\vec\tau, \mathbf{e})^T,$$

where $\vec\tau = \left(\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\right)_{i=1,\dots,n'}$ and $\mathbf{e}$ is the vector of size n' whose all components are equal to 1. Thus, we have $J_\rho(\vec\tau, \mathbf{e}) = \Big[Id_{n'}, -Diag(\vec\tau)\Big]$, denoting by $Id_{n'}$ the identity matrix of size n' and by $Diag(\vec\tau)$ the diagonal matrix of size n' whose diagonal elements are the $\tau_{1,2|\mathbf{Z}'_i}$, for $i = 1, \dots, n'$. To be specific, we get

$$\mathbb{H} = M_\infty(g^*) - Diag(\vec\tau) M_\infty(g^*,1) - M_\infty(g^*,1) Diag(\vec\tau) + Diag(\vec\tau) M_\infty(1) Diag(\vec\tau).$$

For i, j in $\{1, \dots, n'\}$ and using the symmetry of the function g^*, we obtain

$$[M_\infty(g^*)]_{i,j} = \frac{4\int K^2 \mathbb{1}_{\{\mathbf{Z}'_i = \mathbf{Z}'_j\}}}{f_{\mathbf{Z}}(\mathbf{Z}'_i)} \mathbb{E}[g^*(\mathbf{X}_1, \mathbf{X}) g^*(\mathbf{X}_2, \mathbf{X}) | \mathbf{Z} = \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}'_i],$$

$$[Diag(\vec\tau) M_\infty(g^*,1)]_{i,j} = \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i} \frac{4\int K^2 \mathbb{1}_{\{\mathbf{Z}'_i = \mathbf{Z}'_j\}}}{f_{\mathbf{Z}}(\mathbf{Z}'_i)} \mathbb{E}[g^*(\mathbf{X}_1, \mathbf{X}) | \mathbf{Z} = \mathbf{Z}_1 = \mathbf{Z}'_i]$$
$$= \frac{4\int K^2 \mathbb{1}_{\{\mathbf{Z}'_i = \mathbf{Z}'_j\}}}{f_{\mathbf{Z}}(\mathbf{Z}'_i)} \tau^2_{1,2|\mathbf{Z}=\mathbf{Z}'_i} = [M_\infty(g^*,1) Diag(\vec\tau)]_{i,j} = [Diag(\vec\tau) M_\infty(1) Diag(\vec\tau)]_{i,j}.$$

As a consequence, we obtain

$$[\mathbb{H}]_{i,j} = \frac{4\int K^2 \mathbb{1}_{\{\mathbf{Z}'_i=\mathbf{Z}'_j\}}}{f_{\mathbf{Z}}(\mathbf{Z}'_i)}\Big(\mathbb{E}[g^*(\mathbf{X}_1,\mathbf{X})g^*(\mathbf{X}_2,\mathbf{X})|\mathbf{Z}=\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{Z}'_i] - \tau^2_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\Big). \;\square$$

A.5 Proof of Lemma 22

Let us evaluate the variance-covariance matrix $M_{n,n'} := [Cov(\hat{U}_{n,i},\hat{U}_{n,j})]_{1\leq i,j\leq n'}$. Note that

$$\mathbb{E}\Big[g_{n,i}\big(\mathbf{X}_j,\mathbf{Z}_j\big)\Big] = \mathbb{E}\big[\hat{U}_{n,i}\big] = \mathbb{E}\big[U_{n,i}(g)\big], \text{ and that}$$

$$\Big((nh^p)^{1/2}\Big(\hat{U}_{n,i} - \mathbb{E}\big[U_{n,i}(g)\big]\Big)\Big)_{i=1,\dots,n'} = \frac{2h^{p/2}}{n^{1/2}}\sum_{j=1}^n \Big(g_{n,i}\big(\mathbf{X}_j,\mathbf{Z}_j\big) - \mathbb{E}\big[U_{n,i}(g)\big]\Big)_{i=1,\dots,n'},$$

that is a sum of independent vectors. Thus, for every i,j in $\{1,\dots,n'\}$,

$$Cov(\hat{U}_{n,i},\hat{U}_{n,j}) = 4n^{-1}Cov\Big(g_{n,i}\big(\mathbf{X},\mathbf{Z}\big), g_{n,j}\big(\mathbf{X},\mathbf{Z}\big)\Big), \text{ and}$$

$$\begin{aligned}
&\mathbb{E}\Big[g_{n,i}\big(\mathbf{X},\mathbf{Z}\big), g_{n,j}\big(\mathbf{X},\mathbf{Z}\big)\Big] \\
&= \int K_h(\mathbf{Z}'_i - \mathbf{z})K_h(\mathbf{Z}'_j - \mathbf{z})\frac{\mathbb{E}\big[g(\mathbf{X},\mathbf{x})K_h(\mathbf{Z}'_i - \mathbf{Z})\big]\mathbb{E}\big[g(\mathbf{X},\mathbf{x})K_h(\mathbf{Z}'_j - \mathbf{Z})\big]}{\mathbb{E}[K_h(\mathbf{Z}'_i - \mathbf{Z})]^2\mathbb{E}[K_h(\mathbf{Z}'_j - \mathbf{Z})]^2} f_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z})d\mathbf{x}\,d\mathbf{z} \\
&\sim \frac{1}{h^p f^2_{\mathbf{Z}}(\mathbf{Z}'_i)f^2_{\mathbf{Z}}(\mathbf{Z}'_j)}\int g\big(\mathbf{x}_1,\mathbf{x}\big)g\big(\mathbf{x}_2,\mathbf{x}\big)K_h(\mathbf{Z}'_i - \mathbf{z})K_h(\mathbf{Z}'_j - \mathbf{z})K_h(\mathbf{Z}'_i - \mathbf{w}_1)K_h(\mathbf{Z}'_j - \mathbf{w}_2) \\
&\times\; f_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z})f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{w}_1)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{w}_2)d\mathbf{x}\,d\mathbf{z}\,d\mathbf{x}_1\,d\mathbf{w}_1\,d\mathbf{x}_2\,d\mathbf{w}_2 \\
&\sim \frac{1}{h^p f^2_{\mathbf{Z}}(\mathbf{Z}'_i)f^2_{\mathbf{Z}}(\mathbf{Z}'_j)}\int g\big(\mathbf{x}_1,\mathbf{x}\big)g\big(\mathbf{x}_2,\mathbf{x}\big)K(\mathbf{u}_1)K(\mathbf{u}_2)K(\mathbf{u})K(\frac{\mathbf{Z}'_j - \mathbf{Z}'_i}{h} + \mathbf{u})f_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{Z}'_i - h\mathbf{u}) \\
&\times\; f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{Z}'_i - h\mathbf{u}_1)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{Z}'_j - h\mathbf{u}_2)d\mathbf{x}\,d\mathbf{u}\,d\mathbf{x}_1\,d\mathbf{u}_1\,d\mathbf{x}_2\,d\mathbf{u}_2.
\end{aligned}$$

If $i \neq j$ and K is compactly supported, the latter term is zero when n is sufficiently large, and

$$Cov(\hat{U}_{n,i},\hat{U}_{n,j}) = -4n^{-1}\mathbb{E}[U_{n,i}]\mathbb{E}[U_{n,j}] \sim -4n^{-1}\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_j} = o\big((nh^p)^{-1}\big).$$

Otherwise, $i = j$ and we get

$$\mathbb{E}\Big[\Big(g_{n,i}\big(\mathbf{X},\mathbf{Z}\big)\Big)^2\Big] \sim \frac{1}{h^p f_{\mathbf{Z}}^4(\mathbf{Z}_i')} \int g\big(\mathbf{x}_1,\mathbf{x}\big) g\big(\mathbf{x}_2,\mathbf{x}\big) K(\mathbf{u}_1) K(\mathbf{u}_2) K^2(\mathbf{u}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{Z}_i')$$
$$\cdot f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{Z}_i') f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{Z}_i') d\mathbf{x}\, d\mathbf{u}\, d\mathbf{x}_1\, d\mathbf{u}_1\, d\mathbf{x}_2\, d\mathbf{u}_2$$
$$\sim \frac{\int K^2}{h^p f_{\mathbf{Z}}(\mathbf{Z}_i')} \cdot \int g\big(\mathbf{x}_1,\mathbf{x}\big) g\big(\mathbf{x}_2,\mathbf{x}\big) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}_i'}(\mathbf{x}) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}_i'}(\mathbf{x}_1) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}_i'}(\mathbf{x}_2) d\mathbf{x}\, d\mathbf{x}_1\, d\mathbf{x}_2,$$

by Bochner's lemma. As $\mathbb{E}\Big[g_{n,i}\big(\mathbf{X}_1,\mathbf{Z}_1\big)\Big] = O(1)$, we get

$$Var\Big[g_{n,i}\big(\mathbf{X}_j,\mathbf{Z}_j\big)\Big] \sim \frac{\int K^2}{h^p f_{\mathbf{Z}}(\mathbf{Z}_i')} \int g\big(\mathbf{x}_1,\mathbf{x}\big) g\big(\mathbf{x}_2,\mathbf{x}\big) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}_i'}(\mathbf{x}) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}_i'}(\mathbf{x}_1) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}_i'}(\mathbf{x}_2) d\mathbf{x}\, d\mathbf{x}_1\, d\mathbf{x}_2.$$

We have proved that, for every $i, j \in \{1, \dots, n'\}$,

$$nh^p [M_{n,n'}]_{i,j} \to \frac{4\int K^2 \mathbb{1}_{\{\mathbf{Z}_i'=\mathbf{Z}_j'\}}}{f_{\mathbf{Z}}(\mathbf{Z}_i')} \int g\big(\mathbf{x}_1,\mathbf{x}\big) g\big(\mathbf{x}_2,\mathbf{x}\big) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}_i'}(\mathbf{x}) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}_i'}(\mathbf{x}_1) f_{\mathbf{X}|\mathbf{Z}=\mathbf{Z}_i'}(\mathbf{x}_2)\, d\mathbf{x}\, d\mathbf{x}_1\, d\mathbf{x}_2,$$

as $n \to \infty$. Therefore, $nh^p M_{n,n'}$ tends to M_∞.

We now verify Lyapunov's condition with third moments, so that the usual multivariate central limit theorem would apply. It is then sufficient to show that

$$\Big(\frac{h^{p/2}}{n^{1/2}}\Big)^3 \sum_{j=1}^n \mathbb{E}\left[\Big| g_{n,i}\big(\mathbf{X}_j,\mathbf{Z}_j\big) - \mathbb{E}\big[U_{n,i}(g)\big]\Big|^3\right] = o(1). \tag{15}$$

For any $j = 1, \dots, n$, we have

$$\mathbb{E}\left[\Big| g_{n,i}\big(\mathbf{X}_j,\mathbf{Z}_j\big) - \mathbb{E}\big[U_{n,i}(g)\big]\Big|^3\right]$$
$$\sim \int \left| \frac{1}{f_{\mathbf{Z}}^2(\mathbf{Z}_i')} \int g(\mathbf{x}_1,\mathbf{x}) K_h(\mathbf{Z}_i' - \mathbf{z}_1) K_h(\mathbf{Z}_i' - \mathbf{z}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{z}_1) d\mathbf{x}_1\, d\mathbf{z}_1 - \mathbb{E}\big[U_{n,i}(g)\big] \right|^3 f_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z}) d\mathbf{x}\, d\mathbf{z}.$$

We do the change of variable $\mathbf{z}_1 = \mathbf{Z}'_i - h\mathbf{t}_1$ and $\mathbf{z} = \mathbf{Z}'_i - h\mathbf{t}$. We get

$$
\begin{aligned}
&\mathbb{E}\left[\left|g_{n,i}(\mathbf{X}_j, \mathbf{Z}_j) - \mathbb{E}\big[U_{n,i}(g)\big]\right|^3\right] \\
\sim\ & h^{-2p} \int \left| \frac{1}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)} \int g(\mathbf{x}_1, \mathbf{x}) K(\mathbf{t}_1) K(\mathbf{t}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1, \mathbf{Z}'_i - h\mathbf{t}_1) d\mathbf{x}_1\, d\mathbf{t}_1 \right. \\
-\ & \left. h^p \mathbb{E}\big[U_{n,i}(g)\big] \right|^3 f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{Z}'_i - h\mathbf{t}) d\mathbf{x}\, d\mathbf{t} = O(h^{-2p}),
\end{aligned}
$$

because of Bochner's lemma, under our assumptions. Therefore, we have obtained

$$
\Big(\frac{h^{p/2}}{n^{1/2}}\Big)^3 \sum_{j=1}^n \mathbb{E}\left[\left|g_{n,i}(\mathbf{X}_j, \mathbf{Z}_j) - \mathbb{E}\big[U_{n,i}(g)\big]\right|^3\right] = O(h^{3p/2} n^{-3/2} n h^{-2p}) = O((nh^p)^{-1/2}) = o(1).
$$

Therefore, we have checked Lyapunov's condition and the result follows. $\square$

B Non-asymptotic proofs for $\hat{\beta}$

In this section, we will use the notation $\mathbf{u} := \hat{\beta} - \beta^*$ and $\xi = [\xi_{i,n}]_{i=1,\dots,n'}$, $\xi_{i,n} = Y_i - (\mathbb{Z}'\beta)_i$.

B.1 Two technical lemmas

Lemma 23. *We have* $||\mathbb{Z}'\mathbf{u}||^2_{n'} \leq \lambda|\mathbf{u}|_1 + \dfrac{1}{n'} \langle \xi\ ,\ \mathbb{Z}'\mathbf{u}\rangle$.

Proof : As $\hat{\beta}$ is optimal, through the Karush-Kuhn-Tucker conditions, we have $(1/n')\mathbb{Z}'^T(\mathbf{Y} - \mathbb{Z}'\hat{\beta}) \in \partial\big(\lambda|\hat{\beta}|_1\big)$, where $\partial\big(\lambda|\hat{\beta}|_1\big)$ is the subdifferential of the norm $\lambda|\cdot|_1$ evaluated at $\hat{\beta}$. The dual norm of $|\cdot|_1$ is $|\cdot|_\infty$, so there exists $\mathbf{v}$ such that $|\mathbf{v}|_\infty \leq 1$ and $(1/n')\mathbb{Z}'^T(\mathbf{Y} - \mathbb{Z}'\hat{\beta}) + \lambda\mathbf{v} = 0$. We deduce successively $\mathbb{Z}'^T\mathbb{Z}'(\beta^* - \hat{\beta})/n' + \mathbb{Z}'^T\xi/n' + \lambda\mathbf{v} = 0$,

$$
\begin{gathered}
\frac{1}{n'}|\mathbb{Z}'(\beta^* - \hat{\beta})|_2^2 + \frac{1}{n'}(\beta^* - \hat{\beta})^T\mathbb{Z}'^T\xi + \lambda(\beta^* - \hat{\beta})^T\mathbf{v} = 0, \text{ and finally} \\
||\mathbb{Z}'(\beta^* - \hat{\beta})||^2_{n'} \leq \frac{1}{n'}\Big\langle \mathbb{Z}'(\hat{\beta} - \beta^*)\ ,\ \xi \Big\rangle + \lambda|\beta^* - \hat{\beta}|_1.\ \square
\end{gathered}
$$

Lemma 24. *We have* $|\mathbf{u}_{\mathcal{S}^c}|_1 \leq |\mathbf{u}_{\mathcal{S}}|_1 + \dfrac{2}{\lambda n'} \langle \xi\ ,\ \mathbb{Z}'\mathbf{u}\rangle$.

Proof : By definition, $\hat\beta$ is a minimizer of $||\mathbf{Y}-\mathbb{Z}'\beta||^2_{n'}+\lambda|\beta|_1$. Therefore, we have

$$||\mathbf{Y}-\mathbb{Z}'\hat\beta||^2_{n'}+\lambda|\hat\beta|_1 \leq ||\mathbf{Y}-\mathbb{Z}'\beta^*||^2_{n'}+\lambda|\beta^*|_1.$$

After some algebra, we derive $||Y-\mathbb{Z}'\hat\beta||^2_{n'}-||Y-\mathbb{Z}'\beta^*||^2_{n'} \leq \lambda\big(|(\beta^*-\hat\beta)_{\mathcal{S}}|_1-|(\hat\beta-\beta^*)_{\mathcal{S}^C}|_1\big)$. Moreover, the mapping $\beta\mapsto||\mathbf{Y}-\mathbb{Z}'\beta||^2_{n'}$ is convex and its gradient at β^* is $-2\,\mathbb{Z}'^T(\mathbf{Y}-\mathbb{Z}'\beta^*)/n'=-2\,\mathbb{Z}'^T\xi/n'$. So, we obtain

$$||\mathbf{Y}-\mathbb{Z}'\hat\beta||^2_{n'}-||\mathbf{Y}-\mathbb{Z}'\beta^*||^2_{n'} \geq \frac{-2}{n'}\Big\langle \mathbb{Z}'^T\xi\ ,\ \hat\beta-\beta^*\Big\rangle.$$

Finally, combining the two previous equations, we obtain

$$\frac{-2}{n'}\Big\langle \mathbb{Z}'^T\xi\ ,\ \hat\beta-\beta^*\Big\rangle \leq \lambda\big(|(\beta^*-\hat\beta)_{\mathcal{S}}|_1-|(\hat\beta-\beta^*)_{\mathcal{S}^C}|_1\big).\ \square$$

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, $||\mathbb{Z}'(\hat\beta-\beta^*)||_{n'}\leq\dfrac{4(\gamma+1)t\sqrt{s}}{\kappa(s,3)}$ and $|\hat\beta-\beta^*|_q\leq\dfrac{4^{2/q}(\gamma+1)ts^{1/q}}{\kappa^2(s,3)}$, for every $1\leq q\leq 2$.

Proof : Under the first assumption, we have the upper bound

$$\frac{1}{n'}\left\langle \mathbb{Z}'^T\xi\ ,\ \mathbf{u}\right\rangle \leq |\mathbf{u}|_1\max_{j=1,\dots,p'}\left|\frac{1}{n'}\sum_{i=1}^{n'}Z'_{i,j}\xi_{i,n}\right|\leq|\mathbf{u}|_1 t.$$

We first show that $\mathbf{u}$ belongs to the cone $\big\{\delta\in\mathbb{R}^{p'}:|\delta_{\mathcal{S}^C}|_1\leq 3|\delta_{\mathcal{S}}|_1, Card(\mathcal{S})\leq s\big\}$, so that we will be able to use the $RE(s,3)$ assumption with $J_0=\mathcal{S}$. From Lemma 24, $|\mathbf{u}_{\mathcal{S}^C}|_1\leq|\ u_{\mathcal{S}}|_1+2t|\mathbf{u}|_1/\lambda$. With our choice of λ, we deduce $|\mathbf{u}_{\mathcal{S}^C}|_1\leq|\mathbf{u}_{\mathcal{S}}|_1+2|\mathbf{u}|_1/\gamma$. Using the decomposition $|\mathbf{u}|_1=|\mathbf{u}_{\mathcal{S}^C}|_1+|\mathbf{u}_{\mathcal{S}}|_1$, we get $|\mathbf{u}_{\mathcal{S}^C}|_1\leq|\mathbf{u}_{\mathcal{S}}|_1(\gamma+2)/(\gamma-2)\leq 3|\mathbf{u}_{\mathcal{S}}|_1$.

As a consequence, we have $|\mathbf{u}|_1=|\mathbf{u}_{\mathcal{S}^C}|_1+|\mathbf{u}_{\mathcal{S}}|_1\leq 4|\mathbf{u}_{\mathcal{S}}|_1\leq 4\sqrt{s}|\mathbf{u}|_2\leq 4\sqrt{s}\,||\mathbb{Z}'\mathbf{u}||_{n'}/\kappa(s,3)$.

By Lemma 23,

$$||\mathbb{Z}'\mathbf{u}||_{n'}^2 \leq \lambda|\mathbf{u}|_1 + \frac{1}{n'}\langle \xi \ , \ \mathbb{Z}'\mathbf{u}\rangle \leq \lambda|\mathbf{u}|_1 + |\mathbf{u}|_1 t \leq |\mathbf{u}|_1(\gamma+1)t \leq \frac{4\sqrt{s}}{\kappa(s,3)}\,||\mathbb{Z}'\mathbf{u}||_{n'}(\gamma+1)t$$

We can now simplify and we get

$$||\mathbb{Z}'\mathbf{u}||_{n'} \leq \frac{4(\gamma+1)t}{\kappa(s,3)}\sqrt{s}, \quad |\mathbf{u}|_2 \leq \frac{4(\gamma+1)t}{\kappa^2(s,3)}\sqrt{s}, \text{ and } |\mathbf{u}|_1 \leq \frac{16(\gamma+1)t}{\kappa^2(s,3)}\, s.$$

Now, we compute a general bound for $|\mathbf{u}|_q$, with $1 \leq q \leq 2$, using the Hölder norm interpolation inequality:

$$|\mathbf{u}|_q \leq |\mathbf{u}|_1^{2/q-1}|\mathbf{u}|_2^{2-2/q} \leq \frac{4^{2/q}(\gamma+1)ts^{1/q}}{\kappa^2(s,3)}\cdot \quad \Box$$

B.2 Proof of Theorem 8

Using Proposition 3, for every $t_1, t_2 > 0$ such that $C_{K,\alpha}h^\alpha/\alpha! + t_1 < f_{\mathbf{Z},min}/2$, with probability greater than $1 - 2n' \exp\Big(-\big(nh^p t_1^2\big) \,/\, \big(2f_{\mathbf{Z},max}\int K^2 + (2/3)C_K t_1)\big)\Big) - 2n' \exp\Big(-(n-1)h^{2p}t_2^2 f_{\mathbf{Z},min}^4$ $/\, \big(4f_{\mathbf{Z},max}^2(\int K^2)^2 + (8/3)C_K^2 f_{\mathbf{Z},min}^2 t_2\big)\Big)$, we have

$$\begin{aligned}\max_{j=1,\dots,p'}\left|\frac{1}{n'}\sum_{i=1}^{n'} Z'_{i,j}\xi_{i,n}\right| &\leq C_\psi \max_{i=1,\dots,n'}\big|\xi_{i,n}\big| \leq C_\psi C_{\Lambda'} \max_{i=1,\dots,n'}\big|\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\big| \\ &\leq 4C_\psi C_{\Lambda'}\bigg(1 + \frac{16f_{\mathbf{Z},max}^2}{f_{\mathbf{Z},min}^3}\Big(\frac{C_{K,\alpha}h^\alpha}{\alpha!} + t_1\Big)\bigg)\bigg(\frac{C_{\mathbf{XZ},\alpha}h^\alpha}{f_{\mathbf{Z},min}^2\alpha!} + t_2\bigg).\end{aligned}$$

We choose $t_1 := f_{\mathbf{Z},min}/2$ so that, because of Condition (7), we get

$$C_{K,\alpha}h^\alpha/\alpha! + t_1 \leq f_{\mathbf{Z},min}.$$

Now we choose $t_2 := t f^2_{\mathbf{Z},min}/\{8 C_\psi C_{\Lambda'} (f^2_{\mathbf{Z},min} + 16 f^2_{\mathbf{Z},max})\}$. By Condition (8), $C_{\mathbf{XZ},\alpha} h^\alpha / (f^2_{\mathbf{Z},min} \alpha!) \leq t_2$, so that we have

$$4C_\psi C_{\Lambda'} \left(1 + \frac{16 f^2_{\mathbf{Z},max}}{f^2_{\mathbf{Z},min}}\right) \times \left(\frac{C_{\mathbf{XZ},\alpha} h^\alpha}{f^2_{\mathbf{Z},min} \alpha!} + t_2\right) \leq 8 t_2 C_\psi C_{\Lambda'} \left(1 + \frac{16 f^2_{\mathbf{Z},max}}{f^2_{\mathbf{Z},min}}\right) \leq t.$$

As a consequence, we obtain that

$$\begin{aligned}
\mathbb{P}\left(\max_{j=1,\dots,p'} \left| \frac{1}{n'} \sum_{i=1}^{n'} Z'_{i,j} \xi_{i,n} \right| > t \right) \leq 2n' \exp\left(- \frac{n h^p f^2_{\mathbf{Z},min}}{8 f_{\mathbf{Z},max} \int K^2 + (4/3) C_K f_{\mathbf{Z},min}} \right) \\
+ 2n' \exp\left(- \frac{(n-1) h^{2p} t^2}{C_2 + C_3 t} \right),
\end{aligned}$$

and we can apply Lemma 25 to get the claimed result. □

C Asymptotic proofs in n for fixed n'

C.1 Proof of Lemma 10

Using the definition (2) of $\hat{\beta}_{n,n'}$, we get

$$\begin{aligned}
\hat{\beta}_{n,n'} &:= \arg \min_{\beta \in \mathbb{R}^{p'}} \left[\frac{1}{n'} \sum_{i=1}^{n'} \left(\Lambda(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i}) - \boldsymbol{\psi}(\mathbf{Z}'_i)^T \beta \right)^2 + \lambda_{n,n'} |\beta|_1 \right] \\
&= \arg \min_{\beta \in \mathbb{R}^{p'}} \left[\frac{1}{n'} \sum_{i=1}^{n'} \left(\xi_{i,n} + \boldsymbol{\psi}(\mathbf{Z}'_i)^T \beta^* - \boldsymbol{\psi}(\mathbf{Z}'_i)^T \beta \right)^2 + \lambda_{n,n'} |\beta|_1 \right] \\
&= \arg \min_{\beta \in \mathbb{R}^{p'}} \left[\frac{1}{n'} \sum_{i=1}^{n'} \xi^2_{i,n} + \frac{2}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol{\psi}(\mathbf{Z}'_i)^T (\beta^* - \beta) + \frac{1}{n'} \sum_{i=1}^{n'} \left(\boldsymbol{\psi}(\mathbf{Z}'_i)^T (\beta^* - \beta) \right)^2 + \lambda_{n,n'} |\beta|_1 \right] \\
&= \arg \min_{\beta \in \mathbb{R}^{p'}} \left[\frac{2}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol{\psi}(\mathbf{Z}'_i)^T (\beta^* - \beta) + \frac{1}{n'} \sum_{i=1}^{n'} \left(\boldsymbol{\psi}(\mathbf{Z}'_i)^T (\beta^* - \beta) \right)^2 + \lambda_{n,n'} |\beta|_1 \right] \quad \square
\end{aligned}$$

C.2 Proof of Theorem 13

Let us define $r_{n,n'} := (nh^p_{n,n'})^{1/2}$, $\mathbf{u} := r_{n,n'}(\beta - \beta^*)$ and $\hat{\mathbf{u}}_{n,n'} := r_{n,n'}(\hat{\beta}_{n,n'} - \beta^*)$, so that $\hat{\beta}_{n,n'} = \beta^* + \hat{\mathbf{u}}_{n,n'}/r_{n,n'}$. By Lemma 10, $\hat{\beta}_{n,n'} = \arg\min_{\beta \in \mathbb{R}^{p'}} \mathbb{G}_{n,n'}(\beta)$. We have therefore

$$\begin{aligned}
\hat{\mathbf{u}}_{n,n'} &= \arg\min_{\mathbf{u}\in\mathbb{R}^{p'}} \left[\frac{-2}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol{\psi}(\mathbf{Z}'_i)^T \frac{\mathbf{u}}{r_{n,n'}} + \frac{1}{n'} \sum_{i=1}^{n'} \left(\boldsymbol{\psi}(\mathbf{Z}'_i)^T \frac{\mathbf{u}}{r_{n,n'}} \right)^2 + \lambda_{n,n'} \left| \beta^* + \frac{\mathbf{u}}{r_{n,n'}} \right|_1 \right] \\
&= \arg\min_{\mathbf{u}\in\mathbb{R}^{p'}} \mathbb{F}_{n,n'}(\mathbf{u}),
\end{aligned}$$

where, for every $\mathbf{u} \in \mathbb{R}^{p'}$,

$$\mathbb{F}_{n,n'}(\mathbf{u}) := \frac{-2r_{n,n'}}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol{\psi}(\mathbf{Z}'_i)^T \mathbf{u} + \frac{1}{n'} \sum_{i=1}^{n'} \left(\boldsymbol{\psi}(\mathbf{Z}'_i)^T \mathbf{u} \right)^2 + \lambda_{n,n'} r^2_{n,n'} \Big(\big|\beta^* + \frac{\mathbf{u}}{r_{n,n'}}\big|_1 - \big|\beta^*\big|_1 \Big).$$

Note that, by Corollary 12, we have

$$\frac{2r_{n,n'}}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol{\psi}(\mathbf{Z}'_i)^T \mathbf{u} = \frac{2}{n'} \sum_{i=1}^{n'} \sum_{j=1}^{p'} r_{n,n'} \xi_{i,n} \psi_j(\mathbf{Z}'_i) u_j \xrightarrow{D} \frac{2}{n'} \sum_{i=1}^{n'} \sum_{j=1}^{p'} W_i \psi_j(\mathbf{Z}'_i) u_j.$$

We also have, for any (fixed) $\mathbf{u}$ and when n is large enough,

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

Therefore $\lambda_{n,n'} r^2_{n,n'} \Big(\big|\beta^* + \frac{\mathbf{u}}{r_{n,n'}}\big|_1 - \big|\beta^*\big|_1 \Big) \to \ell \sum_{i=1}^{p'} \big(|u_i| \mathbb{1}_{\{\beta^*_i = 0\}} + u_i \text{ sign}(\beta^*_i) \mathbb{1}_{\{\beta^*_i \neq 0\}} \big)$.

We have shown that $\mathbb{F}_{n,n'}(\mathbf{u}) \xrightarrow{D} \mathbb{F}_{\infty,n'}(\mathbf{u})$. Those functions are convex, hence the conclusion follows from the convexity argument. $\square$

C.3 Proof of Proposition 14

The proof closely follows Proposition 1 in Zou [19]. It starts by noting that

$$\mathbb{P}\left(\mathcal{S}_n = \mathcal{S}\right) \leq \mathbb{P}\left(\hat{\beta}_j = 0, \ \forall j \notin \mathcal{S}\right).$$

Because of the weak limit of $\hat{\beta}$ (Theorem 13 and the notations therein), this implies

$$\limsup_n \mathbb{P}\left(\hat{\beta}_j = 0,\ \forall j \notin \mathcal{S}\right) \leq \mathbb{P}\left(u_j^* = 0,\ \forall j \notin \mathcal{S}\right).$$

If $\ell = 0$, then $\mathbf{u}^*$ is asymptotically normal, and the latter probability is zero. Otherwise $\ell \neq 0$, and define $\vec{W}_{\boldsymbol{\psi}} := 2\sum_{i=1}^{n'} W_i \boldsymbol{\psi}(\mathbf{Z}_i')/n'$. The KKT conditions applied to $\mathbb{F}_{\infty,n'}$ provide

$$\vec{W}_{\boldsymbol{\psi}} + \frac{2}{n'}\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{Z}_i')\boldsymbol{\psi}(\mathbf{Z}_i')^T \mathbf{u}^* + \ell \mathbf{v}^* = 0,$$

for some vector $\mathbf{v}^* \in \mathbb{R}^p$ whose components v_j^* are less than one in absolute value when $j \notin \mathcal{S}$, and $v_j^* =$ sign(β_j^*) when $j \in \mathcal{S}$. If $u_j^* = 0$ for all $j \notin \mathcal{S}$, we deduce

$$(\vec{W}_{\boldsymbol{\psi}})_{\mathcal{S}} + \left[\frac{2}{n'}\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{Z}_i')\boldsymbol{\psi}(\mathbf{Z}_i')^T\right]_{\mathcal{S},\mathcal{S}} \mathbf{u}_{\mathcal{S}}^* + \ell\ \mathrm{sign}(\beta_{\mathcal{S}}^*) = 0, \text{ and} \tag{16}$$

$$\left|(\vec{W}_{\boldsymbol{\psi}})_{\mathcal{S}^c} + \left[\frac{2}{n'}\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{Z}_i')\boldsymbol{\psi}(\mathbf{Z}_i')^T\right]_{\mathcal{S}^c,\mathcal{S}} \mathbf{u}_{\mathcal{S}}^*\right| \leq \ell, \tag{17}$$

componentwise and with obvious notations. Combining the two latter equations provides

$$\left|(\vec{W}_{\boldsymbol{\psi}})_{\mathcal{S}^c} - \left[\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{Z}_i')\boldsymbol{\psi}(\mathbf{Z}_i')^T\right]_{\mathcal{S}^c,\mathcal{S}} \left[\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{Z}_i')\boldsymbol{\psi}(\mathbf{Z}_i')^T\right]_{\mathcal{S},\mathcal{S}}^{-1} \left((\vec{W}_{\boldsymbol{\psi}})_{\mathcal{S}} + \ell\ \mathrm{sign}(\beta_{\mathcal{S}}^*)\right)\right| \leq \ell, \tag{18}$$

componentwise. Since the latter event is of probability strictly lower than one, this is still the case for the event $\left\{u_j^* = 0,\ \forall j \notin \mathcal{S}\right\}$. $\square$

C.4 Proof of Theorem 15

The beginning of the proof is similar to the proof of Theorem 13. With obvious notations, $\check{\mathbf{u}}_{n,n'} = \arg\min_{\mathbf{u}\in\mathbb{R}^{p'}} \check{\mathbb{F}}_{n,n'}(\mathbf{u})$, where for every $\mathbf{u} \in \mathbb{R}^{p'}$,

$$\check{\mathbb{F}}_{n,n'}(\mathbf{u}) := \frac{-2r_{n,n'}}{n'}\sum_{i=1}^{n'} \xi_{i,n}\boldsymbol{\psi}(\mathbf{Z}_i')^T\mathbf{u} + \frac{1}{n'}\sum_{i=1}^{n'}\left(\boldsymbol{\psi}(\mathbf{Z}_i')^T\mathbf{u}\right)^2 + \mu_{n,n'}r_{n,n'}^2 \sum_{i=1}^{p'} \frac{1}{|\tilde{\beta}_i|^\delta}\left(|\beta_i^* + \frac{u_i}{r_{n,n'}}| - |\beta_i^*|\right).$$

If $\beta_i^* \neq 0$, then

$$\frac{\mu_{n,n'} r_{n,n'}^2}{|\tilde{\beta}_i|^\delta}\Big(|\beta_i^* + \frac{u_i}{r_{n,n'}}| - |\beta_i^*|\Big) = \frac{\mu_{n,n'} r_{n,n'}}{|\tilde{\beta}_i|^\delta} u_i \ \text{sign}(\beta_i^*) = \frac{\ell}{|\beta_i^*|^\delta} u_i \ \text{sign}(\beta_i^*) + o_P(1).$$

If $\beta_i^* = 0$, then

$$\frac{\mu_{n,n'} r_{n,n'}^2}{|\tilde{\beta}_i|^\delta}\Big(|\beta_i^* + \frac{u_i}{r_{n,n'}}| - |\beta_i^*|\Big) = \frac{\mu_{n,n'} r_{n,n'} \nu_n^\delta}{|\nu_n \tilde{\beta}_i|^\delta} |u_i|.$$

By assumption $\nu_n \tilde{\beta}_i = O_p(1)$, and the latter term tends to the infinity in probability iff $u_i \neq 0$. As a consequence, if there exists some $i \notin \mathcal{S}$ s.t. $u_i \neq 0$, then $\check{\mathbb{F}}_{n,n'}(\mathbf{u})$ tends to the infinity. Otherwise, $u_i = 0$ when $i \notin \mathcal{S}$ and $\check{\mathbb{F}}_{n,n'}(\mathbf{u}) \to \check{\mathbb{F}}_{\infty,n'}(\mathbf{u}_{\mathcal{S}})$. Since $\check{\mathbb{F}}_{\infty,n'}$ is convex, we deduce (Kato [10]) that $\check{\mathbf{u}}_{\mathcal{S}} \to \mathbf{u}^*_{\mathcal{S}}$, and $\check{\mathbf{u}}_{\mathcal{S}^c} \to 0_{\mathcal{S}^c}$, proving the asymptotic normality of $\check{\beta}_{n,n',\mathcal{S}}$.

Now, let us prove the oracle property. If $j \in \mathcal{S}$, then $\check{\beta}_j$ tends to β_j in probability and $\mathbb{P}(j \in \mathcal{S}_n) \to 1$. It suffices to show that $\mathbb{P}(j \in \mathcal{S}_n) \to 0$ when $j \notin \mathcal{S}$. If $j \notin \mathcal{S}$ and $j \in \mathcal{S}_n$, the KKT conditions on $\check{\mathbb{F}}_{n,n'}$ provide

$$\frac{-2r_{n,n'}}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol{\psi}(\mathbf{Z}'_i)_j + \frac{2}{n'} \sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{Z}'_i)_j \boldsymbol{\psi}(\mathbf{Z}'_i)^T \check{\mathbf{u}}_{n,n'} = -\frac{\mu_{n,n'} r_{n,n'} \nu_n^\delta}{|\nu_n \tilde{\beta}_j|^\delta} sign(\check{u}_j).$$

Due to the asymptotic normality of $\check{\beta}$ (that implies the one of $\check{\mathbf{u}}_{n,n'}$), the left hand side of the previous equation is asymptotically normal, when $\ell = 0$. On the other side, the r.h.s. tends to the infinity in probability because $\nu_n \tilde{\beta}_j = O_P(1)$. Therefore, the probability of the latter event tends to zero when $n \to \infty$. $\square$

D Proofs in the double asymptotic framework

D.1 Proof of Theorem 16

By Lemma 10, we have $\hat{\beta}_{n,n'} = \arg\min_{\beta \in \mathbb{R}^{p'}} \mathbb{G}_{n,n'}(\beta)$, where

$$\mathbb{G}_{n,n'}(\beta) := \frac{2}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol{\psi}(\mathbf{Z}'_i)^T (\beta^* - \beta) + \frac{1}{n'} \sum_{i=1}^{n'} \big(\boldsymbol{\psi}(\mathbf{Z}'_i)^T (\beta^* - \beta)\big)^2 + \lambda_{n,n'} |\beta|_1.$$

Define also $\mathbb{G}_{\infty,n'}(\beta) := \sum_{i=1}^{n'} \big(\boldsymbol{\psi}(\mathbf{Z}'_i)^T(\beta^* - \beta)\big)^2/n' + \lambda_0|\beta|_1$. We have

$$\big|\mathbb{G}_{n,n'}(\beta) - \mathbb{G}_{\infty,n'}(\beta)\big| \leq \left|\frac{2}{n'}\sum_{i=1}^{n'} \xi_{i,n}\boldsymbol{\psi}(\mathbf{Z}'_i)^T(\beta^* - \beta)\right| + |\lambda_{n,n'} - \lambda_0| \cdot |\beta|_1.$$

By assumption, the second term on the r.h.s. converges to 0. We now show that the first term on the r.h.s. is negligible. Indeed, for every $\epsilon > 0$,

$$\mathbb{P}\left(\Big\|\frac{1}{n'}\sum_{i=1}^{n'} \xi_{i,n}\boldsymbol{\psi}(\mathbf{Z}'_i)\Big\| > \epsilon\right) \leq \mathbb{P}\left(\frac{\|C_{\Lambda'}\|}{n'}\sum_{i=1}^{n'} |\hat{\tau}_{\mathbf{Z}'_i} - \tau_{\mathbf{Z}'_i}| \times \|\boldsymbol{\psi}(\mathbf{Z}'_i)\| > \epsilon\right)$$
$$\leq \textstyle\sum_{i=1}^{n'} \mathbb{P}\left(|\hat{\tau}_{\mathbf{Z}'_i} - \tau_{\mathbf{Z}'_i}| > C_2\epsilon\right),$$

where C_2 is the constant $(\|C_{\Lambda'}\| \times \|C_\psi\|)^{-1}$.

Apply Proposition 3 with the $t = t' = 1$, and get

$$\mathbb{P}\Big(|\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}} - \tau_{1,2|\mathbf{Z}=\mathbf{z}}| > \epsilon\Big) \leq 4\exp\Big(-Cst \cdot nh^{2p}\Big),$$

for some constant $Cst > 0$. Thus, $\sum_{i=1}^{n'} \xi_{i,n}\boldsymbol{\psi}(\mathbf{Z}'_i)/n' = o_{\mathbb{P}}(1)$, and $\mathbb{G}_{n,n'}(\beta) = \mathbb{G}_{\infty,n'}(\beta) + o_{\mathbb{P}}(1)$ for every β.

Since $\sum_{i=1}^{n'} \psi(\mathbf{Z}'_i)\psi(\mathbf{Z}'_i)^T/n'$ tends towards a matrix $M_{\psi,\mathbf{Z}'}$, deduce that $\mathbb{G}_{\infty,n'}(\beta)$ tends to $\mathbb{G}_{\infty,\infty}(\beta)$ when $n' \to \infty$. Therefore, for all $\beta \in \mathbb{R}^{p'}$, $\mathbb{G}_{n,n'}(\beta)$ weakly tends to $\mathbb{G}_{\infty,\infty}(\beta)$. By the convexity argument, we deduce that $\arg\min_\beta \mathbb{G}_{n,n'}(\beta)$ weakly converges to $\arg\min_\beta \mathbb{G}_{\infty,\infty}(\beta)$. Since the latter minimizer is non random, the same convergence is true in probability. $\Box$

D.2 Proof of Theorem 17

We start as in the proof of Theorem 13. Define $\tilde{r}_{n,n'} := (nn'h^p_{n,n'})^{1/2}$, $\mathbf{u} := \tilde{r}_{n,n'}(\beta - \beta^*)$ 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 $\mathbf{u} \in \mathbb{R}^{p'}$,

$$\mathbb{F}_{n,n'}(\mathbf{u}) := \frac{-2\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'} \xi_{i,n}\boldsymbol{\psi}(\mathbf{Z}'_i)^T\mathbf{u} + \frac{1}{n'}\sum_{i=1}^{n'}\big(\boldsymbol{\psi}(\mathbf{Z}'_i)^T\mathbf{u}\big)^2 + \lambda_{n,n'}\tilde{r}^2_{n,n'}\Big(\big|\beta^* + \frac{\mathbf{u}}{\tilde{r}_{n,n'}}\big|_1 - \big|\beta^*\big|_1\Big), \quad (19)$$

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

Lemma 26. *Under the same assumptions as in Theorem 17,* $T_1 := (\tilde{r}_{n,n'}/n') \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol{\psi}(\mathbf{Z}'_i)$ *tends in law towards a Gaussian random vector* $\mathcal{N}(0, V_2)$.

This lemma is proved in Section D.3. It yields the weak limit of the random vector T_1, and will help to control the first term in Equation (19), which is simply $-2T_1^T \mathbf{u}$.

Using Assumption 2.8(iii) and for every $\mathbf{u} \in \mathbb{R}^{p'}$, we get

$$\frac{1}{n'} \sum_{i=1}^{n'} \left(\boldsymbol{\psi}(\mathbf{Z}'_i)^T \mathbf{u} \right)^2 \to \int_{\mathbf{z}'} \left(\boldsymbol{\psi}(\mathbf{z}')^T \mathbf{u} \right)^2 f_{\mathbf{Z}',\infty} d\mathbf{z}'. \tag{20}$$

Note that this is 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,

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

Therefore, by Assumption 2.8(ii)(b), for every $\mathbf{u} \in \mathbb{R}^{p'}$,

$$\lambda_{n,n'} \tilde{r}_{n,n'}^2 \left(\left|\beta^* + \frac{\mathbf{u}}{\tilde{r}_{n,n'}}\right|_1 - \left|\beta^*\right|_1 \right) \to 0. \tag{21}$$

Combining Lemma 26 and Equations (19-21), and defining the function $\mathbb{F}_{\infty,\infty}$ by

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

where $\tilde{W} \sim \mathcal{N}(0, V_2)$, we get that every finite-dimensional margin of $\mathbb{F}_{n,n'}$ converges weakly to the corresponding margin of $\mathbb{F}_{\infty,\infty}$, defined by

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

where $\tilde{W} \sim \mathcal{N}(0, V_2)$. Now, applying the convexity lemma, we get

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

Since $\mathbb{F}_{\infty,\infty}(\mathbf{u})$ is a continuously differentiable convex function, we apply the first-order condition $\nabla \mathbb{F}_{\infty,\infty}(\mathbf{u}) = 0$, which yields $2\tilde{\mathbf{W}} + 2 \int \boldsymbol{\psi}(\mathbf{z}')\boldsymbol{\psi}(\mathbf{z}')^T \mathbf{u}_{\infty,\infty} f_{\mathbf{Z}',\infty}(\mathbf{z}') d\mathbf{z}' = 0$. As a consequence $\mathbf{u}_{\infty,\infty} = -V_1^{-1}\tilde{\mathbf{W}} \sim \mathcal{N}(0, \tilde{V}_{as})$, using Assumption 2.8(iv). We finally obtain $\tilde{r}_{n,n'}\big(\hat{\beta}_{n,n'} - \beta^*\big) \xrightarrow{D} \mathcal{N}\big(0, \tilde{V}_{as}\big)$, as claimed. $\square$

D.3 Proof of Lemma 26 : convergence of T_1

Using a Taylor expansion of order 1, we have

$$T_1 := \frac{\tilde{r}_{n,n'}}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol{\psi}(\mathbf{Z}'_i) = \frac{\tilde{r}_{n,n'}}{n'} \sum_{i=1}^{n'} \Big(\Lambda\big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\big) - \Lambda\big(\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\big) \Big) \boldsymbol{\psi}(\mathbf{Z}'_i) = T_2 + T_3,$$

where the main term is

$$T_2 := \frac{\tilde{r}_{n,n'}}{n'} \sum_{i=1}^{n'} \Lambda'\big(\tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\big) \big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\big) \boldsymbol{\psi}(\mathbf{Z}'_i),$$

and the remainder is

$$T_3 := \frac{\tilde{r}_{n,n'}}{n'} \sum_{i=1}^{n'} \alpha_{3,i} \big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i}\big)^2 \boldsymbol{\psi}(\mathbf{Z}'_i), \tag{22}$$

with $\forall i = 1, \dots, n'$, $|\alpha_{3,i}| \leq C_{\Lambda''}/2$, by Assumption 2.8(v).

Using the definition (5) of $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$, the definition (4) of the weights $w_{i,n}(\mathbf{z})$ and the notation $\overline{\boldsymbol{\psi}}(\mathbf{z}) := \Lambda'\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\big)\boldsymbol{\psi}(\mathbf{z})$, rewrite $T_2 =: T_4 + T_5$, where

$$\begin{aligned} T_4 := \frac{\tilde{r}_{n,n'}}{n'n^2} \sum_{i=1}^{n'} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} & \frac{K_h(\mathbf{Z}'_i - \mathbf{Z}_{j_1}) K_h(\mathbf{Z}'_i - \mathbf{Z}_{j_2})}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)} \\ & \times \Big(g^*(\mathbf{X}_{j_1}, \mathbf{X}_{j_2}) - \mathbb{E}\big[g^*(\mathbf{X}_{j_1}, \mathbf{X}_{j_2}) | \mathbf{Z}_{j_1} = \mathbf{Z}_{j_2} = \mathbf{Z}'_i\big] \Big) \overline{\boldsymbol{\psi}}(\mathbf{Z}'_i) \end{aligned} \tag{23}$$

$$T_5 := \frac{\tilde r_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_1})K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_2})\bigg(\frac{1}{\hat f_{\mathbf{Z}}(\mathbf{Z}'_i)^2}-\frac{1}{f_{\mathbf{Z}}(\mathbf{Z}'_i)^2}\bigg)$$
$$\times\;\Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2}) - \mathbb{E}\big[g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2})|\mathbf{Z}_{j_1}=\mathbf{Z}_{j_2}=\mathbf{Z}'_i\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i). \tag{24}$$

The random variable T_4 can be seen (see Equation (23)) as a sum of (indexed by i) U-statistics of order 2. Through its Hájek projection, we decompose it into a main term $T_{4,1}$, a remainder term $T_{4,2}$ and the bias $T_{4,3}$: $T_4 = T_{4,1}+T_{4,2}+T_{4,3}$, where

$$T_{4,1} := \frac{\tilde r_{n,n'}}{n'n}\sum_{i=1}^{n'}\sum_{j=1}^{n}\Bigg(g_{n,i}\big(\mathbf{X}_j,\mathbf{Z}_j\big) - \mathbb{E}\bigg[\frac{K_h(\mathbf{Z}'_i-\mathbf{Z}_1)K_h(\mathbf{Z}'_i-\mathbf{Z}_2)}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}g^*(\mathbf{X}_1,\mathbf{X}_2)\bigg]\Bigg)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i),$$
$$T_{4,2} := \frac{\tilde r_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\bigg(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2})\frac{K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_1})K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_2})}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)} - g_{n,i}\big(\mathbf{X}_{j_1},\mathbf{Z}_{j_1}\big)\bigg)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i),$$
$$T_{4,3} := \frac{\tilde r_{n,n'}}{n'}\sum_{i=1}^{n'}\Bigg(\mathbb{E}\bigg[\frac{K_h(\mathbf{Z}'_i-\mathbf{Z}_1)K_h(\mathbf{Z}'_i-\mathbf{Z}_2)}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}g^*(\mathbf{X}_1,\mathbf{X}_2)\bigg] - \mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{Z}'_i\big]\Bigg)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i),$$

and $g_{n,i}\big(\mathbf{x},\mathbf{z}\big) := \mathbb{E}\big[g^*(\mathbf{X},\mathbf{x})K_h(\mathbf{Z}'_i-\mathbf{Z})K_h(\mathbf{Z}'_i-\mathbf{z})\big]\,/\,f^2_{\mathbf{Z}}(\mathbf{Z}'_i)$. We easily check that $g_{n,i}\big(\mathbf{x},\mathbf{z}\big) =: g_{n,i,1}\big(\mathbf{x},\mathbf{z}\big)+g_{n,i,2}\big(\mathbf{x},\mathbf{z}\big)$, where

$$g_{n,i,1}\big(\mathbf{x},\mathbf{z}\big) = \int g^*(\mathbf{x}_1,\mathbf{x})\frac{K_h(\mathbf{Z}'_i-\mathbf{z})}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{Z}'_i)d\mathbf{x}_1, \text{ and} \tag{25}$$
$$g_{n,i,2}\big(\mathbf{x},\mathbf{z}\big) = \int g^*(\mathbf{x}_1,\mathbf{x})\frac{K(\mathbf{t})K_h(\mathbf{Z}'_i-\mathbf{z})}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}\Big(f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{Z}'_i+h\mathbf{t}) - f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{Z}'_i)\Big)d\mathbf{x}_1 d\mathbf{t}. \tag{26}$$

Note that

$$\mathbb{E}\bigg[\frac{K_h(\mathbf{Z}'_i-\mathbf{Z}_1)K_h(\mathbf{Z}'_i-\mathbf{Z}_2)}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}g^*(\mathbf{X}_1,\mathbf{X}_2)\bigg] = \mathbb{E}\Big[g_{n,i}\big(\mathbf{X}_j,\mathbf{Z}_j\big)\Big] = \mathbb{E}\Big[g_{n,i,1}\big(\mathbf{X}_j,\mathbf{Z}_j\big)+g_{n,i,2}\big(\mathbf{X}_j,\mathbf{Z}_j\big)\Big],$$

implying $T_{4,1} = T_{4,1,1}+T_{4,1,2}$, by setting

$$T_{4,1,k} := \frac{\tilde r_{n,n'}}{n'n}\sum_{i=1}^{n'}\sum_{j=1}^{n}\Big(g_{n,i,k}\big(\mathbf{X}_j,\mathbf{Z}_j\big) - \mathbb{E}\Big[g_{n,i,k}\big(\mathbf{X}_j,\mathbf{Z}_j\big)\Big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i),\ k=1,2. \tag{27}$$

Now, we decompose the term T_5, as defined in Equation (24). For every $i=1,\dots,n'$, a usual

Taylor expansion yields

$$\frac{1}{\hat{f}^2_{\mathbf{Z}}(\mathbf{Z}'_i)} - \frac{1}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)} = \frac{1}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}\{\frac{1}{\left(1+\dfrac{\hat{f}_{\mathbf{Z}}(\mathbf{Z}'_i)-f_{\mathbf{Z}}(\mathbf{Z}'_i)}{f_{\mathbf{Z}}(\mathbf{Z}'_i)}\right)^2} - 1\} = -2\frac{\hat{f}_{\mathbf{Z}}(\mathbf{Z}'_i)-f_{\mathbf{Z}}(\mathbf{Z}'_i)}{f^3_{\mathbf{Z}}(\mathbf{Z}'_i)} + T_{7,i},$$

where

$$T_{7,i} = \frac{3}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}(1+\alpha_{7,i})^{-4}\left(\frac{\hat{f}_{\mathbf{Z}}(\mathbf{Z}'_i)-f_{\mathbf{Z}}(\mathbf{Z}'_i)}{f_{\mathbf{Z}}(\mathbf{Z}'_i)}\right)^2, \text{ for some } |\alpha_{7,i}| \leq \left|\frac{\hat{f}_{\mathbf{Z}}(\mathbf{Z}'_i)-f_{\mathbf{Z}}(\mathbf{Z}'_i)}{f_{\mathbf{Z}}(\mathbf{Z}'_i)}\right|. \tag{28}$$

Therefore, we get the decomposition $T_5 = -2\,T_6 + T_7$, where

$$\begin{aligned} T_6 &:= \frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} K_h(\mathbf{Z}'_i - \mathbf{Z}_{j_1})K_h(\mathbf{Z}'_i - \mathbf{Z}_{j_2}) \cdot \frac{\hat{f}_{\mathbf{Z}}(\mathbf{Z}'_i)-f_{\mathbf{Z}}(\mathbf{Z}'_i)}{f^3_{\mathbf{Z}}(\mathbf{Z}'_i)}, \\ &\qquad \cdot \Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2}) - \mathbb{E}\big[g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2})|\mathbf{Z}_{j_1}=\mathbf{Z}_{j_2}=\mathbf{Z}'_i\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i) && (29) \\ T_7 &:= \frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} K_h(\mathbf{Z}'_i - \mathbf{Z}_{j_1})K_h(\mathbf{Z}'_i - \mathbf{Z}_{j_2}) \cdot T_{7,i} \\ &\qquad \cdot \Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2}) - \mathbb{E}\big[g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2})|\mathbf{Z}_{j_1}=\mathbf{Z}_{j_2}=\mathbf{Z}'_i\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i). && (30) \end{aligned}$$

Summing up all the previous equations, we get

$$T_1 = T_{4,1,1} + T_{4,1,2} + T_{4,2} + T_{4,3} - 2 \cdot T_6 + T_7 + T_3, \tag{31}$$

Lemma 27. *Under the assumptions 2.3, 2.4, 2.7, 2.8(i), 2.8 and (iii), if* $n'nh^p_{n,n'} \to \infty$, *if* $f_{\mathbf{Z}}$ *and* $f_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\cdot)$ *are continuous on* $\mathcal{Z}$ *for every* $\mathbf{x}$, *then* $T_{4,1,1} \xrightarrow{D} \mathcal{N}(0, V_2)$.

Afterwards, we will prove that all the remainders terms $T_{4,1,2}$, $T_{4,2}$, $T_{4,3}$, T_6, T_7 and T_3 are negligible, i.e. tends to zero in probability. These results are respectively proved in Subsections D.4, D.5, D.6, D.7, D.8, D.9 and D.10. Combining all these elements with Equation (31) yields $T_1 \xrightarrow{D} \mathcal{N}(0, V_2)$, as claimed. □

D.4 Proof of Lemma 27

Let us prove the convergence of $T_{4,1,1}$ to a non-degenerate Gaussian limit. Using the notation

$$\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j) := \frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'}\int g^*(\mathbf{x}_1,\mathbf{X}_j)\frac{K_h(\mathbf{Z}'_i-\mathbf{Z}_j)}{f_{\mathbf{Z}}(\mathbf{Z}'_i)}f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_i)d\mathbf{x}_1\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i),$$

and combining Equations (27) and (25), we have

$$T_{4,1,1} = \frac{1}{n}\sum_{j=1}^{n}\Big(\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j)-\mathbb{E}\big[\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j)\big]\Big).$$

For a fixed $j\in\{1,\dots,n\}$, we have

$$\begin{aligned}
\mathbb{E}[\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j)] &= \mathbb{E}\Big[\frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'}\int g^*(\mathbf{x}_1,\mathbf{X}_j)\frac{K_h(\mathbf{Z}'_i-\mathbf{Z}_j)}{f_{\mathbf{Z}}(\mathbf{Z}'_i)}f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_i)d\mathbf{x}_1\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)\Big]\\
&= \frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'}\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)\int g^*(\mathbf{x}_1,\mathbf{x}_2)\frac{K_h(\mathbf{Z}'_i-\mathbf{z}_2)}{f_{\mathbf{Z}}(\mathbf{Z}'_i)}f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_i)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{z}_2)d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{z}_2\\
&= \frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'}\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)\int g^*(\mathbf{x}_1,\mathbf{x}_2)\frac{K(\mathbf{t})}{f_{\mathbf{Z}}(\mathbf{Z}'_i)}f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_i)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{Z}'_i-h\mathbf{t})\,d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{t}\\
&\sim \frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'}\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)\int g^*(\mathbf{x}_1,\mathbf{x}_2)f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_i)f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{Z}=\mathbf{Z}'_i)\,d\mathbf{x}_1\,d\mathbf{x}_2,
\end{aligned}$$

using the continuity of $f_{\mathbf{X},\mathbf{Z}}(\cdot,\cdot)$ w.r.t. its second argument, which is guaranteed by Assumption 2.5.

By simple calculations, we obtain

$$\begin{aligned}
Var[T_{4,1,1}] &= \frac{1}{n^2}\sum_{j=1}^{n}Var[\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j)] = \frac{1}{n}Var[\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j)]\\
&\sim \frac{1}{n}\mathbb{E}\Bigg[\Big(\frac{\tilde{r}_{n,n'}}{n'}\sum_{i_1=1}^{n'}\int g^*(\mathbf{x}_1,\mathbf{X}_j)\frac{K_h(\mathbf{Z}'_{i_1}-\mathbf{Z}_j)}{f_{\mathbf{Z}}(\mathbf{Z}'_{i_1})}f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_{i_1})d\mathbf{x}_1\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_1})\Big)\\
&\qquad\cdot\Big(\frac{\tilde{r}_{n,n'}}{n'}\sum_{i_2=1}^{n'}\int g^*(\mathbf{x}_2,\mathbf{X}_j)\frac{K_h(\mathbf{Z}'_{i_2}-\mathbf{Z}_j)}{f_{\mathbf{Z}}(\mathbf{Z}'_{i_2})}f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{Z}=\mathbf{Z}'_{i_2})d\mathbf{x}_2\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_2})\Big)^T\Bigg]
\end{aligned}$$

$$\sim \frac{\tilde{r}^2_{n,n'}}{nn'^2}\sum_{i_1=1}^{n'}\sum_{i_2=1}^{n'}\frac{h^{-p}}{f_{\mathbf{Z}}(\mathbf{Z}'_{i_2})}\int K(\mathbf{t})K\left(\mathbf{t}+\frac{\mathbf{Z}'_{i_2}-\mathbf{Z}'_{i_1}}{h}\right)d\mathbf{t}\times\int g^*(\mathbf{x}_1,\mathbf{x}_3)g^*(\mathbf{x}_2,\mathbf{x}_3)$$
$$\cdot f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_{i_1})f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{Z}=\mathbf{Z}'_{i_2})f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_3|\mathbf{Z}=\mathbf{Z}'_{i_1})\,d\mathbf{x}_1d\mathbf{x}_2d\mathbf{x}_3\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_1})\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_2})^T.$$

Under the condition that $i_1\neq i_2\implies|\mathbf{Z}'_{i_2}-\mathbf{Z}'_{i_1}|>2h$, that is Assumption 2.8(i), all the extra-diagonal terms of the latter sum are zero. Using Assumption 2.8(iii), we get

$$Var[T_{4,1,1}]\sim\frac{\tilde{r}^2_{n,n'}}{nn'^2h^p}\sum_{i_1=1}^{n'}\frac{\int K^2}{f_{\mathbf{Z}}(\mathbf{Z}'_{i_1})}\int g^*(\mathbf{x}_1,\mathbf{x}_3)g^*(\mathbf{x}_2,\mathbf{x}_3)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_1})\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_1})^T$$
$$\times f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_{i_1})f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{Z}=\mathbf{Z}'_{i_1})f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_3|\mathbf{Z}=\mathbf{Z}'_{i_1})\,d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{x}_3$$
$$\sim\int K^2\int g^*(\mathbf{x}_1,\mathbf{x}_3)g^*(\mathbf{x}_2,\mathbf{x}_3)\overline{\boldsymbol{\psi}}(\mathbf{z}')\overline{\boldsymbol{\psi}}(\mathbf{z}')^T$$
$$\times f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{z}')f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{Z}=\mathbf{z}')f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_3|\mathbf{Z}=\mathbf{z}')\frac{f_{\mathbf{Z}',\infty}(\mathbf{z}')}{f_{\mathbf{Z}}(\mathbf{z}')}\,d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{x}_3\,d\mathbf{z}'.$$

Since the terms $\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j)$ are independent (given the sample $(\mathbf{Z}'_i)_{i\geq1}$), it only remains to check Lyapunov's third's moment condition and the proof is finished by applying the usual Central Limit Theorem. The analysis of the third moment of $T_{4,1,1}$ is similar to the study of the previous variance: for any $j=1,\dots,n$,

$$\mathbb{E}\Big[\big\|\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j)^{\otimes3}\big\|_\infty\Big]=\mathbb{E}\left[\left\|\Big(\frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'}\int g^*(\mathbf{x}_1,\mathbf{X}_j)\frac{K_h(\mathbf{Z}'_i-\mathbf{Z}_j)}{f_{\mathbf{Z}}(\mathbf{Z}'_i)}f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_i)d\mathbf{x}_1\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)\Big)^{\otimes3}\right\|_\infty\right]$$
$$\leq\frac{\tilde{r}^3_{n,n'}}{n'^3}\sum_{i_1=1}^{n'}\sum_{i_2=1}^{n'}\sum_{i_3=1}^{n'}\big\|\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_1})\otimes\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_2})\otimes\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_3})\big\|_\infty$$
$$\times\int\Big|g^*(\mathbf{x}_1,\mathbf{x}_4)g^*(\mathbf{x}_2,\mathbf{x}_4)g^*(\mathbf{x}_3,\mathbf{x}_4)\frac{K_h(\mathbf{Z}'_{i_1}-\mathbf{z}_4)K_h(\mathbf{Z}'_{i_2}-\mathbf{z}_4)K_h(\mathbf{Z}'_{i_3}-\mathbf{z}_4)}{f_{\mathbf{Z}}(\mathbf{Z}'_{i_1})f_{\mathbf{Z}}(\mathbf{Z}'_{i_2})f_{\mathbf{Z}}(\mathbf{Z}'_{i_3})}$$
$$\times f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_{i_1})f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{Z}=\mathbf{Z}'_{i_2})f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_3|\mathbf{Z}=\mathbf{Z}'_{i_3})f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_4,\mathbf{z}_4)\Big|d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{x}_3\,d\mathbf{x}_4\,d\mathbf{z}_4.$$

Through the change of variable $\mathbf{z}_4 = \mathbf{Z}'_{i_1} - h\mathbf{t}$, we get

$$\begin{aligned}
\mathbb{E}\Big[\big\|\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j)^{\otimes 3}\big\|_\infty\Big] &\leq \frac{\tilde{r}^3_{n,n'}}{n'^3 h^{2p}} \sum_{i_1=1}^{n'}\sum_{i_2=1}^{n'}\sum_{i_3=1}^{n'} \Big\|\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_1}) \otimes \overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_2}) \otimes \overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_3})\Big\|_\infty \\
&\times \int \Big| g^*(\mathbf{x}_1,\mathbf{x}_4) g^*(\mathbf{x}_2,\mathbf{x}_4) g^*(\mathbf{x}_3,\mathbf{x}_4) \frac{K(\mathbf{t}) K\left(\dfrac{\mathbf{Z}'_{i_2} - \mathbf{Z}'_{i_1}}{h} + \mathbf{t}\right) K\left(\dfrac{\mathbf{Z}'_{i_3} - \mathbf{Z}'_{i_1}}{h} + \mathbf{t}\right)}{f_{\mathbf{Z}}(\mathbf{Z}'_{i_1}) f_{\mathbf{Z}}(\mathbf{Z}'_{i_2}) f_{\mathbf{Z}}(\mathbf{Z}'_{i_3})} \\
&\times f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1 | \mathbf{Z} = \mathbf{Z}'_{i_1}) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2 | \mathbf{Z} = \mathbf{Z}'_{i_2}) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_3 | \mathbf{Z} = \mathbf{Z}'_{i_3}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_4, \mathbf{Z}'_{i_1} - h\mathbf{t}) \Big| \, d\mathbf{x}_1\, d\mathbf{x}_2\, d\mathbf{x}_3\, d\mathbf{x}_4\, d\mathbf{t}.
\end{aligned}$$

Due to Assumption 2.8(i), only the terms for which $i_1 = i_2 = i_3$ are nonzero, this yields

$$\begin{aligned}
\mathbb{E}\Big[\big\|\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j)^{\otimes 3}\big\|_\infty\Big] &\leq \frac{\tilde{r}^3_{n,n'}}{n'^3 h^{2p}} \sum_{i=1}^{n'} \|\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i) \otimes \overline{\boldsymbol{\psi}}(\mathbf{Z}'_i) \otimes \overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)\|_\infty \int \Big| g^*(\mathbf{x}_1,\mathbf{x}_4) g^*(\mathbf{x}_2,\mathbf{x}_4) g^*(\mathbf{x}_3,\mathbf{x}_4) \\
&\times \frac{K(\mathbf{t})^3}{f_{\mathbf{Z}}(\mathbf{Z}'_i)^3} f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1 | \mathbf{Z} = \mathbf{Z}'_i) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2 | \mathbf{Z} = \mathbf{Z}'_i) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_3 | \mathbf{Z} = \mathbf{Z}'_i) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_4, \mathbf{Z}'_i - h\mathbf{t}) \Big| d\mathbf{x}_1\, d\mathbf{x}_2\, d\mathbf{x}_3\, d\mathbf{x}_4\, d\mathbf{t} \\
&\leq \frac{\tilde{r}^3_{n,n'}}{n'^3 h^{2p}} C_K^2 \sum_{i=1}^{n'} \|\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i) \otimes \overline{\boldsymbol{\psi}}(\mathbf{Z}'_i) \otimes \overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)\|_\infty \int |K| \frac{f_{\mathbf{Z},max}}{f_{\mathbf{Z}}(\mathbf{Z}'_i)^3},
\end{aligned}$$

since g^* is bounded by one, by integrating w.r.t. the $\mathbf{x}_k$ variables, $k = 1, \dots, 4$. Since $\overline{\boldsymbol{\psi}}$ is bounded and $f_{\mathbf{Z}}(\mathbf{Z}'_i) > f_{\mathbf{Z},min}$ for every i, we obtain

$$\frac{1}{n^3} \sum_{j=1}^{n} \mathbb{E}\big[\|\boldsymbol{\omega}_n(\mathbf{X}_j,\mathbf{Z}_j)^{\otimes 3}\|_\infty\big] = O\left(\frac{\tilde{r}^3_{n,n'}}{n^2 n'^2 h^{2p}}\right) = O\left((nn'h^p)^{-1/2}\right) = o(1),$$

applying Assumption 2.8(ii)(a). Therefore, we have checked Lyapunov's condition and the result follows. $\square$

D.5 Convergence of $T_{4,1,2}$ to 0

Applying the Taylor-Lagrange formula to the function

$$\phi_{\mathbf{x}_1,\mathbf{u},i}(t) := f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1, \mathbf{Z}'_i + th\mathbf{u}),\ t \in [0,1], \mathbf{x}_1 \in \mathbb{R}^2, \mathbf{u} \in \mathbb{R}^p, i \geq 1,$$

we get from Equation (26) that

$$
\begin{aligned}
g_{n,i,2}(\mathbf{x},\mathbf{z}) &= \int g^*(\mathbf{x}_1,\mathbf{x})\frac{K(\mathbf{u})K_h(\mathbf{Z}'_i-\mathbf{z})}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}\Big(f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{Z}'_i+h\mathbf{u})-f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{Z}'_i)\Big)d\mathbf{x}_1d\mathbf{u} \\
&= \frac{K_h(\mathbf{Z}'_i-\mathbf{z})}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}\int g^*(\mathbf{x}_1,\mathbf{x})K(\mathbf{u})\Big(\phi_{\mathbf{x}_1,\mathbf{u},i}(1)-\phi_{\mathbf{x}_1,\mathbf{u},i}(0)\Big)d\mathbf{x}_1d\mathbf{u} \\
&= \frac{K_h(\mathbf{Z}'_i-\mathbf{z})}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}\int g^*(\mathbf{x}_1,\mathbf{x})K(\mathbf{u})\Big(\sum_{k=1}^{\alpha-1}\frac{1}{k!}\phi^{(k)}_{\mathbf{x}_1,\mathbf{u},i}(0)+\frac{1}{\alpha!}\phi^{(\alpha)}_{\mathbf{x}_1,\mathbf{u},i}(t_{\mathbf{x}_1,\mathbf{u},i})\Big)d\mathbf{x}_1d\mathbf{u}.
\end{aligned}
$$

By Assumption 2.1, for every $i=1,\dots,\alpha-1$, $\int_{\mathbb{R}^p}K(\mathbf{u})\phi^{(i)}_{\mathbf{x}_1,\mathbf{u},i}(0)d\mathbf{u}=0$. Therefore,

$$
\begin{aligned}
g_{n,i,2}(\mathbf{x},\mathbf{z}) &= \frac{K_h(\mathbf{Z}'_i-\mathbf{z})}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}\int g^*(\mathbf{x}_1,\mathbf{x})K(\mathbf{u})\Big(\frac{1}{\alpha!}\phi^{(\alpha)}_{\mathbf{x}_1,\mathbf{u},i}(t_{\mathbf{x}_1,\mathbf{u},i})\Big)d\mathbf{x}_1d\mathbf{u} \\
&= \frac{h^\alpha K_h(\mathbf{Z}'_i-\mathbf{z})}{\alpha! f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}\int g^*(\mathbf{x}_1,\mathbf{x})K(\mathbf{u})\Big(\sum_{i_1,\dots,i_\alpha=1}^{p}u_{i_1}\dots u_{i_\alpha}\frac{\partial^\alpha f_{\mathbf{X},\mathbf{Z}}}{\partial z_{i_1}\dots\partial z_{i_\alpha}}\Big(\mathbf{x}_1,\mathbf{z}+t_{\mathbf{x}_1,\mathbf{u},i}h\mathbf{u}\Big)\Big)d\mathbf{x}_1d\mathbf{u}.
\end{aligned}
$$

In other words, $g_{n,i,2}(\mathbf{x},\mathbf{z})=h^\alpha K_h(\mathbf{Z}'_i-\mathbf{z})\chi(\mathbf{x},\mathbf{z})/\{\alpha! f^2_{\mathbf{Z}}(\mathbf{Z}'_i)\}$, for some measurable χ that is upper bounded when $\mathbf{z}\in\mathcal{Z}$ and $\mathbf{x}\in\mathbb{R}^2$. By the same calculations as in D.4, we easily show that $Var(T_{4,1,2})=O(h^{2\alpha})$. Therefore, $T_{4,1,2}=o_{\mathbb{P}}(1)$. $\square$

D.6 Convergence of $T_{4,2}$ to 0

Let us remind the definition of $T_{4,2}$.

$$
\begin{aligned}
T_{4,2} &:= \frac{\tilde r_{n,n'}}{n^2n'}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2})K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_1})K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_2}) \\
&\qquad -\int g^*(\mathbf{x}_1,\mathbf{X}_{j_1})K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_1})K_h(\mathbf{Z}'_i-\mathbf{z}_1)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{z}_1)d\mathbf{x}_1d\mathbf{z}_1\Big)\frac{\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}. \\
&:= \frac{\tilde r_{n,n'}}{n^2n'}\sum_{i=1}^{n'}\sum_{j_1,j_2=1,j_1\neq j_2}^{n}\{W_{i,j_1,j_2}-\mathbb{E}[W_{i,1,2}|\mathbf{Z}_{j_1}]\}+\frac{\tilde r_{n,n'}}{n^2n'}\sum_{i=1}^{n'}\sum_{j=1}^{n}\{W_{i,j,j}-\mathbb{E}[W_{i,1,2}|\mathbf{Z}_{j_1}]\} \\
&:= T_{4,2,1}+T_{4,2,2},
\end{aligned}
$$

with obvious notations. The second term on the r.h.s. corresponds to the diagonal of the double sum and $T_{4,2,2}=O_P(\sqrt{n'}(nh^p)^{-1})=o_P(1)$ under our assumptions. It remains to prove that $T_{4,2,1}=o_P(1)$. Since it is centered, it is sufficient to show that its variance tend to zero when

$n \to \infty$.

$$Var(T_{4,2,2}) = \frac{\tilde{r}^2_{n,n'}}{(n^2 n')^2} \sum_{i_1=1}^{n'} \sum_{i_2=1}^{n'} \sum_{j_1,j_2=1, j_1\neq j_2}^{n} \sum_{j_3,j_4=1, j_3\neq j_4}^{n}$$
$$\left\{ W_{i_1,j_1,j_2} - \mathbb{E}[W_{i_1,1,2}|\mathbf{Z}_{j_1}] \right\} \left\{ W_{i_2,j_3,j_4} - \mathbb{E}[W_{i_2,1,2}|\mathbf{Z}_{j_1}] \right\}.$$

It is easy to see that the latter expectations are zero when all indices j_1, j_2, j_3 and j_4 are different. Even more, it is necessary that $j_1 \in \{j_3, j_4\}$ or $j_2 \in \{j_3, j_4\}$. For instance, assume $j_1 = j_3$. If, moreover, $j_4 \notin \{j_1, j_2\}$, then

$$\begin{aligned}
&\mathbb{E}\Big[\big(W_{i_1,j_1,j_2} - \mathbb{E}[W_{i_1,j_1,j_2}|\mathbf{Z}_{j_1}] \big) \times \big(W_{i_2,j_1,j_4} - \mathbb{E}[W_{i_2,j_1,j_4}|\mathbf{Z}_{j_1}] \big) \Big] \\
&= \ \mathbb{E}\Big[\mathbb{E}[\big(W_{i_1,j_1,j_2} - \mathbb{E}[W_{i_1,j_1,j_2}|\mathbf{Z}_{j_1}] \big) | \mathbf{Z}_{j_1}, \mathbf{Z}_{j_4}] \times \big(W_{i_2,j_1,j_4} - \mathbb{E}[W_{i_2,j_1,j_4}|\mathbf{Z}_{j_1}] \big) \Big] = 0.
\end{aligned}$$

Therefore, it is necessary to impose a second identity among the four indices j_k, $k = 1, \ldots, 4$. For instance, assume $j = j_3 =: j$ and $j_2 = j_4 =: j'$. Then, the corresponding expectations are given by four terms of the type

$$\begin{aligned}
\mathbb{E}[W_{i_1,j,j'} W_{i_2,j,j'}] &= \frac{\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_1}) \overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_2})}{f^2_{\mathbf{Z}}(\mathbf{Z}'_{i_1}) f^2_{\mathbf{Z}}(\mathbf{Z}'_{i_2})} \int g^*(\mathbf{x}_j, \mathbf{x}_{j'})^2 K_h(\mathbf{Z}'_{i_1} - \mathbf{Z}'_j) K_h(\mathbf{Z}'_{i_1} - \mathbf{z}_{j'}) K_h(\mathbf{Z}'_{i_2} - \mathbf{Z}'_j) \\
\times \ & K_h(\mathbf{Z}'_{i_2} - \mathbf{z}_{j'}) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{z}_j) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{j'}, \mathbf{z}_{j'}) d\mathbf{x}_j \, d\mathbf{z}_j \, d\mathbf{x}_{j'} \, d\mathbf{z}_{j'} \\
= \ & \frac{\overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_1}) \overline{\boldsymbol{\psi}}(\mathbf{Z}'_{i_2})}{f^2_{\mathbf{Z}}(\mathbf{Z}'_{i_1}) f^2_{\mathbf{Z}}(\mathbf{Z}'_{i_2})} \int g^*(\mathbf{x}_j, \mathbf{x}_{j'})^2 K(\mathbf{t}_j) K(\mathbf{t}_{j'}) K_h(\frac{(\mathbf{Z}'_{i_2} - \mathbf{Z}'_{i_1})}{h} + \mathbf{t}_j) K_h(\frac{(\mathbf{Z}'_{i_2} - \mathbf{Z}'_{i_1})}{h} + \mathbf{t}_{j'}) \\
\times \ & f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_j, \mathbf{Z}'_{i_1} - h\mathbf{t}_j) f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_{j'}, \mathbf{Z}'_{i_1} - h\mathbf{t}_{j'}) \, d\mathbf{x}_j \, d\mathbf{t}_j \, d\mathbf{x}_{j'} \, d\mathbf{t}_{j'} / h^{2p},
\end{aligned}$$

that is nonzero only when $\mathbf{Z}'_{i_1} = \mathbf{Z}'_{i_2}$. The other terms can be dealt similarly. Therefore, the number of nonzero terms on the r.h.s. of the latter equation is less than $n'n^2$. We deduce

$$Var(T_{4,2,1}) = O\left(\frac{\tilde{r}^2_{n,n'}}{h^{2p} n^2 n'} \right) = O\left(\frac{1}{nh^p} \right)$$

Therefore, by Assumption 2.8(ii)(a), $T_{4,2,1}$ converges to 0 in $L_2(\mathbb{P})$ and then $T_{4,2}$ tends to zero in probability, as claimed. $\square$

D.7 Convergence of $T_{4,3}$ to 0

Let us remind the definition of $T_{4,3}$.

$$
\begin{aligned}
T_{4,3} &:= \frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'}\Bigg(\mathbb{E}\Big[\frac{K_h(\mathbf{Z}'_i-\mathbf{Z}_1)K_h(\mathbf{Z}'_i-\mathbf{Z}_2)}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}g^*(\mathbf{X}_1,\mathbf{X}_2)\Big] - \mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{Z}'_i\big]\Bigg)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i) \\
&= \frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'}\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)\Bigg(\int \frac{K_h(\mathbf{Z}'_i-\mathbf{z}_1)K_h(\mathbf{Z}'_i-\mathbf{z}_2)}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)}g^*(\mathbf{x}_1,\mathbf{x}_2)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{z}_1)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{z}_2)\,d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{z}_1\,d\mathbf{z}_2 \\
&\qquad - \int g^*(\mathbf{x}_1,\mathbf{x}_2)f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_i)f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{Z}=\mathbf{Z}'_i)d\mathbf{x}_1 d\mathbf{x}_2\Bigg) \\
&= \frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'}\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)\int K(\mathbf{t}_1)K(\mathbf{t}_2)\Bigg(\frac{f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{Z}'_i+h\mathbf{t}_1)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{Z}'_i+h\mathbf{t}_2)}{f^2_{\mathbf{Z}}(\mathbf{Z}'_i)} \\
&\qquad - f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z}=\mathbf{Z}'_i)f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{Z}=\mathbf{Z}'_i)\Bigg)g^*(\mathbf{x}_1,\mathbf{x}_2)\,d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{t}_1\,d\mathbf{t}_2.
\end{aligned}
$$

Apply an α-order Taylor expansion to the function $(\mathbf{t}_1,\mathbf{t}_2)\mapsto f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_1,\mathbf{Z}'_i+h\mathbf{t}_1)f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}_2,\mathbf{Z}'_i+h\mathbf{t}_2)$. Since we work with α-order kernels, it is easy to see that $T_{4,3}=O_{\mathbb{P}}(h^{2\alpha}\tilde{r}_{n,n'})=o_{\mathbb{P}}(1)$, under Assumpion 2.8(ii)(a). $\square$

D.8 Convergence of T_6 to 0

Replacing $\hat{f}_{\mathbf{Z}}$ in the definition of T_6 above by the normalized sum of the kernels, we get

$$
\begin{aligned}
T_6 = &\frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\frac{K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_1})K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_2})}{f^3_{\mathbf{Z}}(\mathbf{Z}'_i)}\big(\mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{Z}'_i)]-f_{\mathbf{Z}}(\mathbf{Z}'_i)\big) \\
&\times\Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2})-\mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{Z}'_i\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i) \\
&+\frac{\tilde{r}_{n,n'}}{n'n^3}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\sum_{j_3=1}^{n}\frac{K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_1})K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_2})}{f^3_{\mathbf{Z}}(\mathbf{Z}'_i)}\big(K_h(\mathbf{Z}'_i-\mathbf{Z}_{j_3})-\mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{Z}'_i)]\big) \\
&\times\Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2})-\mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{Z}'_i\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i) =: T_{6,1}+T_{6,2}.
\end{aligned}
$$

The first term $T_{6,1}$ is a bias term. By Assumption 2.1,

$$
\sup_{i=1,\dots,n'}\big|\mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{Z}'_i)]-f_{\mathbf{Z}}(\mathbf{Z}'_i)\big| = O(h^\alpha).
$$

The sum of the diagonal terms in $T_{6,1}$ is

$$-\frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j=1}^{n}\frac{K_h^2(\mathbf{Z}_i'-\mathbf{Z}_j)}{f_{\mathbf{Z}}^3(\mathbf{Z}_i')}\big(\mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{Z}_i')]-f_{\mathbf{Z}}(\mathbf{Z}_i')\big)\mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{Z}_i'\big]\overline{\boldsymbol{\psi}}(\mathbf{Z}_i')=O_{\mathbb{P}}\Big(\frac{\tilde{r}_{n,n'}h^{\alpha}}{nh^p}\Big).$$

The sum of the extra-diagonal terms in $T_{6,1}$ is a centered r.v.

$$\begin{aligned}\overline{T}_{6,1}:=\frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{1\leq j_1\neq j_2\leq n}&\frac{K_h(\mathbf{Z}_i'-\mathbf{Z}_{j_1})K_h(\mathbf{Z}_i'-\mathbf{Z}_{j_2})}{f_{\mathbf{Z}}^3(\mathbf{Z}_i')}\big(\mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{Z}_i')]-f_{\mathbf{Z}}(\mathbf{Z}_i')\big)\\ &\times\Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2})-\mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{Z}_i'\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{Z}_i').\end{aligned}$$

Note that $\mathbf{z}\mapsto f_{\mathbf{Z}}(\mathbf{z})$ and $\mathbf{z}\mapsto\mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{z}_1,\mathbf{Z}_2=\mathbf{z}_2\big]$ is α-times continuously differentiable on $\mathcal{Z}$, because of 2.1 and 2.5. By α-order Taylor expansions of such terms, they yield some factors h^{α}. It is easy to check that the variance of $\overline{T}_{6,1}$ is of order $\tilde{r}_{n,n'}^2h^{2\alpha}/(n^2h^{2p})$. Therefore,

$$T_{6,1}=O_{\mathbb{P}}\Big(\frac{\tilde{r}_{n,n'}h^{\alpha}}{nh^p}\Big)=O_{\mathbb{P}}\Big(\frac{(n')^{1/2}h^{\alpha}}{\sqrt{nh^p}}\Big)=o_{\mathbb{P}}(1).$$

Concerning $T_{6,2}$, we can assume that the indices j_1, j_2 and j_3 are pairwise distinct. Indeed, the cases of one or two identities among such indices can be easily dealt. They yield an upper bound that is $O_{\mathbb{P}}(\tilde{r}_{n,n'}h^{\alpha}/(nh^p))$ as above, and they are negligible. Once we remove such terms from the triple sum defining $T_{6,2}$, we get the centered r.v. $\overline{T}_{6,2}$. Let us calculate the second moment of $\overline{T}_{6,2}$.

$$\begin{aligned}\mathbb{E}\big[\overline{T}_{6,2}^2\big]&:=\frac{nn'h^p}{n'^2n^6}\sum_{i_1=1}^{n'}\sum_{i_2=1}^{n'}\sum_{1\leq j_1\neq j_2\neq j_3\leq n}\sum_{1\leq j_4\neq j_5\neq j_6\leq n}\mathbb{E}\Bigg[\frac{K_h(\mathbf{Z}_{i_1}'-\mathbf{Z}_{j_1})K_h(\mathbf{Z}_{i_1}'-\mathbf{Z}_{j_2})}{f_{\mathbf{Z}}^3(\mathbf{Z}_{i_1}')}\\ &\times\big(K_h(\mathbf{Z}_{i_1}'-\mathbf{Z}_{j_3})-\mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{Z}_{i_1}')]\big)\Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2})-\mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{Z}_{i_1}'\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{Z}_{i_1}')\\ &\times\frac{K_h(\mathbf{Z}_{i_2}'-\mathbf{Z}_{j_4})K_h(\mathbf{Z}_{i_2}'-\mathbf{Z}_{j_5})}{f_{\mathbf{Z}}^3(\mathbf{Z}_{i_2}')}\big(K_h(\mathbf{Z}_{i_2}'-\mathbf{Z}_{j_6})-\mathbb{E}[\hat{f}_{\mathbf{Z}}(\mathbf{Z}_{i_2}')]\big)\\ &\qquad\times\Big(g^*(\mathbf{X}_{j_4},\mathbf{X}_{j_5})-\mathbb{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{Z}_{i_2}'\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{Z}_{i_2}')^T\Bigg]\\ &=:\frac{nn'h^p}{n'^2n^6}\sum_{i_1,i_2=1}^{n'}\sum_{1\leq j_1\neq j_2\neq j_3\leq n}\sum_{1\leq j_4\neq j_5\neq j_6\leq n}E_{i_1,i_2,j_1-j_6}.\end{aligned}$$

When all the indices of the latter sums are different, the latter expectation is zero. Non zero terms above are obtained only when j_3 and j_6 are equal to some other indices. In the case $j_3 = j_6$ and no other identity among the indices, obtain two extra factors h^α through α-order limited expansions of $\mathbf{z} \mapsto \mathbb{E}\big[g^*(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{Z}_1 = \mathbf{z}_1, \mathbf{Z}_2 = \mathbf{z}_2\big]$. This yields an order $O(nn'h^{p+2\alpha}/(nh^p))$. When j_3 and j_6 are equal to other indices ($j_3 = j_3$ and $j_6 = j_2$, e.g.), we lose h^p but we still benefit from the two latter factors h^α. This yields an upper bound $O(nn'h^{p+2\alpha}/(n^2h^{2p}) = o(1)$. The other situations can be managed similarly. We get

$$\mathbb{E}\big[\overline{T}_{6,2}^2\big] = O\Big(\frac{nn'h^{p+2\alpha}}{nh^p}\Big) = o(1).$$

Globally, we obtain $T_6 \to 0$ in probability under Assumptions 2.8(ii)(a). $\square$

D.9 Convergence of T_7 to 0

Since $\sup_{i=1,\dots,n'} |\hat{f}_{\mathbf{Z}}(\mathbf{Z}'_i) - f_{\mathbf{Z}}(\mathbf{Z}'_i)| = o_{\mathbb{P}}(1)$, note that

$$\sup_{i=1,\dots,n'} |T_{7,i}| \leq \frac{6}{f_{\mathbf{Z},min}^4} \sup_{i=1,\dots,n'} |\hat{f}_{\mathbf{Z}}(\mathbf{Z}'_i) - f_{\mathbf{Z}}(\mathbf{Z}'_i)|^2,$$

with a probability arbitrarily close to one. Apply Lemma 19 with a fixed $t > 0$ and $\mathbf{z} = \mathbf{Z}'_i$ for each $i = 1, \dots, n'$

$$\mathbb{P}\bigg(\sup_{i=1,\dots,n'} |T_{7,i}| \geq \frac{3}{f_{\mathbf{Z},min}^4}\Big(\frac{C_{K,\alpha}h^\alpha}{\alpha!} + t\Big)^2\bigg) \geq 2n' \exp\bigg(-\frac{nh^pt^2}{2f_{\mathbf{Z},max}\int K^2 + (2/3)C_K t}\bigg).$$

By setting $t = h^\alpha$, deduce $\sup_{i=1,\dots,n'} |T_{7,i}| = O_P(h^{2\alpha})$ since $nh^{p+2\alpha}/\ln n' \to \infty$. Then,

$$|T_7| \leq \frac{\tilde{r}_{n,n'}}{n'n^2} \sup_i |T_{7,i}| \sum_{i=1}^{n'} \sum_{j_1=1}^{n} \sum_{j_2=1}^{n} |K|_h(\mathbf{Z}'_i - \mathbf{Z}_{j_1})|K|_h(\mathbf{Z}'_i - \mathbf{Z}_{j_2})|\overline{\boldsymbol{\psi}}(\mathbf{Z}'_i)|.$$

The expectation of the double sum is $O(1)$. Then, by Markov's inequality, we deduce $T_7 = O_{\mathbb{P}}(\tilde{r}_{n,n'} \sup_i |T_{7,i}|) = O_{\mathbb{P}}\big((n'/(nh^p))^{1/2}\big)$. Therefore, $T_T = o_{\mathbb{P}}(1)$. $\square$

D.10 Convergence of T_3 to 0

For every $\epsilon > 0$, by Markov's inequality,

$$\mathbb{P}(|T_3| > \epsilon) \leq \frac{C_{\Lambda''} \tilde{r}_{n,n'}}{2n'\epsilon} \sum_{i=1}^{n'} \mathbb{E}\big[\big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i} \big)^2 \big] \boldsymbol{\psi}(\mathbf{Z}'_i).$$

An approximated calculation of $\mathbb{E}\big[\big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{Z}'_i} \big)^2 \big]$ can be obtained following the steps of the proof of Theorem 6. Indeed, it can be easily seen that the order of magnitude of the latter expectation is the same as the variance of $U_{n,i}(g^*)$, and then of its Hájek projection $\hat{U}_{n,i}(g)$. Since the latter variance is $O((nh^p)^{-1})$, we get

$$\mathbb{P}(|T_3| > \epsilon) \leq B_1 \frac{\tilde{r}_{n,n'}}{nh^p \epsilon},$$

for some constant B_1. Since $n'/(nh^p) \to 0$, we get $T_3 = o_{\mathbb{P}}(1)$, as claimed. $\square$