Journal of Multivariate Analysis 178 (2020) 104610

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

On Kendall's regression

Alexis Derumigny [a], Jean-David Fermanian [b,*]

[a] University of Twente, 5 Drienerlolaan, 7522 NB Enschede, Netherlands
[b] Crest-Ensae, 5 av. Henry le Chatelier, 91764 Palaiseau cedex, France

ARTICLE INFO

Article history:
Received 8 August 2019
Received in revised form 11 March 2020
Accepted 11 March 2020
Available online 23 March 2020

AMS 2010 subject classifications:
primary 62H20
secondary 62F12

Keywords:
Conditional dependence measures
Conditional Kendall's tau
Kernel smoothing
Regression-type models

ABSTRACT

Conditional Kendall's tau is a measure of dependence between two random variables, conditionally on some covariates. We assume a regression-type relationship between conditional Kendall's tau and some covariates, in a parametric setting with a large number of transformations of a small number of regressors. This model may be sparse, and the underlying parameter is estimated through a penalized criterion and a two-step inference procedure. We prove non-asymptotic bounds with explicit constants that hold with high probabilities. We derive the consistency of the latter estimator, its asymptotic law and some oracle properties. Some simulations and applications to real data conclude the paper.

1. Introduction

In dependence modeling, it is common to work with scalar dependence measures which are margin-free. They can be used to quantify the positive or negative relationship between two random variables X_1 and X_2. One of the most popular of them is Kendall's tau, a dependence measure defined by

$$\tau_{1,2} := \Pr\big((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) > 0\big) - \Pr\big((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) < 0\big),$$

where $(X_{i,1}, X_{i,2})$, $i \in \{1, 2\}$ are i.i.d. copies of (X_1, X_2), see [13]. When a covariate $\mathbf{Z}$ is available, it is natural to work with the conditional version of this measure, i.e., the conditional Kendall's tau. It is defined as

$$\tau_{1,2|\mathbf{Z}=\mathbf{z}} := \Pr\big((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) > 0 \big| \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\big) - \Pr\big((X_{1,1} - X_{2,1})(X_{1,2} - X_{2,2}) < 0 \big| \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}\big),$$

where $(X_{i,1}, X_{i,2}, \mathbf{Z}_i)$, $i \in \{1, 2\}$ are i.i.d. copies of $(X_1, X_2, \mathbf{Z})$. Here, the goal will be to study to what extent a p-dimensional covariate $\mathbf{z}$ can affect the dependence between the two variables of interest X_1 and X_2.

The latter conditional dependence measure $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ may be written as a functional of the conditional copula of (X_1, X_2) given $\mathbf{Z} = \mathbf{z}$. Most often, it is difficult to have an intuition about the functional link between a general conditional copula (or some associated measures of dependence) and the underlying explanatory variables. Sometimes, it is even unclear whether such covariates have an influence on the dependence between the variables of interest. This is the so-called "simplifying assumption", well-known in the world of copula modeling (see [5] and the references therein). This issue is

* Corresponding author.
E-mail addresses: a.f.f.derumigny@utwente.nl (A. Derumigny), jean-david.fermanian@ensae.fr (J.-D. Fermanian).

https://doi.org/10.1016/j.jmva.2020.104610

particularly crucial with pair-copula constructions, as pointed out in [1,10,12], among others. In our case, we will evaluate an explicit and flexible link between some dependence measure, the Kendall's tau, and the vector of covariates: we focus on the function $\mathbf{z} \mapsto \tau_{1,2|\mathbf{Z}=\mathbf{z}}$ for $\mathbf{z} \in \mathscr{Z}$, where $\mathscr{Z}$ denotes a compact subset of $\mathbb{R}^p$ that is included in the support of $\mathbf{Z}$ and so that the density $f_{\mathbf{Z}}$ of $\mathbf{Z}$ is bounded from below on $\mathscr{Z}$. In order to simplify notation, the reference to the conditioning event $\mathbf{Z} \in \mathscr{Z}$ will be most often omitted. As a sub-product of our model, we will be able to provide a test of the "simplifying assumption".

First note that a Kendall's tau takes its values in the interval $[-1, 1]$, and not on the whole real line. Nevertheless, for some known increasing function $\Lambda : [-1, 1] \to \mathbb{R}$, assumed to be continuously differentiable on the open interval $(-1, 1)$, the function $\mathbf{z} \mapsto \Lambda\left(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\right)$ takes values on up to the whole real line potentially. Typical transforms are $\Lambda(\tau) = \ln\big((1+\tau)/(1-\tau)\big)$ (the Fisher transform) or $\Lambda(\tau) = \ln\Big(-\ln\big((1-\tau)/2\big)\Big)$. A simple idea would be to assume a linear relationship $\Lambda\left(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\right) = \mathbf{z}^\top \boldsymbol{\beta}_0$, for all $\mathbf{z} \in \mathscr{Z}$, with an unknown parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^p$. To increase the flexibility of this model, we propose rather to decompose the function $\mathbf{z} \mapsto \tau_{1,2|\mathbf{Z}=\mathbf{z}}$ on some functional basis $(\psi_i)_{i\geq 1}$, as any element of a space of functions from $\mathscr{Z}$ to $\mathbb{R}$. We will assume that only a finite number of elements are necessary to represent this function. This means the model is now

$$\Lambda\left(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\right) = \sum_{i=1}^{p'} \psi_i(\mathbf{z})\beta^*_i = \boldsymbol{\psi}(\mathbf{z})^\top \boldsymbol{\beta}^*, \tag{1}$$

for all $\mathbf{z} \in \mathscr{Z}$, with $p' > 0$ and a "true" unknown parameter $\boldsymbol{\beta}^* \in \mathbb{R}^{p'}$. The function $\boldsymbol{\psi}(\cdot) := \big(\psi_1(\cdot), \ldots, \psi_{p'}(\cdot)\big)^\top$ from $\mathbb{R}^p$ to $\mathbb{R}^{p'}$ is known and corresponds to deterministic transformations of the covariates $\mathbf{z}$. Since our model (1) assumes a linear relationship between some transforms of $\mathbf{z}$ and the unobservable "explained" quantity $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$, it is called a "Kendall's regression".

In practice, it is not easy to have intuition about which kind of basis to use, especially in our framework of conditional dependence measurement. Therefore, the most simple solution is to use a lot of different functions : polynomials, exponentials, sinuses and cosinuses, indicator functions, etc. They allow to take into account potential non-linearities and even discontinuities of conditional Kendall's taus with respect to $\mathbf{z}$. For the sake of identifiability, we only require their linear independence, as seen in the following proposition (whose straightforward proof is omitted).

Proposition 1. *The parameter $\boldsymbol{\beta}^*$ in Model (1) is identifiable if and only if the functions $(\psi_1, \ldots, \psi_{p'})$ are linearly independent $\Pr_{\mathbf{Z}}$-a.e. in the sense that, for any given vector $\mathbf{t} = (t_1, \ldots, t_{p'}) \in \mathbb{R}^{p'}$, $\Pr_{\mathbf{Z}}\big(\boldsymbol{\psi}(\mathbf{Z})^\top \mathbf{t} = 0\big) = 1$ implies $\mathbf{t} = 0$.*

With such a large choice among flexible classes of functions, it is unlikely we will be able to guess the right ones ex ante. Therefore, it will be necessary to consider a large number of functions ψ_i under a sparsity constraint: the cardinality of $\mathscr{S}$, the set of non-zero components of $\boldsymbol{\beta}^*$, is less than some $s \in \{1, \ldots, p'\}$. It is denoted by $|\mathscr{S}| = |\boldsymbol{\beta}^*|_0$, where $|\cdot|_0$ yields the number of non-zero components of any vector in $\mathbb{R}^{p'}$. Note that, in our framework and hereafter, p and p' will be held fixed. Typically, choose $p' = 10$ or $p' = 30$ while the original dimension p is small ($p \leq 3$). This corresponds to the decomposition of a function, defined on a small-dimension domain, in a mildly large basis. We will not consider models for which the number of basis functions tends to the infinity with n, as in sieves approaches (see [4] and the references therein). Even possible, such extensions would very significantly modify our asymptotic results and they are left for further studies.

Note that, in Model (1), there is no noise perturbing the variable of interest. In the same way, a classical linear model $Y = \mathbf{X}^\top\boldsymbol{\beta}^* + \varepsilon$ can be rewritten as $\mathrm{E}[Y|\mathbf{X} = \mathbf{x}] = \mathbf{x}^\top\boldsymbol{\beta}^*$ without any explicit noise. In our case, $\Lambda\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\big)$ is just a deterministic function of the variable $\mathbf{z}$. But contrary to more usual models, randomness will come here from the fact the "explained variable" - the conditional Kendall's tau $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ - is not observed in (1). Therefore, a direct estimation of the parameter $\boldsymbol{\beta}^*$ (for example, by the ordinary least squares, or by the Lasso) is unfeasible. In other words, even if the function $\mathbf{z} \mapsto \Lambda\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\big)$ is deterministic, finding the best $\boldsymbol{\beta}$ in Model (1) is far from being just a numerical analysis problem since the function to be decomposed is unknown.

For the sake of conditional Kendall's tau inference, a dataset $(X_{i,1}, X_{i,2}, \mathbf{Z}_i)$, $i \in \{1, \ldots, n\}$ is available. As a first stage, we will approximate the unknown quantity $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ by an estimate $\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}}$. More precisely, we fix a finite collection of points $\mathbf{z}'_1, \ldots, \mathbf{z}'_{n'} \in \mathscr{Z}^{n'}$ and calculate $\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ for each of these points. Then, as a second stage, $\hat{\boldsymbol{\beta}}$ is estimated as the minimizer of the l_1-penalized criterion

$$\hat{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{p'}} \left\{ \frac{1}{n'} \sum_{i=1}^{n'} \Big(\Lambda(\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}) - \boldsymbol{\psi}(\mathbf{z}'_i)^\top \boldsymbol{\beta} \Big)^2 + \lambda |\boldsymbol{\beta}|_1 \right\}, \tag{2}$$

where λ is a positive tuning parameter (that may depend on n and n'), and $|\cdot|_q$ denotes the l_q norm, for $1 \leq q \leq \infty$. Note that even if we study the general case with any $\lambda \geq 0$, the properties of the unpenalized estimator can be derived by choosing the particular case $\lambda = 0$. Moreover, the points $\mathbf{z}'_j$, $j \in \{1, \ldots, n'\}$ may be arbitrarily chosen in $\mathscr{Z}$, and then in the support of $\mathbf{Z}$. In particular, even possible, it is not mandatory to select them into the sample $(\mathbf{z}_i)_{i\in\{1,\ldots,n\}}$, nor to draw them along the (estimated) law $f_{\mathbf{Z}}$.

Several nonparametric estimators of $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_j}$ [7,9,15] can potentially be used. We refer to [7] for a detailed analysis of their statistical properties, which are mainly recalled as lemmas in Appendix A. These estimators (called "first-step estimators") are of the form

$$\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}} := \sum_{i=1}^{n}\sum_{j=1}^{n} w_{i,n}(\mathbf{z})w_{j,n}(\mathbf{z})g^*(\mathbf{X}_i, \mathbf{X}_j), \tag{3}$$

where g^* is a bounded function, $\mathbf{X}_i := (X_{i,1}, X_{i,2})$ for $i \in \{1,\dots,n\}$ and $w_{i,n}(\mathbf{z}) := K_h(\mathbf{Z}_i - \mathbf{z})/\sum_{j=1}^n K_h(\mathbf{Z}_j - \mathbf{z})$, $h = h(n) > 0$ denoting the bandwidth sequence, K being a kernel on $\mathbb{R}^p$ and $K_h(\mathbf{z}) = K(\mathbf{z}/h)/h^p$. In the same way, the conditional Kendall's tau can be rewritten as $\tau_{1,2|\mathbf{Z}=\mathbf{z}} = \mathrm{E}[g^*(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}]$ for the same choices of g^*. Possible choices of g^* are given in Appendix A.

It may seem surprising to invoke nonparametric estimators of $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ for the purpose of estimating a parametric model of the latter map. Actually, in this article, we propose an improvement over the first-step estimators, that we will call "two-step estimator" and which is defined by $\hat{\tau}^{2-step}_{1,2|\mathbf{Z}=\mathbf{z}} := \Lambda^{-1}\big(\boldsymbol{\psi}(\mathbf{z})^\top\hat{\boldsymbol{\beta}}\big)$, where $\hat{\boldsymbol{\beta}}$ is given in (2). Then, the parametric (two-step) estimator $\hat{\tau}^{2-step}_{1,2|\mathbf{Z}=\mathbf{z}}$ of the mapping $\mathbf{z} \mapsto \tau_{1,2|\mathbf{Z}=\mathbf{z}}$ has several nice features compared to the nonparametric (first-step) estimator $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$:

- The display of a mapping $\mathbf{z} \mapsto \tau_{1,2|\mathbf{Z}=\mathbf{z}}$ becomes difficult as soon as the dimension p of $\mathbf{Z}$ becomes larger than two. Therefore, when there are more than two regressors, it is often more practical to discuss the significance of some coefficients $\hat{\boldsymbol{\beta}}$ instead of plotting surfaces (or some of their sections).
- In many settings, the semi-parametric estimator $\hat{\boldsymbol{\beta}}$ will inherit the $\sqrt{nh^p}$ rate of convergence of the first stage nonparametric estimators $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i}$, as expected. Nonetheless, when n and n' both tend to the infinity and for a convenient choice of some tuning parameters, we are able to obtain $\sqrt{n}$-estimators of $\boldsymbol{\beta}$ and then of $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$, under model (1): see Corollary 12.
- Once an estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}^*$ has been computed, the prediction of all the conditional Kendall's tau's for m new values of $\mathbf{z}$, which is just the computation of $\Lambda^{-1}\big(\boldsymbol{\psi}(\mathbf{z})^\top\hat{\boldsymbol{\beta}}\big)$ can be done in time $O(ms)$. This is much faster than what is required with kernel-based estimators of conditional Kendall's tau ($O(mn^2)$), as soon as $s \le n^2$. See a deeper discussion of computation time in the supplementary material.
- Estimating Model (1) not only provides an estimator of the conditional Kendall's tau $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$, but also easily provides estimators of the marginal effects of $\mathbf{z}$ as by-product. For example, given $\mathbf{z} \in \mathcal{Z}$, the marginal effect of z_1, i.e., $\partial\tau_{1,2|\mathbf{Z}=\mathbf{z}}(\mathbf{z})/\partial z_1$, can be directly estimated by $\big(\partial_{z_1}\boldsymbol{\psi}(\mathbf{z})\big)^\top\hat{\boldsymbol{\beta}}/\Lambda'\big(\boldsymbol{\psi}(\mathbf{z})^\top\hat{\boldsymbol{\beta}}\big)$, assuming that $\boldsymbol{\psi}$ and Λ are differentiable respectively at $\mathbf{z}$ and $\boldsymbol{\psi}(\mathbf{z})^\top\hat{\boldsymbol{\beta}}$. Such sensitivities can be useful in many applications. They would be calculated a lot more easily than with "brute-force" non-parametric estimates.

In Section 2, we state non-asymptotic results for the estimator $\hat{\boldsymbol{\beta}}$ that hold with high probability. In Section 3, its asymptotic properties are stated. In particular, we will study the cases when n' is fixed and $n \to \infty$, and when both indices tend to the infinity (but still keeping p and p' as fixed). In the latter case, it is possible to reach the optimal rate of convergence, i.e., $\hat{\boldsymbol{\beta}}$ may become $\sqrt{n}$-asymptotically normal. We also give some oracle properties and suggest a related adaptive estimator. Section 4 illustrates the estimator $\hat{\boldsymbol{\beta}}$ on real data. The proofs have been postponed into the appendix. An extensive simulation study is provided as supplementary material.

Remark 1. Instead of a fixed design setting $(\mathbf{z}'_i)_{i\in\{1,\dots,n'\}}$ 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 (more or less easily) 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 (Section 3.1). When n and n' will tend to the infinity (Section 3.3), both designs are encompassed de facto because we will assume the weak convergence of the empirical distribution associated to the sample $(\mathbf{z}'_i)_{i\in\{1,\dots,n'\}}$, when $n' \to \infty$. In practical terms, we have a large amount of freedom in the choice of the $\mathbf{z}'_i$. In particular, it will not be necessary to select equally spaced $\mathbf{z}'_i$ points. A clever choice would surely be to select n' points among the initial sample, or to draw such points along an estimated distribution of the realizations $(\mathbf{z}_i)_{i\in\{1,\dots,n\}}$. But afterwards, our theoretical results will be valid "given such points $(\mathbf{z}'_i)_{i\in\{1,\dots,n'\}}$".

Remark 2. Note that the estimation of conditional Kendall's tau can also be seen as a classification problem in the space of weighted pairs of observations, leading to several alternative estimators that have been studied in [6]. They are of the type

$$\tilde{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{\beta}\in\mathbb{R}^{p'}}\left\{\frac{1}{n(n-1)}\sum_{i,j=1,i\neq j}^{n}\ell\left(g(\mathbf{X}_i,\mathbf{X}_j),\,\Lambda^{-1}\big(\boldsymbol{\psi}(\mathbf{z}'_i)^\top\boldsymbol{\beta}\big)\right)K_h(\mathbf{Z}_i-\mathbf{Z}_j)+\lambda|\boldsymbol{\beta}|_1\right\}, \tag{4}$$

for a particular map $g(\cdot) : \mathbb{R}^4 \to \{0, 1\}$ and for different loss functions ℓ. In particular, the latter approach does not require any choice of the set $(\mathbf{z}'_i)_{i\in\{1,\dots,n'\}}$ (fixed or random design).

2. Finite-distance bounds on $\hat{\boldsymbol{\beta}}$

Our first goal is to prove finite-distance bounds in probability for the estimator $\hat{\boldsymbol{\beta}}$. Let $\mathbf{Z}'$ be the matrix of size $n' \times p'$ whose lines are $\boldsymbol{\psi}(\mathbf{z}'_i)^\top$, $i \in \{1, \ldots, n'\}$, and let $\mathbf{Y} \in \mathbb{R}^{n'}$ be the column vector whose components are $Y_i = \Lambda(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i})$, $i \in \{1, \ldots, n'\}$. For a vector $\mathbf{v} \in \mathbb{R}^{n'}$, denote by $\|\mathbf{v}\|_{n'} := |\mathbf{v}|_2/\sqrt{n'}$ its empirical norm. We can then rewrite the criterion (2) as $\hat{\boldsymbol{\beta}} := \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{p'}} \Big[||\mathbf{Y} - \mathbf{Z}'\boldsymbol{\beta}||^2_{n'} + \lambda|\boldsymbol{\beta}|_1 \Big]$, where $\mathbf{Y}$ and $\mathbf{Z}'$ may be considered as "observed", so that the practical problem is reduced to a standard Lasso estimation procedure. Define some "residuals" by $\xi_{i,n} := \Lambda(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i}) - \boldsymbol{\psi}(\mathbf{z}'_i)^\top\boldsymbol{\beta}^* = \Lambda(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i}) - \Lambda(\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i})$, for $i \in \{1, \ldots, n'\}$. Note that these $\xi_{i,n}$ are not "true residuals" in the sense that they do not depend on the estimator $\hat{\boldsymbol{\beta}}$, but on the true parameter $\boldsymbol{\beta}^*$. 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 $\hat{\boldsymbol{\beta}}$, assume the *Restricted Eigenvalue* (RE) condition, introduced by [3]. For $c_0 > 0$ and $s \in \{1, \ldots, p\}$, assume

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

$$\kappa(s, c_0) := \min_{\substack{J_0 \subset \{1,\ldots,p'\} \\ Card(J_0) \le s}} \; \min_{\substack{\boldsymbol{\delta} \in \mathbb{R}^{p'}, \boldsymbol{\delta} \neq 0 \\ |\boldsymbol{\delta}_{J_0^C}|_1 \le c_0 |\boldsymbol{\delta}_{J_0}|_1}} \frac{|\mathbf{Z}'\boldsymbol{\delta}|_2}{\sqrt{n'}|\boldsymbol{\delta}|_2} > 0,$$

where for any vector $\boldsymbol{\delta} \in \mathbb{R}^{p'}$ and any $A \subset \{1, \ldots, p'\}$, we denote by $\boldsymbol{\delta}_A := (\delta_a)_{a \in A}$ the subvector of $\boldsymbol{\delta}$ that stacks the components whose index belongs to A. Similarly, for any $p \times q$ matrix $\mathbf{M}$, we will denote by $\mathbf{M}_{A,B}$ the sub-matrix whose rows i and columns j are restricted to $A \subset \{1, \ldots, p\}$ and $B \subset \{1, \ldots, q\}$ respectively. Note that this condition is very mild, and is satisfied with a high probability for a large class of random matrices: see [2, Section 8.1] for references and a discussion. Moreover, it precludes the size n' from being too small compared to the number of unknown parameters p'.

Assumption 1. The function $\mathbf{z} \mapsto \boldsymbol{\psi}(\mathbf{z})$ is bounded on $\mathscr{Z}$ by a constant $C_{\boldsymbol{\psi}}$. Moreover, $\Lambda(\cdot)$ is continuously differentiable. Let $\mathcal{T}$ be the range of $\mathbf{z} \mapsto \tau_{1,2|\mathbf{Z}=\mathbf{z}}$, from $\mathscr{Z}$ towards $[-1, 1]$. On an open neighborhood of $\mathcal{T}$, the derivative of $\Lambda(\cdot)$ is bounded by a constant $C_{\Lambda'}$.

Assumption 2. (a) The kernel K is bounded, and set $\|K\|_\infty =: C_K$. (b) It is symmetrical in the sense that $K(\mathbf{u}) = K(-\mathbf{u})$ for every $\mathbf{u} \in \mathbb{R}^p$ and satisfies $\int K = 1$, $\int |K| < \infty$, $\int K^2 < \infty$. (c) This kernel is of order α for some integer $\alpha > 1$: for all $j \in \{1, \ldots, \alpha - 1\}$ and every indices $i_1, \ldots, i_j \in \{1, \ldots, p\}$, $\int K(\mathbf{u}) \prod_{k=1}^j u_{i_k} \, d\mathbf{u} = 0$. (d) Moreover, $\mathrm{E}[K_h(\mathbf{Z} - \mathbf{z})] > 0$ for every $\mathbf{z} \in \mathscr{Z}$ and $h > 0$. Set $\tilde{K}(\cdot) := K^2(\cdot)/\int K^2$ and $\|\tilde{K}\|_\infty =: C_{\tilde{K}}$.

Assumption 3. $f_{\mathbf{Z}}$ is α-times continuously differentiable on an open neighborhood of $\mathscr{Z}$ and there exists a constant $C_{K,\alpha} > 0$ s.t., for all $\mathbf{z} \in \mathscr{Z}$,

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

Moreover, $C_{\tilde{K},2}$ denotes a similar constant replacing K by $\tilde{K}$ and α by two.

Assumption 4. There exist two positive constants $f_{\mathbf{Z},min}$ and $f_{\mathbf{Z},max}$ such that, for every $\mathbf{z} \in \mathscr{Z}$, $f_{\mathbf{Z},min} \le f_{\mathbf{Z}}(\mathbf{z}) \le f_{\mathbf{Z},max}$.

Assumption 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, $\mathbf{z} \in \mathscr{Z}$. For every $0 \le k \le \alpha$ and every $1 \le i_1, \ldots, i_\alpha \le p$, let

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

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

$$\int |K|(\mathbf{u})|K|(\mathbf{v}) \sum_{k=0}^{\alpha} \binom{\alpha}{k} \sum_{i_1,\ldots,i_\alpha=1}^{p} \mathcal{H}_{k,\vec{\imath}}(\mathbf{u}, \mathbf{v}, \mathbf{x}_1, \mathbf{x}_2, \mathbf{z}) \prod_{w=1}^{k} |u_{i_w}| \prod_{\ell=k+1}^{\alpha} |v_{i_\ell}| \, d\mathbf{u} \, d\mathbf{v} \, d\mathbf{x}_1 \, d\mathbf{x}_2 \le C_{\mathbf{XZ},\alpha}.$$

All the latter regularity conditions are relatively classical. The most demanding one is probably Assumption 4, but, by choosing $\mathscr{Z}$ as a compact subset that is strictly included in $\mathbf{Z}$'s support, this assumption is most of the time satisfied.

Theorem 2 (*Fixed Design Case*). *Suppose that Assumptions 1–5 hold and that the design matrix* $\mathbf{Z}'$ *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*

$$h^\alpha \le \min\Big(C_{h,1}\ ,\ tC_{h,2}\Big),\, C_{h,1} := \frac{f_{\mathbf{Z},min}\alpha!}{4\,C_{K,\alpha}},\, C_{h,2} := \frac{f^4_{\mathbf{Z},min}\alpha!}{8\,C_\psi C_{\Lambda'}(f^2_{\mathbf{Z},min}+8f^2_{\mathbf{Z},max})C_{\mathbf{XZ},\alpha}}\,. \tag{5}$$

Then, we have

$$\Pr\Big(\|\mathbf{Z}'(\hat{\boldsymbol\beta}-\boldsymbol\beta^*)\|_{n'} \le \frac{4(\gamma+1)\tilde t\sqrt{s}}{\kappa(s,3)},\ |\hat{\boldsymbol\beta}-\boldsymbol\beta^*|_q \le \frac{4^{2/q}(\gamma+1)\tilde t s^{1/q}}{\kappa^2(s,3)},\ \textit{for every } 1\le q\le 2\Big)$$
$$\ge 1-2n'\exp\bigg(-\frac{nh^p}{C_2}\bigg)-2n'\exp\bigg(-\frac{(n-1)h^{2p}t}{C_3+C_4t}\bigg)-2n'\exp\bigg(-\frac{nh^p(C_5-C_6h^2)^2}{C_7-C_8h^2}\bigg), \tag{6}$$

where $\tilde t := t + C_1/(nh^p)$ *and* $C_1,\dots,C_8$ *are constants independent of* n, n', h, γ, t *whose explicit values can be found in the proof.*

This theorem, proved in Appendix B, yields some bounds that hold in probability for the prediction error $\|\mathbf{Z}'(\hat{\boldsymbol\beta}-\boldsymbol\beta^*)\|_{n'}$ and for the estimation error $|\hat{\boldsymbol\beta}-\boldsymbol\beta^*|_q$, $1\le q\le 2$, 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 $\hat{\boldsymbol\beta}$ 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 (5): t is limited by a value proportional to h^α. The latter h cannot be chosen as too small, otherwise the probability in (6) will decrease. To be short: low values of h and t yield a sharper upper bound with a lower probability, and the opposite. Therefore, a trade-off has to be found, depending of the kind of result we are interested in.

Clearly, we would like to exhibit the sharpest upper bounds in (6), 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 2 imply $b\alpha\ge a$ (to satisfy (5)) and $1-2a-2pb>0$ (so that the right-hand side of (6) 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 $b\alpha\ge a$ 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 3. *For* $0<\epsilon<1$*, choose the parameters* $\lambda=4t$, $t=(n-1)^{-\alpha(1-\epsilon)/(2\alpha+2p)}$ *and*

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

and n *is sufficiently large so that* (5) *is satisfied. Then, under the assumptions of Theorem 2,*

$$\Pr\Big(\|\mathbf{Z}'(\hat{\boldsymbol\beta}-\boldsymbol\beta^*)\|_{n'} \le \frac{40\tilde t\sqrt{s}}{\kappa(s,3)}\ \textit{and}\ |\hat{\boldsymbol\beta}-\boldsymbol\beta^*|_q \le \frac{10\times 4^{2/q}\tilde t s^{1/q}}{\kappa^2(s,3)},\ \textit{for every } 1\le q\le 2\Big)$$
$$\ge 1-2n'\exp\bigg(-\frac{c_h^p(n-1)^{(2\alpha+p)/(2\alpha+2p)}}{C_2}\bigg)-2n'\exp\bigg(-\frac{c_h^{2p}(n-1)^{2\alpha\epsilon/(2p+2\alpha)}}{C_3+C_4(n-1)^{-\alpha(1-\epsilon)/(2\alpha+2p)}}\bigg)$$
$$-2n'\exp\bigg(-\frac{c_h^p(n-1)^{(2\alpha+p)/(2\alpha+2p)}(C_5-C_6h^2)^2}{C_7-C_8h^2}\bigg),$$

and $\tilde t = (n-1)^{-\alpha(1-\epsilon)/(2\alpha+2p)} + C_1c_h^{-p}(n-1)^{-(2\alpha+p)/(2\alpha+2p)}$.

The finite distance results of this section allow to build confidence intervals for $\boldsymbol\beta$ and then confidence bounds for the whole map $\mathbf{z}\mapsto\tau_{1,2|\mathbf{Z}=\mathbf{z}}$, even if the numerical evaluation of the many constants C_k, $k\in\{1,\dots,8\}$, is rather painful. For instance, under the assumptions and notations of Theorem 2, we know that

$$|\hat{\boldsymbol\beta}-\boldsymbol\beta^*|_q \le \frac{4^{2/q}(\gamma+1)\tilde t s^{1/q}}{\kappa^2(s,3)} := u(t,s),$$

with a probability larger than the r.h.s. of (6). By Hölder's inequality, $|\boldsymbol\psi(\mathbf{z})^\top(\hat{\boldsymbol\beta}-\boldsymbol\beta^*)|$ is smaller than $|\boldsymbol\psi(\mathbf{z})|_{q/(q-1)}|\hat{\boldsymbol\beta}-\boldsymbol\beta^*|_q$, for every $\mathbf{z}$ and $q>1$. We deduce that the true conditional Kendall's tau $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ belongs to the interval

$$\Big[\Lambda^{-1}\Big(\boldsymbol\psi(\mathbf{z})^\top\hat{\boldsymbol\beta}-|\boldsymbol\psi(\mathbf{z})|_{q/(q-1)}u(t,s)\Big),\ \Lambda^{-1}\Big(\boldsymbol\psi(\mathbf{z})^\top\hat{\boldsymbol\beta}+|\boldsymbol\psi(\mathbf{z})|_{q/(q-1)}u(t,s)\Big)\Big]$$

with the latter probability. Moreover, note that a similar bound applies for

$$\|\mathbf{Z}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_{n'}^2 = \frac{1}{(n')^2} \sum_{i=1}^{n'} \Big(\Lambda\big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i}\big) - \Lambda\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}\big) \Big)^2,$$

under our model (1). Nonetheless, this does not provide pointwise confidence intervals but rather the estimation of a L^2-distance between the true curve $\mathbf{z} \mapsto \Lambda(\tau_{1,2|\mathbf{Z}=\mathbf{z}})$ and the estimated one.

Remark 3. It is possible to obtain slightly sharper bound in Theorem 2, by invoking Proposition 4 in [7] instead of our Lemma 15. The price to be paid is to manage non explicit constants and then the impossibility of building confidence intervals. Therefore, we have preferred to keep our current result and some variants could easily be obtained by interested readers.

3. Asymptotic behavior of $\hat{\boldsymbol{\beta}}$

3.1. Asymptotic properties of $\hat{\boldsymbol{\beta}}$ 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 of $\hat{\boldsymbol{\beta}}$ 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 $\hat{\boldsymbol{\beta}}_{n,n'}$ the estimator (2) with $h = h_n$ and $\lambda = \lambda_{n,n'}$. The following lemma, proved in Appendix C, provides another representation of this estimator $\hat{\boldsymbol{\beta}}_{n,n'}$ that will be useful hereafter.

Lemma 4. *We have* $\hat{\boldsymbol{\beta}}_{n,n'} = \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{p'}} \mathbb{G}_{n,n'}(\boldsymbol{\beta})$, *where*

$$\mathbb{G}_{n,n'}(\boldsymbol{\beta}) := \frac{2}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol{\psi}(\mathbf{z}'_i)^\top (\boldsymbol{\beta}^* - \boldsymbol{\beta}) + \frac{1}{n'} \sum_{i=1}^{n'} \big\{ \boldsymbol{\psi}(\mathbf{z}'_i)^\top (\boldsymbol{\beta}^* - \boldsymbol{\beta}) \big\}^2 + \lambda_{n,n'} |\boldsymbol{\beta}|_1. \tag{7}$$

We will invoke a *convexity argument*: "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_∞" (see [11], e.g).

Theorem 5 (*Consistency of* $\hat{\boldsymbol{\beta}}$). *Under Assumption 2, if* $nh_n^p \to \infty$, $\lim K(\mathbf{t})|\mathbf{t}|^p = 0$ *when* $|\mathbf{t}| \to \infty$, $f_{\mathbf{Z}}$ *and* $\mathbf{z} \mapsto \tau_{1,2|\mathbf{Z}=\mathbf{z}}$ *are continuous on* $\mathscr{Z}$, *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,* $\hat{\boldsymbol{\beta}}_{n,n'} \xrightarrow{\mathbb{P}} \boldsymbol{\beta}^{**} := \inf_{\boldsymbol{\beta}} \mathbb{G}_{\infty,n'}(\boldsymbol{\beta})$, *where* $\mathbb{G}_{\infty,n'}(\boldsymbol{\beta}) := \sum_{i=1}^{n'} \big(\boldsymbol{\psi}(\mathbf{z}'_i)^\top (\boldsymbol{\beta}^* - \boldsymbol{\beta}) \big)^2 / n' + \lambda_0 |\boldsymbol{\beta}|_1$ *is assumed to be strictly convex. In particular, if* $\lambda_0 = 0$ *and* $\langle \boldsymbol{\psi}(\mathbf{z}'_1), \dots, \boldsymbol{\psi}(\mathbf{z}'_{n'}) \rangle = \mathbb{R}^{p'}$, *then* $\hat{\boldsymbol{\beta}}_{n,n'} \xrightarrow{\mathbb{P}} \boldsymbol{\beta}^*$.

Proof. By Lemma 16, the first term in the r.h.s. of (7) converges to 0 as $n \to \infty$. The third term in the r.h.s. of (7) converges to $\lambda_0 |\boldsymbol{\beta}|_1$ 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, $\arg\min_{\boldsymbol{\beta}} \mathbb{G}_{n,n'}(\boldsymbol{\beta}) \to \arg\min_{\boldsymbol{\beta}} \mathbb{G}_{\infty,n'}(\boldsymbol{\beta})$ in law. Since we have adopted a fixed design setting, $\boldsymbol{\beta}^{**}$ 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 $\lambda_0 = 0$, $\boldsymbol{\beta}^*$ is the minimum of $\mathbb{G}_{\infty,n'}$ because the vectors $\boldsymbol{\psi}(\mathbf{z}'_i)$, $i \in \{1, \dots, p'\}$ generate the space $\mathbb{R}^{p'}$. Therefore, this implies the consistency of $\hat{\boldsymbol{\beta}}_{n,n'}$. □

Note that the convergence of $\hat{\boldsymbol{\beta}}_{n,n'}$ towards $\boldsymbol{\beta}^*$ necessitates $n' \geq p'$, but the number of design points does not need to grow with n to achieve consistency. To evaluate the limiting behavior of $\hat{\boldsymbol{\beta}}_{n,n'}$, 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 by the following corollary of Lemma 17, for which we need the following assumption.

Assumption 6. (i) $nh_n^p \to \infty$ and $nh_n^{p+2\alpha} \to 0$; (ii) $K(\cdot)$ is compactly supported. (iii) the $\mathbf{z}'_i$ are distinct. (iv) $f_{\mathbf{Z}}$ and $\mathbf{z} \mapsto f_{\mathbf{X},\mathbf{Z}}(\mathbf{x}, \mathbf{z})$ are continuous on $\mathscr{Z}$, for every $\mathbf{x}$.

Corollary 6. *Under Assumptions 2, 5 and 6,* $(nh_n^p)^{1/2} \big[\xi_{1,n}, \dots, \xi_{n',n} \big]^\top$ *tends in law towards a random vector* $\mathcal{N}\big(0, \tilde{\mathbf{H}}\big)$ *given* $(\mathbf{z}'_1, \dots, \mathbf{z}'_{n'})$, *where* $\tilde{\mathbf{H}}$ *is a* $n' \times n'$ *real matrix defined, for every integers* $1 \leq i, j \leq n'$, *by*

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

where $\tilde{g}$ *is the symmetrized version* $\tilde{g}(\mathbf{x}_1, \mathbf{x}_2) := (g^*(\mathbf{x}_1, \mathbf{x}_2) + g^*(\mathbf{x}_2, \mathbf{x}_1))/2$.

Theorem 7 (*Asymptotic Law of the Estimator*). *Under Assumptions 2, 5 and 6, and if $\lambda_{n,n'}(nh^p_{n,n'})^{1/2}$ tends to ℓ when $n \to \infty$, we have $(nh^p_{n,n'})^{1/2}(\hat{\boldsymbol{\beta}}_{n,n'} - \boldsymbol{\beta}^*) \xrightarrow{D} \mathbf{u}^* := \arg\min_{\mathbf{u}\in\mathbb{R}^{p'}} \mathbb{F}_{\infty,n'}(\mathbf{u})$ given $\mathbf{z}'_1, \ldots, \mathbf{z}'_{n'}$, where*

$$\mathbb{F}_{\infty,n'}(\mathbf{u}) := \frac{2}{n'}\sum_{i=1}^{n'}\sum_{j=1}^{p'} W_i\psi_j(\mathbf{z}'_i)u_j + \frac{1}{n'}\sum_{i=1}^{n'}\left(\boldsymbol{\psi}(\mathbf{z}'_i)^\top\mathbf{u}\right)^2 + \ell\sum_{i=1}^{p'}\left(|u_i|\mathbb{1}_{\{\beta^*_i=0\}} + u_i\ \mathrm{sign}(\boldsymbol{\beta}^*_i)\mathbb{1}_{\{\beta^*_i\neq 0\}}\right)$$

is assumed to be strictly convex a.s., with $\mathbf{W} = (W_1, \ldots, W_{n'}) \sim \mathcal{N}\left(0, \tilde{\mathbf{H}}\right)$.

This theorem is proved in Appendix C. 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 $\sum_{i=1}^{n'} W_i\boldsymbol{\psi}(\mathbf{z}'_i) + \sum_{i=1}^{n'}\boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top\mathbf{u} = 0$. Therefore,

$$\mathbf{u}^* = -\Big(\sum_{i=1}^{n'}\boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top\Big)^{-1}\sum_{i=1}^{n'} W_i\boldsymbol{\psi}(\mathbf{z}'_i),$$

when $\Sigma_{n'} := \sum_{i=1}^{n'}\boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top$ is invertible. Then, the limiting law of $(nh^p_{n,n'})^{1/2}(\hat{\boldsymbol{\beta}}_{n,n'} - \boldsymbol{\beta}^*)$ is Gaussian, and its asymptotic covariance is $\mathbf{V}_{as} := \Sigma_{n'}^{-1}\sum_{i,j=1}^{n'}[\tilde{\mathbf{H}}]_{i,j}\boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_j)^\top\Sigma_{n'}^{-1}$.

The previous results on the asymptotic normality of $\hat{\boldsymbol{\beta}}_{n,n'} - \boldsymbol{\beta}^*$ can be used to test $\mathcal{H}_0 : \boldsymbol{\beta}^* = 0$ against the opposite. As said in the introduction, this would constitute a test of the "simplifying assumption", i.e., the fact that the conditional copula of (X_1, X_2) given $\mathbf{Z}$ does not depend on this covariate. Some tests of significance of $\boldsymbol{\beta}^*$ would be significantly simpler than most of the tests of the simplifying assumption that have been proposed in the literature until now. Indeed, the latter ones have been built on nonparametric estimates of conditional copulas and, as sub-products of the weak convergence of the associated processes, the test statistics behaviors are obtained. Therefore, such statistics depend on a preliminary non-parametric estimation of conditional marginal distributions (see [5,15], e.g.), a source of complexities and statistical noise. At the opposite, some tests of $\mathcal{H}_0$ based on $\hat{\boldsymbol{\beta}}_{n,n'}$ do not require this stage, at the cost of a (probably small) loss of power. For instance, in the case of $\ell = 0$, we propose the Wald-type test statistics

$$\mathcal{W}_n := nh^p_{n,n'}(\hat{\boldsymbol{\beta}}_{n,n'} - \boldsymbol{\beta}^*)^\top\mathbf{V}_n(\hat{\boldsymbol{\beta}}_{n,n'} - \boldsymbol{\beta}^*),\ \mathbf{V}_n := \Sigma_{n'}^{-1}\sum_{i,j=1}^{n'}\hat{H}_{i,j}\boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_j)^\top\Sigma_{n'}^{-1},$$

$$\hat{H}_{i,j} := \frac{4\int K^2\mathbb{1}_{\{\mathbf{z}'_i=\mathbf{z}'_j\}}}{\hat{f}_{\mathbf{Z}}(\mathbf{z}'_i)}\left(\Lambda'\left(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i}\right)\right)^2\left\{\mathcal{G}_n(\mathbf{z}'_i) - \hat{\tau}^2_{1,2|\mathbf{Z}=\mathbf{z}'_i}\right\},$$

where $\hat{f}_{\mathbf{Z}}(\mathbf{Z})$ and $\mathcal{G}_n(\mathbf{z})$ denote consistent estimators of $f_{\mathbf{Z}}(\mathbf{z})$ and $\mathrm{E}[\tilde{g}(\mathbf{X}_1, \mathbf{X})\tilde{g}(\mathbf{X}_2, \mathbf{X})|\mathbf{Z} = \mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}]$ respectively. Under $\mathcal{H}_0$, $\mathcal{W}_n$ tends to a chi-square distribution with n' degrees of freedom. For instance, with the same notations as in Section 1, we propose

$$\mathcal{G}_n(\mathbf{z}) = \sum_{i,j,k=1, i\neq j\neq k}^{n} w_{i,n}(\mathbf{z})w_{j,n}(\mathbf{z})w_{k,n}(\mathbf{z})\tilde{g}(\mathbf{X}_i, \mathbf{X}_k)\tilde{g}(\mathbf{X}_j, \mathbf{X}_k).$$

Note that if there is an intercept, i.e., if one of the functions in $\boldsymbol{\psi}$ (say, ψ_1) is constant to 1, it should be removed in the statistics above. The corresponding coefficients of $\hat{\boldsymbol{\beta}}$ should be removed as well. Indeed, in this case the simplifying assumption does not correspond to $\boldsymbol{\beta}^* = 0$, but rather to $\boldsymbol{\beta}^*_{-1} = 0$ where $\boldsymbol{\beta}^*_{-i}$ denotes the vector $\boldsymbol{\beta}^*$ where the ith coefficient has been removed.

3.2. Oracle property and a related adaptive procedure

Let remember that $\mathscr{S} := \{j : \boldsymbol{\beta}^*_j \neq 0\}$ and assume that $|\mathscr{S}| = s < p$ so that the true model depends on a subset of predictors. In the same spirit as [8], we say that an estimator $\hat{\boldsymbol{\beta}}$ satisfies the oracle property if

1. the limiting law of $\hat{\boldsymbol{\beta}}_{\mathscr{S}} - \boldsymbol{\beta}^*_{\mathscr{S}}$ is the same as if the "true" model containing only predictors from $\mathscr{S}$ was chosen and estimation was performed only on that model;
2. we identify the nonzero components of the true parameter $\boldsymbol{\beta}^*$ with probability one when the sample size n is large, i.e., the probability of the event $\left(\{j : \hat{\boldsymbol{\beta}}_j \neq 0\} = \mathscr{S}\right)$ tends to one.

As above, let us fix n' and n will tend to the infinity. Then, denote $\{j : \hat{\boldsymbol{\beta}}_j \neq 0\}$ by $\mathscr{S}_n$, that will implicitly depend on n'. It is well-known that the usual Lasso estimator does not fulfill the oracle property, see [16]. Here, this is still the case. The following proposition is proved in Appendix C.

Proposition 8. *Under Assumptions 2, 5 and 6,* $\limsup_n \Pr(\mathscr{S}_n = \mathscr{S}) = c < 1$.

A usual way of obtaining the oracle property is to modify our estimator in an "adaptive" way. Following [16], consider a preliminary "rough" estimator of $\boldsymbol{\beta}^*$, denoted by $\tilde{\boldsymbol{\beta}}_n$, or more simply $\tilde{\boldsymbol{\beta}}$. Moreover $\nu_n(\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta}^*)$ is assumed to be asymptotically normal, for some deterministic sequence (ν_n) that tends to the infinity. Our adaptive estimator will be denoted by $\check{\boldsymbol{\beta}}_{n,n'}$ (or simply $\check{\boldsymbol{\beta}}$). It is defined as

$$\check{\boldsymbol{\beta}}_{n,n'} := \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{p'}} \left\{ \frac{1}{n'} \sum_{i=1}^{n'} \left(\Lambda(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i}) - \boldsymbol{\psi}(\mathbf{z}'_i)^\top \boldsymbol{\beta} \right)^2 + \mu_{n,n'} \sum_{j=1}^{p'} \frac{|\boldsymbol{\beta}_j|}{|\tilde{\boldsymbol{\beta}}_{n,j}|^\delta} \right\}, \tag{8}$$

for some constant $\delta > 0$ and some positive deterministic sequence $(\mu_{n,n'})$. Hereafter, we still set $\mathscr{S}_n = \{j : \check{\boldsymbol{\beta}}_j \neq 0\}$. The following theorem is proved in Appendix C.

Theorem 9 (*Asymptotic Law of the Adaptive Estimator of* $\boldsymbol{\beta}$). *Under Assumptions 2, 5 and 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*

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

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

is assumed to be strictly convex a.s., with $\mathbf{W} = (W_1, \ldots, W_{n'}) \sim \mathcal{N}\big(0, \tilde{\mathbf{H}}\big)$*. Moreover, when* $\ell = 0$*, the oracle property is fulfilled:* $\Pr(\mathscr{S}_n = \mathscr{S}) \to 1$ *as* $n \to \infty$.

Note that the latter rates of convergence of $\hat{\boldsymbol{\beta}}_{n,n'}$ and $\check{\boldsymbol{\beta}}_{n,n'}$ were "sub-optimal": for a fixed size n', these parameters cannot be estimated at a $\sqrt{n}$ rate. In the next section, we show the latter optimal rate can be achieved by allowing an increasing size n'.

3.3. Asymptotic properties of $\hat{\boldsymbol{\beta}}$ 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 $\mathbf{z}'_1, \ldots, \mathbf{z}'_{n'}, \ldots$. 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 10 (*Consistency of* $\hat{\boldsymbol{\beta}}_{n,n'}$*, Jointly in* (n, n')). *Assume that Assumptions 1–5 are satisfied. Assume that* $\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top / n'$ *converges to a matrix* $\mathbf{M}_{\psi,\mathbf{z}'}$*, 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* $\hat{\boldsymbol{\beta}}_{n,n'} \xrightarrow{\mathbb{P}} \arg\min_{\boldsymbol{\beta} \in \mathbb{R}^{p'}} \mathbb{G}_{\infty,\infty}(\boldsymbol{\beta})$*, as* $(n, n') \to \infty$*, where* $\mathbb{G}_{\infty,\infty}(\boldsymbol{\beta}) := (\boldsymbol{\beta}^* - \boldsymbol{\beta})\mathbf{M}_{\psi,\mathbf{z}'}(\boldsymbol{\beta}^* - \boldsymbol{\beta})^\top + \lambda_0|\boldsymbol{\beta}|_1$ *is assumed to be strictly convex. Moreover, if* $\lambda_0 = 0$ *and* $\mathbf{M}_{\psi,\mathbf{z}'}$ *is invertible, then* $\hat{\boldsymbol{\beta}}_{n,n'}$ *is consistent and tends to the true value* $\boldsymbol{\beta}^*$.

The proof of this theorem is provided in Appendix C. Note that, since the sequence $(\mathbf{z}'_i)$ is deterministic, we just assume the usual convergence of $\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top / n'$ in $\mathbb{R}^{p'^2}$. Moreover, if the "second subset" $(\mathbf{z}'_i)_{i \in \{1,\ldots,n'\}}$ were a random sample (drawn along the law $\Pr_{\mathbf{Z}}$), the latter convergence would be understood "in probability". And if $\Pr_{\mathbf{Z}}$ satisfies the identifiability condition (Proposition 1), then $\mathbf{M}_{\psi,\mathbf{z}'}$ would be invertible and $\hat{\boldsymbol{\beta}}_{n,n'} \to \boldsymbol{\beta}^*$ in probability. Now, we want to go one step further and derive the asymptotic law of the estimator $\hat{\boldsymbol{\beta}}_{n,n'}$.

Assumption 7.

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

(ii) (a) $n'(nh^{p+4\alpha}_{n,n'} + h^{2\alpha}_{n,n'} + (nh^p_{n,n'})^{-1}) \to 0$, (b) $\lambda_{n,n'}(n'\, n\, h^p_{n,n'})^{1/2} \to 0$, (c) $n\, h^{p+\alpha}_{n,n'} / \ln n' \to \infty$.

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

(iv) The matrix $\mathbf{V}_1 := \int \boldsymbol{\psi}(\mathbf{z}')\boldsymbol{\psi}(\mathbf{z}')^\top f_{\mathbf{z}',\infty}(\mathbf{z}')d\mathbf{z}'$ is non-singular.

(v) $\Lambda(\cdot)$ is two times continuously differentiable. Let $\mathcal{T}$ be the range of $\mathbf{z} \mapsto \tau_{1,2|\mathbf{Z}=\mathbf{z}}$, from $\mathscr{Z}$ towards $[-1, 1]$. On an open neighborhood of $\mathcal{T}$, the second derivative of $\Lambda(\cdot)$ is bounded by a constant $C_{\Lambda''}$.

Part (i) of the latter assumption forbids the design points $(\mathbf{z}'_i)_{i\geq 1}$ from being too close to each other and too fast, with respect to the rate of convergence $(h_{n,n'})$ to 0. This can be guaranteed by choosing an appropriate design. For example, if $p = 1$ and $\mathscr{Z} = [0, 1]$, 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 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. Interestingly, it is always possible to choose the asymptotically optimal bandwidth, i.e., $h \propto n^{-1/(2\alpha+p)}$. In this case, we can set $n' = n^a$, with any $a \in]0, 2\alpha/(2\alpha + p)[$ and the constraints are satisfied.

The design points $\mathbf{z}'_i$ are deterministic, similarly to all results in the present paper. For a given n', we can invoke the non-random measure $\mathrm{Pr}_{\mathbf{z}',n'} := n'^{-1} \sum_{i=1}^{n'} \delta_{z'_i}$. Equivalently, all results can be seen as given conditionally on the sample $(\mathbf{z}'_i)_{i\geq 1}$. In (iii), we impose the weak convergence of $\mathrm{Pr}_{\mathbf{z}',n'}$ to a measure that is absolutely continuous w.r.t. the Lebesgue measure. Intuitively, this means we do not want to observe some design points that would be repeated infinitely often (this would result in a Dirac component in $\mathrm{Pr}_{\mathbf{z}',\infty}$). An optimal choice of the density $f_{\mathbf{z}',\infty}$ is not an easy task. Indeed, even if we knew exactly the true density $f_{\mathbf{Z}}$, there is no obvious reasons why we should select the $\mathbf{z}'_i$ along $f_{\mathbf{Z}}$ (at least in the limit). If we want a small asymptotic variance $\tilde{\mathbf{V}}_{as}$ (see below), 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 $\boldsymbol{\psi}(\mathbf{z}')\boldsymbol{\psi}(\mathbf{z}')^\top$ is big.

Part (iv) of the assumption is usual, and ensures that the design is somehow "asymptotically full rank". This matrix $\mathbf{V}_1$ will also appear in the asymptotic variance of $\hat{\boldsymbol{\beta}}_{n,n'}$.

Part (v) allows us to control a remainder term in a Taylor expansion of Λ. Notice that this technical assumption was not necessary in the previous section, where we used the Delta-method on the vector $(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{z}'_i})_{i\in\{1,\dots,n'\}}$. But when the number of terms n' tends to infinity, we have to invoke second derivatives to control remainder terms.

The proof of the next theorem is provided in Appendix D.

Theorem 11 (*Asymptotic Law of $\hat{\boldsymbol{\beta}}_{n,n'}$, Jointly in (n, n')*)**.** *Under Assumptions 2–5 and 7, we have*

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

where $\tilde{\mathbf{V}}_{as} := \mathbf{V}_1^{-1}\mathbf{V}_2\mathbf{V}_1^{-1}$, $\mathbf{V}_1$ is the matrix defined in Assumption 7(iv), and

$$\mathbf{V}_2 := \int K^2 \int (\tilde{g}(\mathbf{x}_1, \mathbf{x}_3)\tilde{g}(\mathbf{x}_2, \mathbf{x}_3) - \tau_{1,2|\mathbf{z}'_1=\mathbf{z}'_2=\mathbf{z}})\Lambda'\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\big)^2\boldsymbol{\psi}(\mathbf{z})\boldsymbol{\psi}(\mathbf{z})^\top$$
$$\times f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{Z} = \mathbf{z})f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{Z} = \mathbf{z})f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_3|\mathbf{Z} = \mathbf{z})\frac{f_{\mathbf{z}',\infty}(\mathbf{z})}{f_{\mathbf{Z}}(\mathbf{z})}\, d\mathbf{x}_1\, d\mathbf{x}_2\, d\mathbf{x}_3\, d\mathbf{z}'.$$

Let us identify the triplets (n', h, λ) so that the assumptions of the latter theorem are fulfilled and the rate of convergence of $\hat{\boldsymbol{\beta}}_{n,n'}$ is maximal. To fix the ideas, set $n' = n^a/\ln n$, $a > 0$, $h_{n,n'} = n^{-b}$, $b > 0$ and $\lambda_{n,n'} = n^{-c}$, $c > 0$. Then, the constraints (i) and (ii) of Assumption 7 are rewritten as $a + 1 \leq b(p + 4\alpha)$, $a \leq \min(2\alpha, p)b$, $b(p + \alpha) < 1$, $a + 1 - bp \leq 2c$ and $a \leq 1 - bp$.

We deduce some triplets (a, b, c) that satisfy the latter constraints and for which the rate of convergence of $\hat{\boldsymbol{\beta}}_{n,n'}$ is $\sqrt{n}$, i.e. is optimal.

Corollary 12. *Set $\alpha = p$, $n' = n^{a^*}/\ln n$, $h_{n,n'} = n^{-b^*}$ and $\lambda_{n,n'} = n^{-c^*}$ with $(a^*, b^*, c^*) := \big(1/4, 1/(4p), 1/2 + \epsilon\big)$ for some $\epsilon > 0$. Under the assumptions of Theorem 11, $n^{1/2}(\hat{\boldsymbol{\beta}}_{n,n'} - \boldsymbol{\beta}^*) \xrightarrow{D} \mathcal{N}(\mathbf{0}, \tilde{\mathbf{V}}_{as})$.*

In Section 3.1, we proved that the rate of convergence of $\hat{\boldsymbol{\beta}}_{n,n'}$ is $1/\sqrt{nh^p}$ if the size n' is fixed. Basically, this could be seen as a consequence of the Delta-method: if $\sqrt{nh^p}\big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}} - \tau_{1,2|\mathbf{Z}=\mathbf{z}}\big)$ is asymptotically normal when $n \to \infty$, this would be the case of $\sqrt{nh^p}(\hat{\boldsymbol{\beta}}_{n,n'} - \boldsymbol{\beta})$ since $\hat{\boldsymbol{\beta}} = \phi_{n'}\big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_1}, \dots, \hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_{n'}}\big)$ with a fixed n' and for some function $\phi_{n'}$. Nonetheless, we now obtain the parametric rate $\sqrt{n}$ when n and n' both tend to the infinity. In this case, the Delta-Method cannot be applied because we have to consider many different transforms $\phi_{n'}$. The parametric rate is then the result of a usual kind of "averaging process" w.r.t. the $\mathbf{z}'_i$, $i \in \{1, \dots, n'\}$, when n' tend to the infinity, but not too quickly compared to n (Assumption 7(ii)).

The asymptotic normality results that we proved in Section 3 easily allow to build pointwise asymptotic confidence intervals. For example, assume $\ell = 0$ and denote by Φ the cdf of a standard Gaussian r.v. and $\sigma(\mathbf{z})^2 := \boldsymbol{\psi}(\mathbf{z})^\top\mathbf{V}_{as}\boldsymbol{\psi}(\mathbf{z})$. Then, by Theorem 7 and for every $\mathbf{z}$, $\boldsymbol{\psi}(\mathbf{z})^\top\boldsymbol{\beta}^*$ belongs to the interval $\big[\boldsymbol{\psi}(\mathbf{z})^\top\hat{\boldsymbol{\beta}}_{n,n'} \pm \Phi^{-1}(\epsilon/2)(nh^p_{n,n'})^{-1/2}\sigma(\mathbf{z})\big]$ with a probability close to $1 - \epsilon$ when $n \to \infty$, $\epsilon \in (0, 1)$. As a consequence, $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ belongs to the interval

$$\Big[\Lambda^{-1}\big(\boldsymbol{\psi}(\mathbf{z})^\top\hat{\boldsymbol{\beta}}_{n,n'} + \Phi^{-1}(\epsilon/2)(nh^p_{n,n'})^{-1/2}\sigma(\mathbf{z})\big)\,,\ \Lambda^{-1}\big(\boldsymbol{\psi}(\mathbf{z})^\top\hat{\boldsymbol{\beta}}_{n,n'} - \Phi^{-1}(\epsilon/2)(nh^p_{n,n'})^{-1/2}\sigma(\mathbf{z})\big)\Big],$$

with a probability close to $1 - \epsilon$, when $n \to \infty$. Similar confidence intervals can be obtained by applying every latter asymptotic normality result. The details are left to the reader.

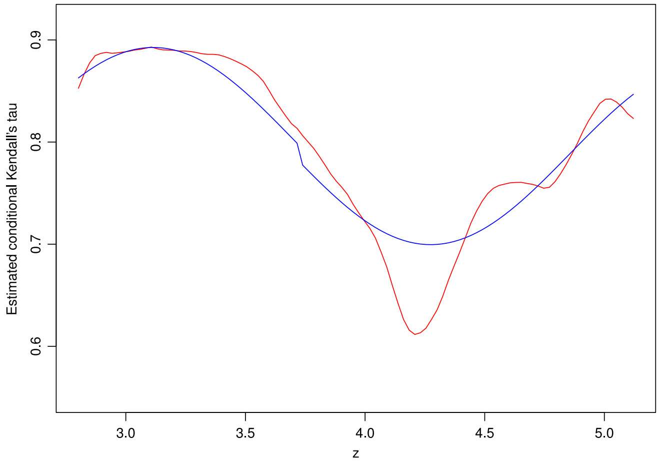

Fig. 1. Estimated conditional Kendall's tau $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$ (red curve), and prediction $\Lambda^{-1}\big(\boldsymbol{\psi}(\mathbf{z})^\top \hat{\boldsymbol{\beta}}\big)$ (blue curve) as a function of $\mathbf{z}$ for the application on real data, where the estimated non-zero coefficients are $\hat{\boldsymbol{\beta}}_1 = 0.78$, $\hat{\boldsymbol{\beta}}_7 = -0.043$, $\hat{\boldsymbol{\beta}}_8 = 0.069$ and $\hat{\boldsymbol{\beta}}_{11} = 0.020$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4. A 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 studied in [9].

We use $n' = 100$, $h = 2\sigma(Z)n^{-1/5}$. 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 $i \in \{1,\dots,5\}$, $\psi_{5+2i}(z) = \cos(2i\pi z)$ and $\psi_{6+2i}(z) = \sin(2i\pi z)$ for $i \in \{1,2\}$, $\psi_{11}(z) = \mathbb{1}\{z \le 0.4\}$, $\psi_{12}(z) = \mathbb{1}\{z \le 0.6\}$. Then, they cover a mix of polynomial, trigonometric and step-functions. The latter functions will be composed with a linear transform to be defined on $[\min(Z), \max(Z)]$). The tuning parameter λ is chosen by the cross-validation algorithm detailed in the supplementary material. The results are displayed in Fig. 1. As expected, the levels of conditional dependence between male and female expectancies are strong overall. Many poor countries suffer from epidemics, malnutrition or even wars. In such cases, life expectancies of both genders are exposed to the same "exogenous" factors, inducing high Kendall's taus. Logically, we observe a monotonic decrease of such Kendall's taus when Z is larger, up to $Z \simeq 4.5$, as already noticed by [9]. Indeed, when countries become richer, more developed and safe, men and women less and less depend on their environment (and on its risks of death, potentially). Nonetheless, when Z become even larger (the richest countries in the world), conditional dependencies between male and female life expectancies interestingly increase again, because men and women behave similarly in terms of way of life. In particular, they can benefit from the same levels of security and health and are exposed to the same lethal risks.

CRediT authorship contribution statement

Alexis Derumigny: Conceptualization, Methodology, Writing - review & editing, Software. **Jean-David Fermanian:** Conceptualization, Methodology, Writing - review & editing.

Acknowledgments

We thank the Editor, the Associate Editor and some referees whose remarks have significantly improved this work. The authors are grateful for helpful discussions with Christian Francq, Johanna Neslehová, Ostap Okhrin, Olivier Scaillet, Alexandre Tsybakov, Jean-Michel Zakoïan, and the participants at the "Copulas and their Applications" workshop (Almeria, Spain, 2017), the Computational and Financial Econometrics 2017 congress, the CREST Financial Econometrics seminar (Feb.2018), the TUM Colloquium in probability (Jan.2019), the seminar of the Geneva Research Center for Statistics (March 2019) and the Dresdner Mathematisches Seminar (April 2019). The authors have been supported by the labex Ecodec (reference project ANR-11-LABEX-0047).

Appendix A. Technical results concerning the first-step estimator

Three possible choices for g^* are given in [7]

$$g_1(\mathbf{X}_i, \mathbf{X}_j) := 4 \cdot \mathbb{1}\big\{X_{i,1} < X_{j,1}, X_{i,2} < X_{j,2}\big\} - 1,$$
$$g_2(\mathbf{X}_i, \mathbf{X}_j) := \mathbb{1}\big\{(X_{i,1} - X_{j,1}).(X_{i,2} - X_{j,2}) > 0\big\} - \mathbb{1}\big\{(X_{i,1} - X_{j,1}).(X_{i,2} - X_{j,2}) < 0\big\},$$
$$g_3(\mathbf{X}_i, \mathbf{X}_j) := 1 - 4 \cdot \mathbb{1}\big\{X_{i,1} \langle X_{j,1}, X_{i,2} \rangle X_{j,2}\big\},$$

where $\mathbb{1}$ is the indicator function. In the following, we assume that we have chosen g^* as one of the g_k for a fixed $k \in \{1, 2, 3\}$. We now recall some useful lemmas which have been proved in [7].

Lemma 13 (*Lemma 12 in [7]*). *Under Assumptions 2–4, for any $t > 0$ and any $\mathbf{z} \in \mathscr{Z}$, the estimator $\hat f_{\mathbf{Z}}(\mathbf{z}) := n^{-1}\sum_{i=1}^n K_h(\mathbf{Z}_i - \mathbf{z})$ satisfies*

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

Lemma 14 (*Proposition 2 in [7]*). *Under Assumptions 2–4 and if $C_{K,\alpha}h^\alpha/\alpha! < f_{\mathbf{Z},min}$, for any $\mathbf{z} \in \mathscr{Z}$, the estimator $\hat f_{\mathbf{Z}}(\mathbf{z})$ is strictly positive with a probability larger than*

$$1 - 2\exp\Big(-nh^p\big(f_{\mathbf{Z},min} - C_{K,\alpha}h^\alpha/\alpha!\big)^2 \big/ \big(2 f_{\mathbf{Z},max}\int K^2 + (2/3)C_K(f_{\mathbf{Z},min} - C_{K,\alpha}h^\alpha/\alpha!)\big)\Big).$$

Lemma 15 (*Proposition 3 in [7]*). *Under Assumptions 2–5, for every $t_1 > 0$ such that $C_{K,\alpha}h^\alpha/\alpha! + t_1 \leq f_{\mathbf{Z},min}/2$, every $t_2 > 0$ and for any $\mathbf{z} \in \mathscr{Z}$, we have*

$$\Pr\left(\left|\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}} - \tau_{1,2|\mathbf{Z}=\mathbf{z}}\right| > \frac{c_k}{f^2_{\mathbf{Z}}(\mathbf{z})}\left(\frac{C_{\mathbf{XZ},\alpha}h^\alpha}{\alpha!} + \frac{3 f_{\mathbf{Z}}(\mathbf{z})\int K^2}{2nh^p} + t_2\right) \times \left(1 + \frac{16 f^2_{\mathbf{Z}}(\mathbf{z})}{f^3_{\mathbf{Z},min}}\left(\frac{C_{K,\alpha}h^\alpha}{\alpha!} + t_1\right)\right)\right)$$
$$\leq 2\exp\left(-\frac{nh^p t_1^2}{2 f_{\mathbf{Z},max}\int K^2 + (2/3)C_K t_1}\right) + 2\exp\left(-\frac{(n-1)h^{2p}t_2^2}{4 f^2_{\mathbf{Z},max}(\int K^2)^2 + (8/3)C_K^2 t_2}\right)$$
$$+ 2\exp\left(-\frac{nh^p(f_{\mathbf{Z}}(\mathbf{z}) - C_{\tilde K,2}h^2)^2}{8 f_{\mathbf{Z},max}\int \tilde K^2 + 4 C_{\tilde K}(f_{\mathbf{Z}}(\mathbf{z}) - C_{\tilde K,2}h^2)/3}\right),$$

with $c_1 := c_3 := 4$ and $c_2 := 2$.

Lemma 16 (*Consistency, Proposition 7 in [7]*). *Under Assumption 2, if $nh_n^p \to \infty$, $\lim K(\mathbf{t})|\mathbf{t}|^p = 0$ when $|\mathbf{t}| \to \infty$, $f_{\mathbf{Z}}$ and $\mathbf{z} \mapsto \tau_{1,2|\mathbf{Z}=\mathbf{z}}$ are continuous on $\mathscr{Z}$, then for any $\mathbf{z} \in \mathscr{Z}$, $\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ tends to $\tau_{1,2|\mathbf{Z}=\mathbf{z}}$ in probability, when $n \to \infty$.*

Lemma 17 (*Joint Asymptotic Normality, Proposition 9 in [7]*). *Under Assumptions 2, 5, 6, for any $\mathbf{z}'_1, \dots, \mathbf{z}'_i$, we have $(nh_n^p)^{1/2}\big(\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}\big)_{i\in\{1,\dots,n'\}} \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{H})$ as $n \to \infty$, where $\mathbf{H}$ is a $n' \times n'$ real matrix defined by*

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

for every $1 \leq i, j \leq n'$, and $(\mathbf{X}, \mathbf{Z})$, $(\mathbf{X}_1, \mathbf{Z}_1)$, $(\mathbf{X}_2, \mathbf{Z}_2)$ are independent copies, where $\tilde g$ is the symmetrized version $\tilde g(\mathbf{x}_1, \mathbf{x}_2) := \big(g^*(\mathbf{x}_1, \mathbf{x}_2) + g^*(\mathbf{x}_2, \mathbf{x}_1)\big)/2$.

Appendix B. Proofs of finite-distance results for $\hat{\boldsymbol\beta}$

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

Lemma 18. *We have $\|\mathbf{Z}'\mathbf{u}\|^2_{n'} \leq \lambda|\mathbf{u}|_1 + \frac{1}{n'}\big\langle \boldsymbol\xi\ ,\ \mathbf{Z}'\mathbf{u}\big\rangle$.*

Proof. As $\hat{\boldsymbol\beta}$ is optimal, through the Karush–Kuhn–Tucker conditions, we have $(1/n')\mathbf{Z}'^\top(\mathbf{Y} - \mathbf{Z}'\hat{\boldsymbol\beta}) \in \partial\big(\lambda|\hat{\boldsymbol\beta}|_1\big)$, where $\partial\big(\lambda|\hat{\boldsymbol\beta}|_1\big)$ is the subdifferential of the norm $\lambda|\cdot|_1$ evaluated at $\hat{\boldsymbol\beta}$. The dual norm of $|\cdot|_1$ is $|\cdot|_\infty$, so there exists $\mathbf{v}$ such that $|\mathbf{v}|_\infty \leq 1$ and $(1/n')\mathbf{Z}'^\top(\mathbf{Y} - \mathbf{Z}'\hat{\boldsymbol\beta}) + \lambda\mathbf{v} = 0$. We deduce successively $\mathbf{Z}'^\top\mathbf{Z}'(\boldsymbol\beta^* - \hat{\boldsymbol\beta})/n' + \mathbf{Z}'^\top\boldsymbol\xi/n' + \lambda\mathbf{v} = 0$, $\frac{1}{n'}|\mathbf{Z}'(\boldsymbol\beta^* - \hat{\boldsymbol\beta})|_2^2 + \frac{1}{n'}(\boldsymbol\beta^* - \hat{\boldsymbol\beta})^\top\mathbf{Z}'^\top\boldsymbol\xi + \lambda(\boldsymbol\beta^* - \hat{\boldsymbol\beta})^\top\mathbf{v} = 0$, and finally $\|\mathbf{Z}'(\boldsymbol\beta^* - \hat{\boldsymbol\beta})\|^2_{n'} \leq \frac{1}{n'}\Big\langle\mathbf{Z}'(\hat{\boldsymbol\beta} - \boldsymbol\beta^*)\ ,\ \boldsymbol\xi\Big\rangle + \lambda|\boldsymbol\beta^* - \hat{\boldsymbol\beta}|_1$. □

Lemma 19. *We have* $|\mathbf{u}_{\mathcal{S}^C}|_1 \le |\mathbf{u}_{\mathcal{S}}|_1 + \dfrac{2}{\lambda n'} \left\langle \boldsymbol{\xi}\ ,\ \mathbf{Z}'\mathbf{u} \right\rangle$.

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

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

After some algebra, we derive $\|\mathbf{Y} - \mathbf{Z}'\hat{\boldsymbol{\beta}}\|_{n'}^2 - \|\mathbf{Y} - \mathbf{Z}'\boldsymbol{\beta}^*\|_{n'}^2 \le \lambda\big(|(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}})_{\mathcal{S}}|_1 - |(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{\mathcal{S}^C}|_1\big)$. Moreover, the mapping $\boldsymbol{\beta} \mapsto \|\mathbf{Y} - \mathbf{Z}'\boldsymbol{\beta}\|_{n'}^2$ is strictly convex and its gradient at $\boldsymbol{\beta}^*$ is $-2\,\mathbf{Z}'^\top(\mathbf{Y} - \mathbf{Z}'\boldsymbol{\beta}^*)/n' = -2\,\mathbf{Z}'^\top\boldsymbol{\xi}/n'$. So, we obtain

$$\|\mathbf{Y} - \mathbf{Z}'\hat{\boldsymbol{\beta}}\|_{n'}^2 - \|\mathbf{Y} - \mathbf{Z}'\boldsymbol{\beta}^*\|_{n'}^2 \ge \frac{-2}{n'}\left\langle \mathbf{Z}'^\top\boldsymbol{\xi}\ ,\ \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \right\rangle.$$

Combining the two previous equations, we get

$$(-2)\left\langle \mathbf{Z}'^\top\boldsymbol{\xi}, \hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^* \right\rangle/n' \le \lambda\big(|(\boldsymbol{\beta}^* - \hat{\boldsymbol{\beta}})_{\mathcal{S}}|_1 - |(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)_{\mathcal{S}^C}|_1\big). \quad \square$$

Lemma 20. *Assume that* $\max_{j\in\{1,\dots,p'\}} \left|\sum_{i=1}^{n'} Z'_{i,j}\xi_{i,n}/n'\right| \le t$, *for some* $t > 0$, *that the assumption* $RE(s,3)$ *is satisfied, and that the tuning parameter is given by* $\lambda = \gamma t$, *with* $\gamma \ge 4$. *Then,* $\|\mathbf{Z}'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*)\|_{n'} \le 4(\gamma+1)t\sqrt{s}/\kappa(s,3)$ *and* $|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^*|_q \le 4^{2/q}(\gamma+1)ts^{1/q}/\kappa^2(s,3)$, *for every* $1 \le q \le 2$.

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

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

We first show that $\mathbf{u}$ belongs to the cone $\left\{\boldsymbol{\delta} \in \mathbb{R}^{p'} : |\boldsymbol{\delta}_{\mathcal{S}^C}|_1 \le 3|\boldsymbol{\delta}_{\mathcal{S}}|_1, Card(\mathcal{S}) \le s\right\}$, so that we will be able to use the $RE(s,3)$ assumption with $J_0 = \mathcal{S}$. From Lemma 19, $|\mathbf{u}_{\mathcal{S}^C}|_1 \le |\mathbf{u}_{\mathcal{S}}|_1 + 2t|\mathbf{u}|_1/\lambda$. With our choice of λ, we deduce $|\mathbf{u}_{\mathcal{S}^C}|_1 \le |\mathbf{u}_{\mathcal{S}}|_1 + 2|\mathbf{u}|_1/\gamma$. Using the decomposition $|\mathbf{u}|_1 = |\mathbf{u}_{\mathcal{S}^C}|_1 + |\mathbf{u}_{\mathcal{S}}|_1$, we get $|\mathbf{u}_{\mathcal{S}^C}|_1 \le |\mathbf{u}_{\mathcal{S}}|_1(\gamma+2)/(\gamma-2) \le 3|\mathbf{u}_{\mathcal{S}}|_1$. As a consequence, we have

$$|\mathbf{u}|_1 = |\mathbf{u}_{\mathcal{S}^C}|_1 + |\mathbf{u}_{\mathcal{S}}|_1 \le 4|\mathbf{u}_{\mathcal{S}}|_1 \le 4\sqrt{s}|\mathbf{u}|_2 \le 4\sqrt{s}\,\|\mathbf{Z}'\mathbf{u}\|_{n'}/\kappa(s,3).$$

By Lemma 18,

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

We can now simplify and we get

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

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

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

Proof of Theorem 2. Using Lemma 15, for every $t_1, t_2 > 0$ such that $C_{K,\alpha}h^\alpha/\alpha! + t_1 \le f_{\mathbf{Z},min}/2$, we have

$$\begin{aligned}
&\Pr\Bigg(\max_{i=1,\dots,n'}\left|\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{Z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{z}}\right| > \max_{i=1,\dots,n'}\frac{4}{f_{\mathbf{Z}}^2(\mathbf{Z}'_i)}\left(\frac{C_{\mathbf{XZ},\alpha}h^\alpha}{\alpha!} + \frac{3f_{\mathbf{Z}}(\mathbf{Z}'_i)\int K^2}{2nh^p} + t_2\right) \times \left\{1 + \frac{16f_{\mathbf{Z}}^2(\mathbf{Z}'_i)}{f_{\mathbf{Z},min}^3}\left(\frac{C_{K,\alpha}h^\alpha}{\alpha!} + t_1\right)\right\}\Bigg) \\
&\le 2n'\exp\left(-\frac{nh^pt_1^2}{2f_{\mathbf{Z},max}\int K^2 + (2/3)C_Kt_1}\right) + 2n'\exp\left(-\frac{(n-1)h^{2p}t_2^2}{4f_{\mathbf{Z},max}^2(\int K^2)^2 + (8/3)C_K^2t_2}\right) \\
&\quad + 2n'\exp\left(-\frac{nh^p(f_{\mathbf{Z}}(\mathbf{z}) - C_{\tilde{K},2}h^2)^2}{8f_{\mathbf{Z},max}\int\tilde{K}^2 + 4C_{\tilde{K}}(f_{\mathbf{Z}}(\mathbf{z}) - C_{\tilde{K},2}h^2)/3}\right).
\end{aligned} \tag{B.1}$$

Therefore, on this event, we have

$$\max_{j\in\{1,\dots,p'\}} \left| \frac{1}{n'} \sum_{i=1}^{n'} Z'_{i,j} \xi_{i,n} \right| \le C_\psi \max_{i\in\{1,\dots,n'\}} \left|\xi_{i,n}\right| \le C_\psi C_{\Lambda'} \max_{i\in\{1,\dots,n'\}} \left|\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}\right|$$
$$\le \frac{4C_\psi C_{\Lambda'}}{f^2_{\mathbf{Z},min}} \left(1 + \frac{16 f^2_{\mathbf{Z},max}}{f^3_{\mathbf{Z},min}} \left\{ \frac{C_{K,\alpha} h^\alpha}{\alpha!} + t_1 \right\}\right) \left(\frac{C_{\mathbf{XZ},\alpha} h^\alpha}{\alpha!} + \frac{3 f_{\mathbf{Z},max} \int K^2}{2nh^p} + t_2 \right).$$

Choose $t_1 := f_{\mathbf{Z},min}/4$ so that, because of Condition (5), we get $C_{K,\alpha}h^\alpha/\alpha! + t_1 \le f_{\mathbf{Z},min}/2$. Still on the same event, we have

$$\max_{j\in\{1,\dots,p'\}} \left| \frac{1}{n'} \sum_{i=1}^{n'} Z'_{i,j} \xi_{i,n} \right| \le \frac{4C_\psi C_{\Lambda'}}{f^2_{\mathbf{Z},min}} \left(1 + \frac{8 f^2_{\mathbf{Z},max}}{f^2_{\mathbf{Z},min}}\right) \left(\frac{C_{\mathbf{XZ},\alpha} h^\alpha}{\alpha!} + \frac{3 f_{\mathbf{Z},max} \int K^2}{2nh^p} + t_2 \right). \tag{B.2}$$

Now, choose $t_2 := t f^4_{\mathbf{Z},min} / \{8 C_\psi C_{\Lambda'} (f^2_{\mathbf{Z},min} + 8 f^2_{\mathbf{Z},max})\}$. By Condition (5), we have $C_{\mathbf{XZ},\alpha} h^\alpha / (f^2_{\mathbf{Z},min} \alpha!) \le t_2$ and $3 f_{\mathbf{Z},max} \int K^2 / (2nh^p) \le t_2$, so that the following inequalities hold:

$$\frac{4C_\psi C_{\Lambda'}}{f^2_{\mathbf{Z},min}} \left(1 + \frac{8 f^2_{\mathbf{Z},max}}{f^2_{\mathbf{Z},min}}\right) \left(\frac{C_{\mathbf{XZ},\alpha} h^\alpha}{\alpha!} + t_2 \right) \le \frac{8 t_2 C_\psi C_{\Lambda'}}{f^2_{\mathbf{Z},min}} \left(1 + \frac{8 f^2_{\mathbf{Z},max}}{f^2_{\mathbf{Z},min}}\right) \le t. \tag{B.3}$$

Replacing the values of t_1 and t_2 by their expression in (B.1) and combining it with inequalities (B.2) and (B.3), this yields

$$\Pr\left(\max_{j\in\{1,\dots,p'\}} \left| \frac{1}{n'} \sum_{i=1}^{n'} Z'_{i,j} \xi_{i,n} \right| > t + \frac{C_1}{nh^p} \right) \le 2n' \exp\left(-\frac{nh^p}{C_2} \right) + 2n' \exp\left(-\frac{(n-1)h^{2p} t}{C_3 + C_4 t} \right) + 2n' \exp\left(-\frac{nh^p (C_5 - C_6 h^2)^2}{C_7 - C_8 h^2} \right),$$

where $C_1 := 6 C_\psi C_{\Lambda'} f_{\mathbf{Z},max} \int K^2 \left(f^2_{\mathbf{Z},min} + 8 f^2_{\mathbf{Z},max} \right) f^{-4}_{\mathbf{Z},min}$, $C_2 := 8(12 f_{\mathbf{Z},max} \int K^2 + C_K f_{\mathbf{Z},min})/(3 f^2_{\mathbf{Z},min})$, $C_3 := 32 C_\psi C_{\Lambda'} (f^2_{\mathbf{Z},min} + 8 f^2_{\mathbf{Z},max}) f^{-2}_{\mathbf{Z},max} (\int K^2)^2$, $C_4 := (64/3) C_K^2 f^{-4}_{\mathbf{Z},min} C_\psi C_{\Lambda'} (f^2_{\mathbf{Z},min} + 8 f^2_{\mathbf{Z},max})$, $C_5 := f_{\mathbf{Z},min}$, $C_6 := C_{\tilde K,2}$, $C_7 := 8 f_{\mathbf{Z},max} \int \tilde K^2 + (4/3) C_{\tilde K} f_{\mathbf{Z},min}$, and $C_8 := C_{\tilde K,2}/3$. Finally, apply Lemma 20 to get the claimed result. □

Appendix C. Proofs of asymptotic results for $\hat{\boldsymbol\beta}_{n,n'}$

Proof of Lemma 4. Using the definition (2) of $\hat{\boldsymbol\beta}_{n,n'}$, we get

$$\hat{\boldsymbol\beta}_{n,n'} := \arg\min_{\boldsymbol\beta \in \mathbb{R}^{p'}} \frac{1}{n'} \sum_{i=1}^{n'} \Big(\Lambda(\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}) - \boldsymbol\psi(\mathbf{z}'_i)^\top \boldsymbol\beta \Big)^2 + \lambda_{n,n'} |\boldsymbol\beta|_1$$
$$= \arg\min_{\boldsymbol\beta \in \mathbb{R}^{p'}} \frac{1}{n'} \sum_{i=1}^{n'} \xi^2_{i,n} + \frac{2}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol\psi(\mathbf{z}'_i)^\top (\boldsymbol\beta^* - \boldsymbol\beta) + \frac{1}{n'} \sum_{i=1}^{n'} \left(\boldsymbol\psi(\mathbf{z}'_i)^\top (\boldsymbol\beta^* - \boldsymbol\beta) \right)^2 + \lambda_{n,n'} |\boldsymbol\beta|_1$$
$$= \arg\min_{\boldsymbol\beta \in \mathbb{R}^{p'}} \frac{2}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol\psi(\mathbf{z}'_i)^\top (\boldsymbol\beta^* - \boldsymbol\beta) + \frac{1}{n'} \sum_{i=1}^{n'} \left(\boldsymbol\psi(\mathbf{z}'_i)^\top (\boldsymbol\beta^* - \boldsymbol\beta) \right)^2 + \lambda_{n,n'} |\boldsymbol\beta|_1. \quad \square$$

Proof of Theorem 7. Let us define $r_{n,n'} := (n h^p_{n,n'})^{1/2}$, $\mathbf{u} := r_{n,n'}(\boldsymbol\beta - \boldsymbol\beta^*)$ and $\hat{\mathbf{u}}_{n,n'} := r_{n,n'}(\hat{\boldsymbol\beta}_{n,n'} - \boldsymbol\beta^*)$, so that $\hat{\boldsymbol\beta}_{n,n'} = \boldsymbol\beta^* + \hat{\mathbf{u}}_{n,n'}/r_{n,n'}$. By Lemma 4, $\hat{\boldsymbol\beta}_{n,n'} = \arg\min_{\boldsymbol\beta \in \mathbb{R}^{p'}} \mathbb{G}_{n,n'}(\boldsymbol\beta)$. We have therefore

$$\hat{\mathbf{u}}_{n,n'} = \arg\min_{\mathbf{u} \in \mathbb{R}^{p'}} \left\{ \frac{(-2)}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol\psi(\mathbf{z}'_i)^\top \frac{\mathbf{u}}{r_{n,n'}} + \frac{1}{n'} \sum_{i=1}^{n'} \left(\boldsymbol\psi(\mathbf{z}'_i)^\top \frac{\mathbf{u}}{r_{n,n'}} \right)^2 + \lambda_{n,n'} \left| \boldsymbol\beta^* + \frac{\mathbf{u}}{r_{n,n'}} \right|_1 \right\},$$

or $\hat{\mathbf{u}}_{n,n'} = \arg\min_{\mathbf{u} \in \mathbb{R}^{p'}} \mathbb{F}_{\infty,\infty}(\mathbf{u})$, where, for every $\mathbf{u} \in \mathbb{R}^{p'}$,

$$\mathbb{F}_{\infty,\infty}(\mathbf{u}) := \frac{(-2) r_{n,n'}}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol\psi(\mathbf{z}'_i)^\top \mathbf{u} + \frac{1}{n'} \sum_{i=1}^{n'} \left(\boldsymbol\psi(\mathbf{z}'_i)^\top \mathbf{u} \right)^2 + \lambda_{n,n'} r^2_{n,n'} \left(\left| \boldsymbol\beta^* + \frac{\mathbf{u}}{r_{n,n'}} \right|_1 - \left| \boldsymbol\beta^* \right|_1 \right).$$

Note that, by Corollary 6, we have

$$\frac{2 r_{n,n'}}{n'} \sum_{i=1}^{n'} \xi_{i,n} \boldsymbol\psi(\mathbf{z}'_i)^\top \mathbf{u} = \frac{2}{n'} \sum_{i=1}^{n'} \sum_{j=1}^{p'} r_{n,n'} \xi_{i,n} \psi_j(\mathbf{z}'_i) u_j \xrightarrow{D} \frac{2}{n'} \sum_{i=1}^{n'} \sum_{j=1}^{p'} W_i \psi_j(\mathbf{z}'_i) u_j.$$

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

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

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

We have shown that $\mathbb{F}_{\infty,\infty}(\mathbf{u}) \xrightarrow{D} \mathbb{F}_{\infty,n'}(\mathbf{u})$. Those functions are strictly convex and $\mathbb{F}_{\infty,n'}(\mathbf{u})$ is uniquely minimized at $\mathbf{u}^*$, hence the conclusion follows from the convexity argument. □

Proof of Proposition 8. The proof closely follows Proposition 1 in [16]. It starts by noting that $\Pr(\mathscr{S}_n = \mathscr{S}) \leq \Pr\left(\hat{\boldsymbol{\beta}}_j = 0, \ \forall j \notin \mathscr{S}\right)$. Because of the weak limit of $\hat{\boldsymbol{\beta}}$ (Theorem 7 and the notation therein), this implies

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

If $\ell = 0$, then $\mathbf{u}^*$ is asymptotically normal, and the latter probability is zero. Otherwise, $\ell \neq 0$ and define the Gaussian random vector $\mathbf{W}_{\boldsymbol{\psi}} := 2\sum_{i=1}^{n'} W_i \boldsymbol{\psi}(\mathbf{z}'_i)/n'$. The Karush–Kuhn–Tucker optimality conditions, when applied to $\mathbb{F}_{\infty,n'}$, provide $\mathbf{W}_{\boldsymbol{\psi}} + 2\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top \mathbf{u}^*/n' + \ell \mathbf{v}^* = 0$, for some vector $\mathbf{v}^* \in \mathbb{R}^p$ whose components v_j^* are less than one in absolute value when $j \notin \mathscr{S}$, and $v_j^* = \text{ sign}(\boldsymbol{\beta}_j^*)$ when $j \in \mathscr{S}$. If $u_j^* = 0$ for all $j \notin \mathscr{S}$, we deduce

$$(\mathbf{W}_{\boldsymbol{\psi}})_{\mathscr{S}} + \left[\frac{2}{n'}\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top\right]_{\mathscr{S},\mathscr{S}} \mathbf{u}^*_{\mathscr{S}} + \ell \text{ sign}(\boldsymbol{\beta}^*_{\mathscr{S}}) = 0, \ \left|(\mathbf{W}_{\boldsymbol{\psi}})_{\mathscr{S}^c} + \left[\frac{2}{n'}\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top\right]_{\mathscr{S}^c,\mathscr{S}} \mathbf{u}^*_{\mathscr{S}}\right| \leq \ell,$$

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

$$\left|(\mathbf{W}_{\boldsymbol{\psi}})_{\mathscr{S}^c} - \left[\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top\right]_{\mathscr{S}^c,\mathscr{S}} \left[\sum_{i=1}^{n'} \boldsymbol{\psi}(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top\right]^{-1}_{\mathscr{S},\mathscr{S}} \left((\mathbf{W}_{\boldsymbol{\psi}})_{\mathscr{S}} + \ell \text{ sign}(\boldsymbol{\beta}^*_{\mathscr{S}})\right)\right| \leq \ell,$$

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

Proof of Theorem 9. The beginning of the proof is similar to the proof of Theorem 7. With obvious notation, $\check{\mathbf{u}}_{n,n'} = \arg\min_{\mathbf{u} \in \mathbb{R}^{p'}} \check{\mathbb{F}}_{\infty,\infty}(\mathbf{u})$, where for every $\mathbf{u} \in \mathbb{R}^{p'}$,

$$\check{\mathbb{F}}_{\infty,\infty}(\mathbf{u}) := \frac{(-2)r_{n,n'}}{n'}\sum_{i=1}^{n'} \xi_{i,n}\boldsymbol{\psi}(\mathbf{z}'_i)^\top \mathbf{u} + \frac{1}{n'}\sum_{i=1}^{n'} \left(\boldsymbol{\psi}(\mathbf{z}'_i)^\top \mathbf{u}\right)^2 + \mu_{n,n'} r_{n,n'}^2 \sum_{i=1}^{p'} \frac{1}{|\tilde{\boldsymbol{\beta}}_i|^\delta}\left(|\boldsymbol{\beta}_i^* + \frac{u_i}{r_{n,n'}}| - |\boldsymbol{\beta}_i^*|\right).$$

If $\boldsymbol{\beta}_i^* \neq 0$, then

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

If $\boldsymbol{\beta}_i^* = 0$, then

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

By assumption $\nu_n \tilde{\boldsymbol{\beta}}_i = O_p(1)$, that tends to the infinity in probability iff $u_i \neq 0$. Thus, if there exists some $i \notin \mathscr{S}$ s.t. $u_i \neq 0$, then $\check{\mathbb{F}}_{\infty,\infty}(\mathbf{u})$ tends to the infinity. Otherwise, $u_i = 0$ when $i \notin \mathscr{S}$ and $\check{\mathbb{F}}_{\infty,\infty}(\mathbf{u}) \to \check{\mathbb{F}}_{\infty,n'}(\mathbf{u}_{\mathscr{S}})$. Since $\check{\mathbb{F}}_{\infty,n'}$ is strictly convex, we deduce from [11] that $\check{\mathbf{u}}_{\mathscr{S}} \to \mathbf{u}^*_{\mathscr{S}}$, and $\check{\mathbf{u}}_{\mathscr{S}^c} \to 0_{\mathscr{S}^c}$, proving the asymptotic normality of $\check{\boldsymbol{\beta}}_{n,n',\mathscr{S}}$.

Now, let us prove the oracle property. If $j \in \mathscr{S}$, then $\check{\boldsymbol{\beta}}_j$ tends to $\boldsymbol{\beta}_j$ in probability and $\Pr(j \in \mathscr{S}_n) \to 1$. It suffices to show that $\Pr(j \in \mathscr{S}_n) \to 0$ when $j \notin \mathscr{S}$. If $j \notin \mathscr{S}$ and $j \in \mathscr{S}_n$, the Karush–Kuhn–Tucker conditions on $\check{\mathbb{F}}_{\infty,\infty}$ provide

$$\frac{(-2)r_{n,n'}}{n'}\sum_{i=1}^{n'} \xi_{i,n}\psi_j(\mathbf{z}'_i) + \frac{2}{n'}\sum_{i=1}^{n'} \psi_j(\mathbf{z}'_i)\boldsymbol{\psi}(\mathbf{z}'_i)^\top \check{\mathbf{u}}_{n,n'} = -\frac{\mu_{n,n'} r_{n,n'} \nu_n^\delta}{|\nu_n \tilde{\boldsymbol{\beta}}_j|^\delta} sign(\check{u}_j).$$

Due to the asymptotic normality of $\check{\boldsymbol{\beta}}$ (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 $\nu_n \tilde{\boldsymbol{\beta}}_j = O_P(1)$. Therefore, the probability of the latter event tends to zero when $n \to \infty$. □

Proof of Theorem 10. By Lemma 4, we have $\hat{\boldsymbol\beta}_{n,n'} = \arg\min_{\boldsymbol\beta\in\mathbb{R}^{p'}} \mathbb{G}_{n,n'}(\boldsymbol\beta)$, where

$$\mathbb{G}_{n,n'}(\boldsymbol\beta) := \frac{2}{n'}\sum_{i=1}^{n'} \xi_{i,n}\boldsymbol\psi(\mathbf{z}'_i)^\top(\boldsymbol\beta^* - \boldsymbol\beta) + \frac{1}{n'}\sum_{i=1}^{n'}\big(\boldsymbol\psi(\mathbf{z}'_i)^\top(\boldsymbol\beta^* - \boldsymbol\beta)\big)^2 + \lambda_{n,n'}|\boldsymbol\beta|_1.$$

Define also $\mathbb{G}_{\infty,n'}(\boldsymbol\beta) := \sum_{i=1}^{n'}\big(\boldsymbol\psi(\mathbf{z}'_i)^\top(\boldsymbol\beta^* - \boldsymbol\beta)\big)^2/n' + \lambda_0|\boldsymbol\beta|_1$. We have

$$\big|\mathbb{G}_{n,n'}(\boldsymbol\beta) - \mathbb{G}_{\infty,n'}(\boldsymbol\beta)\big| \le \left|\frac{2}{n'}\sum_{i=1}^{n'} \xi_{i,n}\boldsymbol\psi(\mathbf{z}'_i)^\top(\boldsymbol\beta^* - \boldsymbol\beta)\right| + |\boldsymbol\beta|_1|\lambda_{n,n'} - \lambda_0|.$$

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 $\epsilon > 0$,

$$\Pr\Big(\Big\|\frac{1}{n'}\sum_{i=1}^{n'}\xi_{i,n}\boldsymbol\psi(\mathbf{z}'_i)\Big\| > \epsilon\Big) \le \Pr\Big(\frac{\|C_{\Lambda'}\|}{n'}\sum_{i=1}^{n'}|\hat\tau_{\mathbf{z}'_i} - \tau_{\mathbf{z}'_i}| \times \big\|\boldsymbol\psi(\mathbf{z}'_i)\big\| > \epsilon\Big) \le \sum_{i=1}^{n'}\Pr\Big(|\hat\tau_{\mathbf{z}'_i} - \tau_{\mathbf{z}'_i}| > Cst\,\epsilon\Big),$$

where Cst is the constant $(\|C_{\Lambda'}\| \times \|C_\psi\|)^{-1}$. Apply Lemma 15 with the $t = f_{\mathbf{Z},min}/4$ and t'/ϵ is a sufficiently small constant. When n is sufficiently large, we get

$$\Pr\Big(|\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}} - \tau_{1,2|\mathbf{Z}=\mathbf{z}}| > Cst\,\epsilon\Big) \le 4\exp\Big(-nh^{2p}Cst'\Big),$$

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

Since $\sum_{i=1}^{n'}\psi(\mathbf{z}'_i)\psi(\mathbf{z}'_j)^\top/n'$ tends towards a matrix $\mathbf{M}_{\psi,\mathbf{z}'}$, deduce that $\mathbb{G}_{\infty,n'}(\boldsymbol\beta)$ tends to $\mathbb{G}_{\infty,\infty}(\boldsymbol\beta)$ when $n' \to \infty$. Therefore, for all $\boldsymbol\beta\in\mathbb{R}^{p'}$, $\mathbb{G}_{n,n'}(\boldsymbol\beta)$ weakly tends to $\mathbb{G}_{\infty,\infty}(\boldsymbol\beta)$. By the convexity argument, we deduce that $\arg\min_{\boldsymbol\beta}\mathbb{G}_{n,n'}(\boldsymbol\beta)$ weakly converges to $\arg\min_{\boldsymbol\beta}\mathbb{G}_{\infty,\infty}(\boldsymbol\beta)$. Since the latter minimizer is non random, the same convergence is true in probability. □

Appendix D. Proof of Theorem 11

Proof. We start as in the proof of Theorem 7. Define $\tilde r_{n,n'} := (nn'h^p_{n,n'})^{1/2}$, $\mathbf{u} := \tilde r_{n,n'}(\boldsymbol\beta - \boldsymbol\beta^*)$ and $\hat{\mathbf{u}}_{n,n'} := \tilde r_{n,n'}(\hat{\boldsymbol\beta}_{n,n'} - \boldsymbol\beta^*)$, so that $\hat{\boldsymbol\beta}_{n,n'} = \boldsymbol\beta^* + \hat{\mathbf{u}}_{n,n'}/\tilde r_{n,n'}$. We define for every $\mathbf{u}\in\mathbb{R}^{p'}$,

$$\mathbb{F}_{\infty,\infty}(\mathbf{u}) := \frac{(-2)\tilde r_{n,n'}}{n'}\sum_{i=1}^{n'}\xi_{i,n}\boldsymbol\psi(\mathbf{z}'_i)^\top\mathbf{u} + \frac{1}{n'}\sum_{i=1}^{n'}\Big\{\boldsymbol\psi(\mathbf{z}'_i)^\top\mathbf{u}\Big\}^2 + \lambda_{n,n'}\tilde r^2_{n,n'}\Big(\big|\boldsymbol\beta^* + \frac{\mathbf{u}}{\tilde r_{n,n'}}\big|_1 - \big|\boldsymbol\beta^*\big|_1\Big), \tag{D.1}$$

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

Lemma 21. *Under the same assumptions as in Theorem 11, $\mathbf{T}_1 := (\tilde r_{n,n'}/n')\sum_{i=1}^{n'}\xi_{i,n}\boldsymbol\psi(\mathbf{z}'_i)$ tends in law towards a Gaussian random vector $\mathcal{N}(\mathbf{0},\mathbf{V}_2)$.*

This lemma is proved below. It will help to control the first term of (D.1), which is simply $-2\mathbf{T}_1^\top\mathbf{u}$. Concerning the second term of (D.1), using Assumption 7(iii), $\sum_{i=1}^{n'}\Big\{\boldsymbol\psi(\mathbf{z}'_i)^\top\mathbf{u}\Big\}^2/n' \to \int\Big\{\boldsymbol\psi(\mathbf{z}')^\top\mathbf{u}\Big\}^2 f_{\mathbf{z}',\infty}\,d\mathbf{z}'$, for every $\mathbf{u}\in\mathbb{R}^{p'}$. This has to be read as a convergence of a sequence of real numbers indexed by $\mathbf{u}$, because the design points $\mathbf{z}'_i$ are deterministic. We also have, for any $\mathbf{u}\in\mathbb{R}^{p'}$ and when n is large enough,

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

Therefore, by Assumption 7(ii)(b), for every $\mathbf{u}\in\mathbb{R}^{p'}$,

$$\lambda_{n,n'}\tilde r^2_{n,n'}\Big(\big|\boldsymbol\beta^* + \frac{\mathbf{u}}{\tilde r_{n,n'}}\big|_1 - \big|\boldsymbol\beta^*\big|_1\Big) \to 0, \tag{D.2}$$

when (n, n') tends to the infinity. Combining Lemma 21, (D.1), (D.2) and defining the function $\mathbb{F}_{\infty,\infty}$ by

$$\mathbb{F}_{\infty,\infty}(\mathbf{u}) := 2\tilde{\mathbf{W}}^\top\mathbf{u} + \int\Big\{\boldsymbol\psi(\mathbf{z}')^\top\mathbf{u}\Big\}^2 f_{\mathbf{z}',\infty}(\mathbf{z}')d\mathbf{z}',\ \mathbf{u}\in\mathbb{R}^{p'},$$

where $\tilde W \sim \mathcal{N}(\mathbf{0},\mathbf{V}_2)$, we obtain that every finite-dimensional margin of $\mathbb{F}_{\infty,\infty}$ converges weakly to the corresponding margin of $\mathbb{F}_{\infty,\infty}$. Now, applying the convexity lemma, we get $\hat{\mathbf{u}}_{n,n'} \xrightarrow{D} \mathbf{u}_{\infty,\infty}$, 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 $\nabla\mathbb{F}_{\infty,\infty}(\mathbf{u}) = \mathbf{0}$, which yields $2\tilde{\mathbf{W}} + 2\int \boldsymbol{\psi}(\mathbf{z}')\boldsymbol{\psi}(\mathbf{z}')^\top \mathbf{u}_{\infty,\infty} f_{\mathbf{z}',\infty}(\mathbf{z}')d\mathbf{z}' = \mathbf{0}$. As a consequence $\mathbf{u}_{\infty,\infty} = -\mathbf{V}_1^{-1}\tilde{\mathbf{W}} \sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{V}}_{as})$, using Assumption 7(iv). We finally obtain $\tilde{r}_{n,n'}\big(\hat{\boldsymbol{\beta}}_{n,n'} - \boldsymbol{\beta}^*\big) \xrightarrow{D} \mathcal{N}\big(\mathbf{0}, \tilde{\mathbf{V}}_{as}\big)$, as claimed. □

Proof of Lemma 21 (*Convergence of* $\mathbf{T}_1$)**.** Using a Taylor expansion, we have

$$\mathbf{T}_1 := \frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'} \xi_{i,n}\boldsymbol{\psi}(\mathbf{z}'_i) = \frac{\tilde{r}_{n,n'}}{n'}\sum_{i=1}^{n'}\Big\{\Lambda\big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i}\big) - \Lambda\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}\big)\Big\}\boldsymbol{\psi}(\mathbf{z}'_i) = \mathbf{T}_2 + \mathbf{T}_3,$$

where the main term is $\mathbf{T}_2 := \tilde{r}_{n,n'}\sum_{i=1}^{n'} \Lambda'\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}\big)\big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}\big)\boldsymbol{\psi}(\mathbf{z}'_i)/n'$, and the remainder is

$$\mathbf{T}_3 := \tilde{r}_{n,n'}\sum_{i=1}^{n'}\alpha_{3,i}\big(\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}'_i} - \tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}\big)^2\boldsymbol{\psi}(\mathbf{z}'_i)/n',$$

with $|\alpha_{3,i}| \le C_{\Lambda''}/2$, $i \in \{1,\dots,n'\}$, by Assumption 7(v). Using the definition (3) of $\hat{\tau}_{1,2|\mathbf{Z}=\mathbf{z}}$, the definition of the weights $w_{i,n}(\mathbf{z})$ and the notation $\overline{\boldsymbol{\psi}}(\mathbf{z}) := \Lambda'\big(\tau_{1,2|\mathbf{Z}=\mathbf{z}}\big)\boldsymbol{\psi}(\mathbf{z})$, we rewrite $\mathbf{T}_2 =: \mathbf{T}_4 + \mathbf{T}_5$, where

$$\mathbf{T}_4 := \frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\frac{K_h(\mathbf{z}'_i - \mathbf{Z}_{j_1})K_h(\mathbf{z}'_i - \mathbf{Z}_{j_2})}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)}\Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2}) - \mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}'_i\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{z}'_i), \tag{D.3}$$

$$\mathbf{T}_5 := \frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}K_h(\mathbf{z}'_i - \mathbf{Z}_{j_1})K_h(\mathbf{z}'_i - \mathbf{Z}_{j_2})\bigg(\frac{1}{\hat{f}_{\mathbf{Z}}(\mathbf{z}'_i)^2} - \frac{1}{f_{\mathbf{Z}}(\mathbf{z}'_i)^2}\bigg)\Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2}) - \mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}'_i\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{z}'_i). \tag{D.4}$$

Note that we can put together the terms (j_1, j_2) and (j_2, j_1). This corresponds to the substitution of g^* by its symmetrized version $\tilde{g}$. In the following, we will therefore assume that g^* has been symmetrized without loss of generality. The random variable $\mathbf{T}_4$ can be seen, see (D.3), as a sum of (indexed by i) U-statistics of order 2. Its Hájek projection (see [14, p. 162]) will yield the asymptotically normal dominant term of $\mathbf{T}_2$. To lighten notation, we denote $\tau_i := \tau_{1,2|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}'_i}$, $f(\cdot,\cdot) = f_{\mathbf{X},\mathbf{Z}}(\cdot,\cdot)$ and

$$g_{i,j_1,j_2} := g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2}) - \mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{z}'_i\big] = g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2}) - \tau_i.$$

Implicitly, all the expectations we will consider are expectations conditionally on the sequence of $\mathbf{z}'_i$, $i \ge 1$.

First, note that, by usual α-order limited expansions, we have

$$\begin{aligned}
\mathrm{E}[\mathbf{T}_4] &= \frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'} n(n-1)\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)}\big(g^*(\mathbf{x}_1,\mathbf{x}_2) - \tau_i\big)\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)f(\mathbf{x}_1,\mathbf{z}_1)f(\mathbf{x}_2,\mathbf{z}_2)\,d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{z}_1\,d\mathbf{z}_2 \\
&\quad - \frac{\tilde{r}_{n,n'}}{n'n}\sum_{i=1}^{n'}\tau_i\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)\int\frac{K_h^2(\mathbf{z}'_i - \mathbf{z})}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)}f(\mathbf{x},\mathbf{z})d\mathbf{x}\,d\mathbf{z} \\
&= \frac{(n-1)\tilde{r}_{n,n'}}{n'n}\sum_{i=1}^{n'}\int\frac{K(\mathbf{t}_1)K(\mathbf{t}_2)}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)}\big(g^*(\mathbf{x}_1,\mathbf{x}_2) - \tau_i\big)\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)f(\mathbf{x}_1,\mathbf{z}'_i - h\mathbf{t}_1)f(\mathbf{x}_2,\mathbf{z}'_i - h\mathbf{t}_1)\,d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{t}_1\,d\mathbf{t}_2 \\
&\quad - \frac{\tilde{r}_{n,n'}}{n'nh^p}\sum_{i=1}^{n'}\tau_i\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)\int\frac{K^2(\mathbf{t})}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)}f_{\mathbf{X},\mathbf{Z}}(\mathbf{x},\mathbf{z}'_i - h\mathbf{t})d\mathbf{x}\,d\mathbf{t} \\
&= \frac{(n-1)\tilde{r}_{n,n'}h^{2\alpha}}{n'n}\sum_{i=1}^{n'}\int\frac{K(\mathbf{t}_1)K(\mathbf{t}_2)}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)}\big(g^*(\mathbf{x}_1,\mathbf{x}_2) - \tau_i\big)\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)d^{(\alpha)}_{\mathbf{Z}}f(\mathbf{x}_1,\mathbf{z}^*_i)\cdot\mathbf{t}_1^{(\alpha)}d^{(\alpha)}_{\mathbf{Z}}f(\mathbf{x}_2,\mathbf{z}^*_i)\cdot\mathbf{t}_2^{(\alpha)}\,d\mathbf{x}_1\,d\mathbf{x}_2\,d\mathbf{t}_1\,d\mathbf{t}_2 \\
&\quad - \frac{\tilde{r}_{n,n'}}{n'nh^p}\sum_{i=1}^{n'}\tau_i\int K^2\int\frac{\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)}f(\mathbf{x},\mathbf{z}^*_i)\,d\mathbf{x} = O\Big(\tilde{r}_{n,n'}h^{2\alpha} + \frac{\tilde{r}_{n,n'}}{nh^p}\Big) = O\Big(\sqrt{nn'h^{p+4\alpha}} + \sqrt{n'/(nh^p)}\Big) = o(1),
\end{aligned}$$

under Assumption 7(ii). Above, we have denoted by $\mathbf{z}^*_i$ some vectors in $\mathbb{R}^p$ s.t. $\|\mathbf{z}'_i - \mathbf{Z}^*_i\|_\infty < 1$. They depend on $\mathbf{z}'_i$, $\mathbf{x}_1$, $\mathbf{x}_2$ or $\mathbf{x}$, respectively.

Moreover, set

$$\mathbf{T}_4 - \mathrm{E}[\mathbf{T}_4] = \frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1,j_2=1}^{n}\zeta_{i,j_1,j_2}, \tag{D.5}$$

$$\zeta_{i,j_1,j_2} = \Big\{ K_h(\mathbf{z}'_i - \mathbf{Z}_{j_1}) K_h(\mathbf{z}'_i - \mathbf{Z}_{j_2}) g_{i,j_1,j_2} - \mathrm{E}\Big[K_h(\mathbf{z}'_i - \mathbf{Z}_{j_1}) K_h(\mathbf{z}'_i - \mathbf{Z}_{j_2}) g_{i,j_1,j_2} \Big] \Big\} \frac{\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)}.$$

Note that $\mathrm{Var}(\mathbf{T}_4) = \mathrm{E}[\mathbf{T}_4 \mathbf{T}_4^\top] + o(1)$ and

$$\mathrm{E}[\mathbf{T}_4 \mathbf{T}_4^\top] = \frac{\tilde{r}^2_{n,n'}}{(n')^2 n^4} \sum_{i_1,i_2=1}^{n'} \sum_{j_1,j_2=1}^{n} \sum_{j_3,j_4=1}^{n} \mathrm{E}[\zeta_{i_1,j_1,j_2} \zeta^\top_{i_2,j_3,j_4}].$$

By independence, $\mathrm{E}[\zeta_{i,j_1,j_2} \zeta^\top_{i,j_3,j_4}] = 0$ when $\{j_1, j_2\} \cap \{j_3, j_4\} = \emptyset$. Otherwise, assume that $j_1 = j_3 = j$ and there are no other identities among the four indices (j_1, j_2, j_3, j_4). Set

$$\overline{\zeta}_i := \mathrm{E}\Big[K_h(\mathbf{z}'_i - \mathbf{Z}_1) K_h(\mathbf{z}'_i - \mathbf{Z}_2) g_{i,1,2} \Big] \frac{\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)}. \tag{D.6}$$

Then, $\mathrm{E}[\zeta_{i_1,j,j_2} \zeta^\top_{i_2,j,j_4}] = \zeta_{i_1,j,j_2,i_2,j,j_4} - \overline{\zeta}_{i_1} \overline{\zeta}^\top_{i_2}$, where

$$\begin{aligned} \zeta_{i_1,j,j_2,i_2,j,j_4} &:= \mathrm{E}\Big[K_h(\mathbf{z}'_{i_1} - \mathbf{Z}_j) K_h(\mathbf{z}'_{i_1} - \mathbf{Z}_{j_2}) K_h(\mathbf{z}'_{i_2} - \mathbf{Z}_j) K_h(\mathbf{z}'_{i_2} - \mathbf{Z}_{j_4}) g_{i_1,j,j_2} g^\top_{i_2,j,j_4} \Big] \frac{\overline{\boldsymbol{\psi}}(\mathbf{z}'_{i_1}) \overline{\boldsymbol{\psi}}(\mathbf{z}'_{i_2})^\top}{f^2_{\mathbf{Z}}(\mathbf{z}'_{i_1}) f^2_{\mathbf{Z}}(\mathbf{z}'_{i_2})} \\ &= \frac{\overline{\boldsymbol{\psi}}(\mathbf{z}'_{i_1}) \overline{\boldsymbol{\psi}}(\mathbf{z}'_{i_2})^\top}{h^p f^2_{\mathbf{Z}}(\mathbf{z}'_{i_1}) f^2_{\mathbf{Z}}(\mathbf{z}'_{i_2})} \int K(\mathbf{t}_1) K(\mathbf{t}_2) K\Big(\frac{\mathbf{z}'_{i_2} - \mathbf{z}'_{i_1}}{h} + \mathbf{t}_1 \Big) K(\mathbf{t}_4) \big(g^*(\mathbf{x}_1, \mathbf{x}_2) - \tau_{i_1} \big) \\ &\quad \times \big(g^*(\mathbf{x}_1, \mathbf{x}_4) - \tau_{i_2} \big) f(\mathbf{x}_1, \mathbf{z}'_{i_1} - h\mathbf{t}_1) f(\mathbf{x}_2, \mathbf{z}'_{i_1} - h\mathbf{t}_2) f(\mathbf{x}_4, \mathbf{z}'_{i_4} - h\mathbf{t}_4)\, d\mathbf{x}_1\, d\mathbf{x}_2\, d\mathbf{x}_4\, d\mathbf{t}_1\, d\mathbf{t}_2\, d\mathbf{t}_4. \end{aligned}$$

By assumption, $\zeta_{i_1,j,j_2,i_2,j,j_4}$ is zero when $i_1 \neq i_2$. Otherwise, when $i_1 = i_2 = i$,

$$\begin{aligned} \zeta_{i,j,j_2,i,j,j_4} &\simeq \frac{\overline{\boldsymbol{\psi}}(\mathbf{z}'_i) \overline{\boldsymbol{\psi}}(\mathbf{z}'_i)^\top}{h^p f_{\mathbf{Z}}(\mathbf{z}'_i)} \int K^2 \int \big(g^*(\mathbf{x}_1, \mathbf{x}_2) - \tau_i \big) \big(g^*(\mathbf{x}_1, \mathbf{x}_4) - \tau_i \big) \\ &\quad \times f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{z}'_i) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{z}'_i) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_4|\mathbf{z}'_i)\, d\mathbf{x}_1\, d\mathbf{x}_2\, d\mathbf{x}_4 := C_{i,1,2,4}/h^p. \end{aligned}$$

It is easy to check that the terms with other identities among the four indices j_k, as ζ_{i,j,j_2,i,j,j_2} or $\zeta_{i,j,j_2,i,j,j}$ will induce negligible remainder terms. Therefore, we get

$$\frac{\tilde{r}^2_{n,n'}}{(n')^2 n^4} \sum_{i_1,i_2=1}^{n'} \sum_{j,j_2,j_4=1}^{n} \zeta_{i_1,j,j_2,i_2,j,j_4} \simeq \frac{1}{n'} \sum_{i=1}^{n'} C_{i,1,2,4}.$$

Concerning the terms induced by the product of two $\overline{\zeta}_i$, note that, by limited expansions,

$$\begin{aligned} \overline{\zeta}_i &= \frac{\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)} \int K_h(\mathbf{z}'_i - \mathbf{z}_1) K_h(\mathbf{z}'_i - \mathbf{z}_2) \big(g^*(\mathbf{x}_1, \mathbf{x}_2) - \tau_i \big) f(\mathbf{x}_1, \mathbf{z}_1) f(\mathbf{x}_2, \mathbf{z}_2)\, d\mathbf{x}_1\, d\mathbf{z}_1\, d\mathbf{x}_2\, d\mathbf{z}_2 \\ &= \frac{h^{2\alpha} \overline{\boldsymbol{\psi}}(\mathbf{z}'_i)}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)} \int K(\mathbf{t}_1) K(\mathbf{t}_2) \big(g^*(\mathbf{x}_1, \mathbf{x}_2) - \tau_i \big) d^{(\alpha)}_{\mathbf{Z}} f(\mathbf{x}_1, \mathbf{z}^*_i) \cdot \mathbf{t}_1^{(\alpha)} d^{(\alpha)}_{\mathbf{Z}} f(\mathbf{x}_2, \mathbf{z}^*_i) \cdot \mathbf{t}_2^{(\alpha)}\, d\mathbf{x}_1\, d\mathbf{t}_1\, d\mathbf{x}_2\, d\mathbf{t}_2, \end{aligned}$$

with the same notation as above. As a consequence, $\sup_i \overline{\zeta}_i = O(h^{2\alpha})$ and

$$\frac{\tilde{r}^2_{n,n'}}{(n')^2 n^4} \sum_{i_1,i_2=1}^{n'} \sum_{j,j_2,j_4=1}^{n} \overline{\zeta}_{i_1} \overline{\zeta}_{i_2} \simeq \frac{\tilde{r}^2_{n,n'}}{n} \Big(\frac{1}{n'} \sum_{i=1}^{n'} \overline{\zeta}_{i,1,2} \Big)^2 = O\Big(\frac{h^{4\alpha} \tilde{r}^2_{n,n'}}{n} \Big) = O(n' h^{4\alpha+p}) = o(1).$$

Therefore, we obtain

$$\frac{\tilde{r}^2_{n,n'}}{(n')^2 n^4} \sum_{i_1,i_2=1}^{n'} \sum_{j,j_2,j_4=1}^{n} \mathrm{E}[\zeta_{i_1,j,j_2} \zeta^\top_{i_2,j,j_4}] \simeq \frac{1}{n'} \sum_{i=1}^{n'} C_{i,1,2,4}.$$

To calculate $\mathrm{E}[\mathbf{T}_4 \mathbf{T}_4^\top]$, there are three other similar terms, that respectively correspond to the cases $j_1 = j_4$, $j_2 = j_3$ or $j_2 = j_4$. Therefore, we deduce

$$\begin{aligned} \mathrm{Var}(\mathbf{T}_4) \simeq \mathrm{E}[\mathbf{T}_4 \mathbf{T}_4^\top] \simeq \frac{4}{n'} \sum_{i=1}^{n'} C_{i,1,2,4} \simeq 4 \int K^2 \int \frac{\overline{\boldsymbol{\psi}}(\mathbf{z}) \overline{\boldsymbol{\psi}}(\mathbf{z})^\top}{f_{\mathbf{Z}}(\mathbf{z})} \int \big(g^*(\mathbf{x}_1, \mathbf{x}_2) - \tau_{1,2|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}} \big) \big(g^*(\mathbf{x}_1, \mathbf{x}_4) - \tau_{1,2|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}} \big) \\ \times f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_1|\mathbf{z}) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_2|\mathbf{z}) f_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}_4|\mathbf{z}) f_{\mathbf{z}',\infty}(\mathbf{z})\, d\mathbf{x}_1\, d\mathbf{x}_2\, d\mathbf{x}_4\, d\mathbf{z}, \end{aligned}$$

that is equal to the so-called variance–covariance matrix $\mathbf{V}_2$. Now assume that $\mathbf{T}_4 - \mathrm{E}[\mathbf{T}_4]$ is asymptotically normal, i.e., $\mathbf{T}_4 - \mathrm{E}[\mathbf{T}_4] \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{V}_2)$. This result will be proved below.

Decompose the term $\mathbf{T}_5$, given in (D.4). For every $i \in \{1, \ldots, n'\}$, a usual Taylor expansion yields

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

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

Therefore, we obtain the decomposition $\mathbf{T}_5 = (-2)\mathbf{T}_6 + \mathbf{T}_7$, where

$$\mathbf{T}_6 := \frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} K_h(\mathbf{z}'_i - \mathbf{Z}_{j_1})K_h(\mathbf{z}'_i - \mathbf{Z}_{j_2})\left(\frac{\hat{f}_{\mathbf{Z}}(\mathbf{z}'_i) - f_{\mathbf{Z}}(\mathbf{z}'_i)}{f^3_{\mathbf{Z}}(\mathbf{z}'_i)}\right)\Big(g^*(\mathbf{X}_{j_1}, \mathbf{X}_{j_2}) - \mathrm{E}\big[g^*(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}'_i\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{z}'_i),$$

$$\mathbf{T}_7 := \frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} K_h(\mathbf{z}'_i - \mathbf{Z}_{j_1})K_h(\mathbf{z}'_i - \mathbf{Z}_{j_2}T_{7,i})\Big(g^*(\mathbf{X}_{j_1}, \mathbf{X}_{j_2}) - \mathrm{E}\big[g^*(\mathbf{X}_1, \mathbf{X}_2)|\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{z}'_i\big]\Big)\overline{\boldsymbol{\psi}}(\mathbf{z}'_i).$$

Summing up all the previous equations, we get $\mathbf{T}_1 = \big(\mathbf{T}_4 - \mathrm{E}[\mathbf{T}_4]\big) - 2\,\mathbf{T}_6 + \mathbf{T}_7 + \mathbf{T}_3 + o(1)$. Afterwards, we will prove that all the remainders terms $\mathbf{T}_6$, $\mathbf{T}_7$ and $\mathbf{T}_3$ are negligible, i.e., they tend to zero in probability. These results are proved below. Combining all these elements with the asymptotic normality of $\mathbf{T}_4$, we get $\mathbf{T}_1 \xrightarrow{D} \mathcal{N}(\mathbf{0}, \mathbf{V}_2)$, as claimed. □

Proof of the asymptotic normality of $\mathbf{T}_4$. We will lead the usual Hájek projection of $\mathbf{T}_4$. To weaken notation, denote $\mathrm{E}[\zeta_{i,j_1,j_2}|\mathbf{X}_{j_1}, \mathbf{Z}_{j_1}] := \mathrm{E}[\zeta_{i,j_1,j_2}|j_1]$. Then, recalling (D.5), we can write $\mathbf{T}_4 - \mathrm{E}[\mathbf{T}_4] = \mathbf{T}_{4,1} + \mathbf{T}_{4,2} + \mathbf{T}_{4,3}$, with

$$\mathbf{T}_{4,1} := \frac{2\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1,j_2=1}^{n} \mathbb{1}(j_1 \neq j_2)\mathrm{E}[\zeta_{i,j_1,j_2}|j_1], \quad \mathbf{T}_{4,2} := \frac{2\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j=1}^{n}\mathrm{E}[\zeta_{i,j,j}|j],$$

$$\mathbf{T}_{4,3} := \frac{\tilde{r}_{n,n'}}{n'n^2}\sum_{i=1}^{n'}\sum_{j_1,j_2=1}^{n}\Big(\zeta_{i,j_1,j_2} - \mathrm{E}[\zeta_{i,j_1,j_2}|j_1] - \mathrm{E}[\zeta_{i,j_1,j_2}|j_2]\Big).$$

We will prove that $\mathbf{T}_{4,2}$ and $\mathbf{T}_{4,3}$ are $o_P(1)$. Therefore, the asymptotic normality of $\mathbf{T}_4$ reduces to the one of $\mathbf{T}_{4,1}$.

Note that $n\mathbf{T}_{4,1}/2(n-1) = \sum_{j=1}^n \boldsymbol{\beta}_{j,n,n'}$, where $\boldsymbol{\beta}_{j,n,n'} := \tilde{r}_{n,n'}\sum_{i=1}^{n'}\mathrm{E}[\zeta_{i,j,0}|j]/(n'n)$, $j \in \{1, \ldots, n\}$, by formally considering a random vector $\mathbf{Z}_0$ that is independent of the other $\mathbf{Z}_j$, $j \ge 1$. Therefore, we get a triangular array of random vectors $(\boldsymbol{\beta}_{j,n,n'})_{j\in\{1,\ldots,n\}}$, s.t., for a fixed n, the variables $\boldsymbol{\beta}_{j,n,n'}$ are mutually independent given the vectors $\mathbf{z}'_i$, $i \ge 1$. Let us check Lyapunov's sufficient condition, that will imply the asymptotic normality of $\mathbf{T}_{4,1}$. In other words, it is sufficient to prove that $\sum_{j=1}^n \|\boldsymbol{\beta}_{j,n,n'}\|^3_\infty \longrightarrow 0$, when n and n' tend to the infinity. Recalling (D.6), we can rewrite

$$\boldsymbol{\beta}_{j,n,n'} = \frac{\tilde{r}_{n,n'}}{n'n}\sum_{i=1}^{n'}\left\{K_h(\mathbf{z}'_i - \mathbf{Z}_j)\frac{\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)}{f^2_{\mathbf{Z}}(\mathbf{z}'_i)}\int K_h(\mathbf{z}'_i - \mathbf{z})\big(g^*(\mathbf{x}, \mathbf{X}_j) - \tau_i\big)f(\mathbf{x}, \mathbf{z})\,d\mathbf{x}\,d\mathbf{z} - \overline{\zeta}_i\right\} := \frac{\tilde{r}_{n,n'}}{n'n}\sum_{i=1}^{n'}\gamma_{i,j},$$

where $\sup_i \overline{\zeta}_i := O(h^{2\alpha})$. Note that

$$\|\boldsymbol{\beta}_{j,n,n'}\|^3_\infty \le p^3\frac{\tilde{r}^3_{n,n'}}{(n')^3n^3}\sum_{i_1,i_2,i_3=1}^{n'}\|\gamma_{i_1,j}\|_\infty\|\gamma_{i_2,j}\|_\infty\|\gamma_{i_3,j}\|_\infty.$$

The terms that involve some products by the means $\overline{\zeta}_{i_k}$, $k \in \{1, 2, 3\}$, are negligible and they may be forgotten here. For some constants Cst, this provides

$$\begin{aligned}\sum_{j=1}^n \mathrm{E}\Big[\|\boldsymbol{\beta}_{j,n,n'}\|^3_\infty\Big] \le\ & \frac{Cst\,\tilde{r}^3_{n,n'}}{(n')^3n^3}\sum_{j=1}^{n}\sum_{i_1,i_2,i_3=1}^{n'}\frac{\|\overline{\boldsymbol{\psi}}\|_\infty(\mathbf{z}'_{i_1})\|\overline{\boldsymbol{\psi}}\|_\infty(\mathbf{z}'_{i_2})\|\overline{\boldsymbol{\psi}}\|_\infty(\mathbf{z}'_{i_3})}{f^2_{\mathbf{Z}}(\mathbf{z}'_{i_1})f^2_{\mathbf{Z}}(\mathbf{z}'_{i_2})f^2_{\mathbf{Z}}(\mathbf{z}'_{i_3})}\\ &\times \mathrm{E}\Big[\Big|K_h(\mathbf{z}'_{i_1} - \mathbf{Z}_j)\int K_h(\mathbf{z}'_{i_1} - \mathbf{z}_1)\big(g^*(\mathbf{x}_1, \mathbf{X}_j) - \tau_{i_1}\big)f(\mathbf{x}_1, \mathbf{z}_1)\,d\mathbf{x}_1\,d\mathbf{z}_1\Big|\\ &\times \Big|K_h(\mathbf{z}'_{i_2} - \mathbf{Z}_j)\int K_h(\mathbf{z}'_{i_2} - \mathbf{z}_2)\big(g^*(\mathbf{x}_2, \mathbf{X}_j) - \tau_{i_2}\big)f(\mathbf{x}_2, \mathbf{z}_2)\,d\mathbf{x}_2\,d\mathbf{z}_2\Big|\\ &\times \Big|K_h(\mathbf{z}'_{i_3} - \mathbf{Z}_j)\int K_h(\mathbf{z}'_{i_3} - \mathbf{z}_3)\big(g^*(\mathbf{x}_3, \mathbf{X}_j) - \tau_{i_3}\big)f(\mathbf{x}_3, \mathbf{z}_3)\,d\mathbf{x}_3\,d\mathbf{z}_3\Big|\Big].\end{aligned}$$

By some now usual changes of variables, the latter expectations are zero when one of the three indices i_1, i_2 and i_3 is different from the others. Thus, the non-zero expectations are obtained when $i_1 = i_2 = i_3$. In the latter case, we get

$$\sum_{j=1}^{n} \|\boldsymbol{\beta}_{j,n,n'}\|_\infty^3 \le \frac{Cst\, \tilde{r}^3_{n,n'}}{(n')^3 n^2} \sum_{i=1}^{n'} \frac{\|\overline{\boldsymbol{\psi}}\|^3_\infty(\mathbf{z}'_i)}{f^6_{\mathbf{Z}}(\mathbf{z}'_i)} \int |K|^3_h(\mathbf{z}'_i - \mathbf{z}) \Big| \int K_h(\mathbf{z}'_i - \mathbf{z}_1)\big(g^*(\mathbf{x}_1, \mathbf{x}) - \tau_i\big) f(\mathbf{x}_1, \mathbf{z}_1)\, d\mathbf{x}_1\, d\mathbf{z}_1 \Big|^3 f(\mathbf{x}, \mathbf{z})\, d\mathbf{x}\, d\mathbf{z}$$
$$\le \frac{Cst\, \tilde{r}^3_{n,n'}}{(n')^3 n^2 h^{2p}} \sum_{i=1}^{n'} \frac{\|\overline{\boldsymbol{\psi}}\|^3_\infty(\mathbf{z}'_i)}{f^6_{\mathbf{Z}}(\mathbf{z}'_i)} \int |K|^3(\mathbf{t}) \Big| \int K(\mathbf{t}_1)\big(g^*(\mathbf{x}_1, \mathbf{x}) - \tau_i\big) f(\mathbf{x}_1, \mathbf{z}'_i - h\mathbf{t}_1)\, d\mathbf{x}_1\, d\mathbf{t}_1 \Big|^3$$
$$\times f(\mathbf{x}, \mathbf{z}'_i - h\mathbf{t})\, d\mathbf{x}\, d\mathbf{t} = O\Big(\frac{\tilde{r}^3_{n,n'}}{(n')^2 n^2 h^{2p}} \Big) = O\Big(\frac{1}{(nn'h^p)^{1/2}} \Big) = o(1).$$

Concerning the remainder terms $\mathbf{T}_{4,2}$ and $\mathbf{T}_{4,3}$, note that $\mathrm{E}[\mathbf{T}_{4,2}] = \mathrm{E}[\mathbf{T}_{4,3}] = 0$. Moreover, since $\mathrm{E}[\zeta_{i,j,j}|j]$ is centered,

$$\mathrm{E}[\mathbf{T}_{4,2}\mathbf{T}_{4,2}^\top] = \frac{4\tilde{r}^2_{n,n'}}{(n')^2 n^4} \sum_{i_1,i_2=1}^{n'} \sum_{j=1}^{n} \mathrm{E}\Big[\mathrm{E}[\zeta_{i_1,j,j}|j] \mathrm{E}[\zeta^\top_{i_2,j,j}|j] \Big].$$

When $i_1 \ne i_2$, some usual changes of variables yield

$$\mathrm{E}\Big[\mathrm{E}[\zeta_{i_1,j,j}|j] \mathrm{E}[\zeta^\top_{i_2,j,j}|j] \Big] = \frac{\overline{\boldsymbol{\psi}}(\mathbf{z}'_{i_1}) \overline{\boldsymbol{\psi}}(\mathbf{z}'_{i_2})^\top}{f^2_{\mathbf{Z}}(\mathbf{z}'_{i_1}) f^2_{\mathbf{Z}}(\mathbf{z}'_{i_2})} \tau_{i_1}\tau_{i_2} \Big(\mathrm{E}\big[K^2_h(\mathbf{z}'_{i_1} - \mathbf{Z}_j) K^2_h(\mathbf{z}'_{i_2} - \mathbf{Z}_j)\big] - \mathrm{E}\big[K^2_h(\mathbf{z}'_{i_1} - \mathbf{Z}_j)\big] \mathrm{E}\big[K^2_h(\mathbf{z}'_{i_2} - \mathbf{Z}_j)\big] \Big)$$
$$= O(h^{-2p}),$$

uniformly w.r.t. i. By a similar reasoning, we can prove that $\sup_i \mathrm{E}\Big[\mathrm{E}[\zeta_{i,j,j}|j] \mathrm{E}[\zeta^\top_{i,j,j}|j] \Big] = O(h^{-3p})$. Therefore,

$$\mathrm{E}[\mathbf{T}_{4,2}\mathbf{T}_{4,2}^\top] = O\Big(\frac{\tilde{r}^2_{n,n'}}{(n')^2 n^4} \big\{ (n')^2 n h^{-2p} + n' n h^{-3p} \big\} \Big) = O\Big(\frac{n'}{n^2 h^p} + \frac{1}{n^2 h^{2p}} \Big) = o(1).$$

Concerning $\mathbf{T}_{4,3}$, this remainder term is centered and

$$\mathrm{E}[\mathbf{T}_{4,3}\mathbf{T}_{4,3}^\top] = \frac{\tilde{r}^2_{n,n'}}{(n')^2 n^4} \sum_{i_1,i_2=1}^{n'} \sum_{j_1,j_2=1}^{n} \sum_{j_3,j_4=1}^{n} \mathrm{E}\Big[\big\{ \zeta_{i_1,j_1,j_2} - \mathrm{E}[\zeta_{i_1,j_1,j_2}|j_1] - \mathrm{E}[\zeta_{i_1,j_1,j_2}|j_2] \big\} \big\{ \zeta_{i_2,j_3,j_4} - \mathrm{E}[\zeta_{i_2,j_3,j_4}|j_3] - \mathrm{E}[\zeta_{i_2,j_3,j_4}|j_4] \big\}^\top \Big]. \tag{D.7}$$

The expectations on the latter r.h.s. are zero when $\{j_1, j_2\} \cap \{j_3, j_4\} = \emptyset$ due to independence and the fact that the terms $\zeta_{i,j,j'}$ are centered. Otherwise, there is at least an identity among the indices j_k, $k \in \{1, \ldots, 4\}$. For instance, assume $j_1 = j_3 = j$ and $j \ne j_2 \ne j_4$. Then,

$$\mathrm{E}\Big[\big\{ \zeta_{i_1,j,j_2} - \mathrm{E}[\zeta_{i_1,j,j_2}|j] - \mathrm{E}[\zeta_{i_1,j,j_2}|j_2] \big\} \big\{ \zeta_{i_2,j,j_4} - \mathrm{E}[\zeta_{i_2,j,j_4}|j] - \mathrm{E}[\zeta_{i_2,j,j_4}|j_4] \big\}^\top \Big]$$
$$= \mathrm{E}\Big[\big\{ \zeta_{i_1,j,j_2} - \mathrm{E}[\zeta_{i_1,j,j_2}|j] \big\} \big\{ \zeta_{i_2,j,j_4} - \mathrm{E}[\zeta_{i_2,j,j_4}|j] \big\}^\top \Big] = \mathrm{E}\Big[\mathrm{E}\Big[\big\{ \zeta_{i_1,j,j_2} - \mathrm{E}[\zeta_{i_1,j,j_2}|j] \big\} \big\{ \zeta_{i_2,j,j_4} - \mathrm{E}[\zeta_{i_2,j,j_4}|j] \big\}^\top \Big| j \Big] \Big]$$
$$= \mathrm{E}\Big[\mathrm{E}\big[\zeta_{i_1,j,j_2} \zeta^\top_{i_2,j,j_4} |j \big] \Big] - \mathrm{E}\Big[\mathrm{E}\big[\zeta_{i_1,j,j_2} |j \big] \mathrm{E}\big[\zeta^\top_{i_2,j,j_4} |j \big] \Big] = 0.$$

Due to the symmetry of the latter cross-products, all cases of a single identity among the j_k, $k \in \{1, \ldots, 4\}$, yield the same result. Therefore, we need (at least) two identities among them to obtain non zero covariances in the calculation of $\mathrm{E}[\mathbf{T}_{4,3}\mathbf{T}_{4,3}^\top]$. Thus, let us assume that $j_1 = j_3$ and $j_2 = j_4$. Then, the corresponding terms in (D.7) are

$$\frac{\tilde{r}^2_{n,n'}}{(n')^2 n^4} \sum_{i_1,i_2=1}^{n'} \sum_{j_1,j_2=1}^{n} \mathrm{E}\Big[\big\{ \zeta_{i_1,j_1,j_2} - \mathrm{E}[\zeta_{i_1,j_1,j_2}|j_1] - \mathrm{E}[\zeta_{i_1,j_1,j_2}|j_2] \big\} \times \big\{ \zeta_{i_2,j_1,j_2} - \mathrm{E}[\zeta_{i_2,j_1,j_2}|j_1] - \mathrm{E}[\zeta_{i_2,j_1,j_2}|j_2] \big\}^\top \Big]$$
$$= \frac{\tilde{r}^2_{n,n'}}{(n')^2 n^4} \sum_{i_1,i_2=1}^{n'} \sum_{j_1,j_2=1}^{n} \Big(\mathrm{E}\big[\zeta_{i_1,j_1,j_2} \zeta^\top_{i_2,j_1,j_2} \big] - 2\mathrm{E}\big[\mathrm{E}[\zeta_{i_1,j_1,j_2}|j_1] \mathrm{E}[\zeta_{i_2,j_1,j_2}|j_1]^\top \big] \Big) =: v_{4,3,1} - v_{4,3,2}.$$

By now usual techniques, we get

$$
\begin{aligned}
v_{4,3,1} &= \frac{\tilde r^2_{n,n'}}{(n')^2 n^4} \sum_{i_1,i_2=1}^{n'} \sum_{j_1,j_2=1}^{n} \mathrm{E}\Big[\zeta_{i_1,j_1,j_2}\zeta^\top_{i_2,j_1,j_2}\Big] \simeq \frac{\tilde r^2_{n,n'}}{(n')^2 n^4} \sum_{i=1}^{n'} \sum_{j_1,j_2=1}^{n} \mathrm{E}\Big[\zeta_{i,j_1,j_2}\zeta^\top_{i,j_1,j_2}\Big] \\
&\simeq \frac{\tilde r^2_{n,n'}}{(n')^2 n^2} \sum_{i=1}^{n'} \frac{\overline{\psi}(\mathbf{z}'_i)\overline{\psi}^\top(\mathbf{z}'_i)}{f^4_{\mathbf{Z}}(\mathbf{z}'_i)} \int K_h^2(\mathbf{z}'_i-\mathbf{z}_1)K_h^2(\mathbf{z}'_i-\mathbf{z}_2)\big(g^*(\mathbf{x}_1,\mathbf{x}_2)-\tau_i\big)^2 f(\mathbf{x}_1,\mathbf{z}_1) \\
&\qquad \times f(\mathbf{x}_2,\mathbf{z}_2)\,d\mathbf{x}_1\,d\mathbf{z}_1\,d\mathbf{x}_2\,d\mathbf{z}_2 = O\Big(\frac{\tilde r^2_{n,n'}}{n'n^2h^{2p}}\Big) = O\Big(\frac{1}{nh^p}\Big) = o(1).
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
v_{4,3,2} &= \frac{\tilde r^2_{n,n'}}{(n')^2 n^4} \sum_{i_1,i_2=1}^{n'} \sum_{j_1,j_2=1}^{n} \mathrm{E}\Big[\mathrm{E}[\zeta_{i_1,j_1,j_2}|j_1]\mathrm{E}[\zeta^\top_{i_2,j_1,j_2}|j_1]\Big] \simeq \frac{\tilde r^2_{n,n'}}{(n')^2 n^4} \sum_{i=1}^{n'} \sum_{j_1,j_2=1}^{n} \mathrm{E}\Big[\mathrm{E}[\zeta_{i,j_1,j_2}|j_1]\mathrm{E}[\zeta^\top_{i,j_1,j_2}|j_1]\Big] \\
&\simeq \frac{\tilde r^2_{n,n'}}{(n')^2 n^2} \sum_{i=1}^{n'} \frac{\overline{\psi}(\mathbf{z}'_i)\overline{\psi}^\top(\mathbf{z}'_i)}{f^4_{\mathbf{Z}}(\mathbf{z}'_i)} \int K_h^2(\mathbf{z}'_i-\mathbf{z}_1)K_h(\mathbf{z}'_i-\mathbf{z}_2)K_h(\mathbf{z}'_i-\mathbf{z}_3)\big(g^*(\mathbf{x}_1,\mathbf{x}_2)-\tau_i\big) \\
&\qquad \times \big(g^*(\mathbf{x}_1,\mathbf{x}_3)-\tau_i\big) f(\mathbf{x}_1,\mathbf{z}_1)f(\mathbf{x}_2,\mathbf{z}_2)f(\mathbf{x}_3,\mathbf{z}_3)\,d\mathbf{x}_1\,d\mathbf{z}_1\,d\mathbf{x}_2\,d\mathbf{z}_2\,d\mathbf{x}_3\,d\mathbf{z}_3 = O\Big(\frac{\tilde r^2_{n,n'}}{n'n^2h^{p}}\Big) = O\Big(\frac{1}{n}\Big) = o(1).
\end{aligned}
$$

Another case of two identities occurs when $j_1 = j_4$ and $j_2 = j_3$, but it can be dealt similarly. Then, we have proved that $\mathrm{E}[\mathbf{T}_{4,3}\mathbf{T}^\top_{4,3}] = o(1)$ and $\mathbf{T}_{4,3} = o_P(1)$. □

Convergence of $\mathbf{T}_6$ to 0. Replacing $\hat f_{\mathbf{Z}}$ in the definition of $\mathbf{T}_6$ above by the normalized sum of the kernels, we get

$$
\begin{aligned}
\mathbf{T}_6 &= \frac{\tilde r_{n,n'}}{n'n^2} \sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n} \frac{K_h(\mathbf{z}'_i-\mathbf{Z}_{j_1})K_h(\mathbf{z}'_i-\mathbf{Z}_{j_2})}{f^3_{\mathbf{Z}}(\mathbf{z}'_i)} \big(\mathrm{E}[\hat f_{\mathbf{Z}}(\mathbf{z}'_i)] - f_{\mathbf{Z}}(\mathbf{z}'_i)\big)\Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2}) - \mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}'_i\big]\Big)\overline{\psi}(\mathbf{z}'_i) \\
&+ \frac{\tilde r_{n,n'}}{n'n^3} \sum_{i=1}^{n'}\sum_{j_1=1}^{n}\sum_{j_2=1}^{n}\sum_{j_3=1}^{n} \frac{K_h(\mathbf{z}'_i-\mathbf{Z}_{j_1})K_h(\mathbf{z}'_i-\mathbf{Z}_{j_2})}{f^3_{\mathbf{Z}}(\mathbf{z}'_i)} \big(K_h(\mathbf{z}'_i-\mathbf{Z}_{j_3}) - \mathrm{E}[\hat f_{\mathbf{Z}}(\mathbf{z}'_i)]\big) \\
&\times \Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2}) - \mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}'_i\big]\Big)\overline{\psi}(\mathbf{z}'_i) =: \mathbf{T}_{6,1} + \mathbf{T}_{6,2}.
\end{aligned}
$$

The first term $\mathbf{T}_{6,1}$ is a bias term. By Assumptions 2–3, $\sup_{i\in\{1,\dots,n'\}} \big|\mathrm{E}[\hat f_{\mathbf{Z}}(\mathbf{z}'_i)] - f_{\mathbf{Z}}(\mathbf{z}'_i)\big| = O(h^\alpha)$. The sum of the diagonal terms in $\mathbf{T}_{6,1}$ is

$$
-\frac{\tilde r_{n,n'}}{n'n^2} \sum_{i=1}^{n'}\sum_{j=1}^{n} \frac{K_h^2(\mathbf{z}'_i-\mathbf{Z}_j)}{f^3_{\mathbf{Z}}(\mathbf{z}'_i)} \big(\mathrm{E}[\hat f_{\mathbf{Z}}(\mathbf{z}'_i)] - f_{\mathbf{Z}}(\mathbf{z}'_i)\big)\mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}'_i\big]\overline{\psi}(\mathbf{z}'_i),
$$

that is $O_P\big(\tilde r_{n,n'}h^\alpha/(nh^p)\big)$. The sum of the extra-diagonal terms in $\mathbf{T}_{6,1}$ is the r.v.

$$
\overline{\mathbf{T}}_{6,1} := \frac{\tilde r_{n,n'}}{n'n^2} \sum_{i=1}^{n'} \sum_{1\le j_1\ne j_2\le n} \frac{K_h(\mathbf{z}'_i-\mathbf{Z}_{j_1})K_h(\mathbf{z}'_i-\mathbf{Z}_{j_2})}{f^3_{\mathbf{Z}}(\mathbf{z}'_i)} \big\{\mathrm{E}[\hat f_{\mathbf{Z}}(\mathbf{z}'_i)] - f_{\mathbf{Z}}(\mathbf{z}'_i)\big\}\Big(g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2}) - \mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}'_i\big]\Big)\overline{\psi}(\mathbf{z}'_i).
$$

Note that $\mathbf{z} \mapsto f_{\mathbf{Z}}(\mathbf{z})$ and $(\mathbf{z}_1,\mathbf{z}_2) \mapsto \mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{z}_1,\mathbf{Z}_2=\mathbf{z}_2\big]$ are α-times continuously differentiable on $\mathscr{Z}$ and $\mathscr{Z}^2$ respectively, because of Assumptions 3 and 5. By α-order Taylor expansions of such terms, they yield some factors h^α. It is easy to check that the expectation of $(\overline{\mathbf{T}}_{6,1})^2$ is of order $\tilde r^2_{n,n'}h^{2\alpha}/(n^2h^{2p})$. Therefore,

$$
\mathbf{T}_{6,1} = O_P\Big(\frac{\tilde r_{n,n'}h^\alpha}{nh^p}\Big) = O_P\Big(\frac{(n')^{1/2}h^\alpha}{\sqrt{nh^p}}\Big) = o_P(1).
$$

Concerning $\mathbf{T}_{6,2}$, 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 $O_P(\tilde r_{n,n'}h^\alpha/(nh^p))$ as above, and they are negligible. Once we remove such terms from the triple sums (indexed by (j_1,j_2,j_3)) defining $\mathbf{T}_{6,2}$, we get the centered r.v. $\overline{\mathbf{T}}_{6,2}$. Let us calculate the second moment of $\overline{\mathbf{T}}_{6,2}$.

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

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, we obtain two extra factors h^α through α-order limited expansions of $(\mathbf{z}_1,\mathbf{z}_2)\mapsto\mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{z}_1,\mathbf{Z}_2=\mathbf{z}_2\big]$. This yields an order $O(nn'h^{p+2\alpha}/(nh^p))$. When j_3 and j_6 are equal to two different indices ($j_3=j_4$ and $j_6=j_2$, e.g.), we lose another factor 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 $\mathrm{E}\big[\overline{T}^2_{6,2}\big]=O\Big(nn'h^{p+2\alpha}/(nh^p)\Big)=o(1)$.

Globally, we obtain $\mathbf{T}_6\to\mathbf{0}$ in probability under Assumption 7(ii)(a). □

Convergence of $\mathbf{T}_7$ to 0. Since $\sup_{i\in\{1,\dots,n'\}}|\hat f_{\mathbf{Z}}(\mathbf{z}'_i)-f_{\mathbf{Z}}(\mathbf{z}'_i)|=o_P(1)$, note that

$$\sup_{i\in\{1,\dots,n'\}}|\mathbf{T}_{7,i}|\le\frac{6}{f^4_{\mathbf{Z},min}}\sup_{i\in\{1,\dots,n'\}}|\hat f_{\mathbf{Z}}(\mathbf{z}'_i)-f_{\mathbf{Z}}(\mathbf{z}'_i)|^2,$$

with a probability arbitrarily close to one. Apply Lemma 13 with a fixed $t>0$ and $\mathbf{z}=\mathbf{z}'_i$ for each $i\in\{1,\dots,n'\}$

$$\Pr\bigg(\sup_{i\in\{1,\dots,n'\}}|\mathbf{T}_{7,i}|\ge\frac{6}{f^4_{\mathbf{Z},min}}\Big(\frac{C_{K,\alpha}h^\alpha}{\alpha!}+t\Big)^2\bigg)\le 2n'\exp\bigg(-\frac{nh^pt^2}{2f_{\mathbf{Z},max}\int K^2+(2/3)C_Kt}\bigg).$$

Set $t\propto h^{\alpha/2}$. Deduce $\sup_{i\in\{1,\dots,n'\}}|\mathbf{T}_{7,i}|=O_P(h^\alpha)$ since $nh^{p+\alpha}/\ln n'\to\infty$ by assumption. Then,

$$|\mathbf{T}_7|\le\frac{\tilde r_{n,n'}}{n'n^2}\sup_i|T_{7,i}|\sum_{i=1}^{n'}|\overline{\boldsymbol{\psi}}(\mathbf{z}'_i)|\sum_{j_1=1}^n\sum_{j_2=1}^n|K|_h(\mathbf{z}'_i-\mathbf{Z}_{j_1})|K|_h(\mathbf{z}'_i-\mathbf{Z}_{j_2})\Big|g^*(\mathbf{X}_{j_1},\mathbf{X}_{j_2})-\mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{Z}_2=\mathbf{z}'_i\big]\Big|.$$

The expectation of the double sum is $O(h^\alpha)$, by an α-order limited expansion of $(\mathbf{z}_1,\mathbf{z}_2)\mapsto\mathrm{E}\big[g^*(\mathbf{X}_1,\mathbf{X}_2)|\mathbf{Z}_1=\mathbf{z}_1,\mathbf{Z}_2=\mathbf{z}_2\big]$. Then, by Markov's inequality, we deduce $\mathbf{T}_7=O_P(\tilde r_{n,n'}\sup_i|T_{7,i}|h^\alpha)=O_P(\tilde r_{n,n'}h^{2\alpha})=O_P\big((n'nh^{p+4\alpha})^{1/2}\big)$, and then $\mathbf{T}_7=o_P(1)$ due to Assumption 7(ii)(a). □

Convergence of $\mathbf{T}_3$ to 0. For every $\epsilon>0$, by Markov's inequality,

$$\Pr(|\mathbf{T}_3|>\epsilon)\le\frac{C_{\Lambda''}\tilde r_{n,n'}}{2n'\epsilon}\sum_{i=1}^{n'}\mathrm{E}\Big[\big(\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}-\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}\big)^2\Big]\times|\boldsymbol{\psi}(\mathbf{z}'_i)|.$$

An approximated calculation of $\mathrm{E}\Big[\big(\hat\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}-\tau_{1,2|\mathbf{Z}=\mathbf{z}'_i}\big)^2\Big]$ can be obtained following the steps of the proof of Lemma 17. 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 $O\big((nh^p)^{-1}\big)$, we get $\Pr(|\mathbf{T}_3|>\epsilon)\le B\tilde r_{n,n'}/(nh^p\epsilon)$, for some constant B. Since $n'/(nh^p)\to 0$, we get $\mathbf{T}_3=o_P(1)$, as claimed. □

Appendix E. Supplementary data

The supplementary file contains a simulation study (with different estimators, DGPs and model specifications), a discussion around the numerical complexity of our algorithms, and some guidelines for choosing the tuning parameter. Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jmva.2020.104610.

References

[1] E. Acar, C. Genest, J. Nešlehová, Beyond simplified pair-copula constructions, J. Multivariate Anal. 110 (2012) 74–90.
[2] P. Bellec, G. Lecué, A. Tsybakov, Slope meets lasso: improved oracle bounds and optimality, Ann. Statist. 46 (6B) (2018) 3603–3642.

[3] P. Bickel, Y. Ritov, A. Tsybakov, Simultaneous analysis of Lasso and Dantzig selector, Ann. Statist. 37 (4) (2009) 1705–1732.
[4] X. Chen, Large sample sieve estimation of semi-nonparametric models, Handb. Econom. 6 (2007) 5549–5632.
[5] A. Derumigny, J.-D. Fermanian, About tests of the ''simplifying'' assumption for conditional copulas, Depend. Model. 5 (1) (2017) 154–197.
[6] A. Derumigny, J.-D. Fermanian, A classification point-of-view about conditional Kendall's tau, Comput. Statist. Data Anal. 135 (2019) 70–94.
[7] A. Derumigny, J.-D. Fermanian, On kernel-based estimation of conditional Kendall's tau: finite-distance bounds and asymptotic behavior, Depend. Model. 7 (1) (2019) 292–321.
[8] J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, J. Amer. Statist. Assoc. 96 (456) (2001) 1348–1360.
[9] I. Gijbels, N. Veraverbeke, M. Omelka, Conditional copulas, association measures and their applications, Comput. Statist. Data Anal. 55 (5) (2011) 1919–1932.
[10] I. Hobæk Haff, K. Aas, A. Frigessi, On the simplified pair-copula construction–simply useful or too simplistic? J. Multivariate Anal. 101 (2010) 1296–1310.
[11] K. Kato, Asymptotics for argmin processes: Convexity arguments, J. Multivariate Anal. 100 (8) (2009) 1816–1829.
[12] M. Kurz, F. Spanhel, Testing the simplifying assumption in high-dimensional vine copulas, 2017, arXiv preprint arXiv:1706.02338.
[13] R. Nelsen, An Introduction to Copulas, Springer Science & Business Media, 2007.
[14] A. Van der Vaart, Asymptotic Statistics, Vol. 3, Cambridge university press, 2000.
[15] N. Veraverbeke, M. Omelka, I. Gijbels, Estimation of a conditional copula and association measures, Scand. J. Stat. 38 (4) (2011) 766–780.
[16] H. Zou, The adaptive lasso and its oracle properties, J. Amer. Statist. Assoc. 101 (476) (2006) 1418–1429.