

# AVERAGE DENSITY: WEAK LIMITS AND INFERENCE IN NON-REGULAR SEMI-PARAMETRIC PROBLEMS

PAUL DELATTE

*University of Southern California, [delatte@usc.edu](mailto:delatte@usc.edu)*

In this paper, we derive novel convergence and inference results for the optimal kernel-based plug-in estimator of the expected value of a density considered in [Giné and Nickl \(2008a\)](#). We show that the estimator remains asymptotically normal when the density is highly irregular and when the bandwidth converges to zero very rapidly. In both cases, however, the convergence rate of the estimator is slower than parametric, the asymptotic variance depends on the kernel, the plug-in variance estimator is inconsistent, and the non-parametric bootstrap fails. While this limits the possibility of inference, we show that the problem is not impossible and demonstrate how to construct variance estimators that are consistent in both regular and non-regular cases. The positive results in this paper thus extend the insights of [Cattaneo, Crump, and Jansson \(2014b\)](#) by providing support for under-smoothing as a robust practice even in problems where the nuisance parameter may be highly irregular. By contrast, the negative results we obtain under weak regularity conditions cast some light on the fundamental limits of inference in semi-parametric problems. They indicate what happens when the high-level conditions of [Newey \(1994\)](#) or [Chen, Linton, and Van Keilegom \(2003\)](#) break down. This should be appreciated in virtue of the paradigmatic simplicity of average density estimation among all semi-parametric problems. From a technical viewpoint, a core contribution of this paper is to connect the traditional "low smoothness asymptotics" as found in [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#) with the "small bandwidth asymptotics" introduced in [Cattaneo, Crump, and Jansson \(2014b\)](#) for kernel-based estimators. This is made possible by the assumption-lean moment bounds we directly derive for the estimator.

## 1 Introduction

### 1.1 Problem and results overview

Problems of inference on low-dimensional parameters in the presence of infinite-dimensional nuisance parameters abound in applications. When the low-dimensional parameter is expressed as a functional of the underlying distribution and the nuisance parameter, an intuitive two-step procedure consists in plugging a non-parametric estimator of the nuisance parameter in an empirical counterpart to the functional defining the low-dimensional parameter. The most standard case obtains when the low-dimensional parameter is expressed through a moment condition whose empirical counterpart is obtained by plugging the empirical measure. This two-step procedure,

which is alternatively known as semi-parametric plug-in estimation by re-substitution, has been extensively studied. The resulting estimators are commonly refined by some easy-to-implement bias-correction (with or without sample splitting). Some strong convergence results based on "high-level" conditions are available for such estimators. In particular, for low-dimensional parameters expressed through moment conditions, it is known that if the non-parametric estimator for the infinite-dimensional nuisance parameter is strongly consistent at a fast enough rate, then the estimator for the low-dimensional parameter is asymptotically linear and has a  $\sqrt{n}$ -normal weak limit independent of the non-parametric estimator; see [Andrews \(1994\)](#), [Newey \(1994\)](#), [Chen, Linton, and Van Keilegom \(2003\)](#), or, more recently, [Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins \(2018\)](#). However, one may be worried that the non-parametric estimator is not as good in practice as required by theory. This may happen for a number of reasons: the true nuisance parameter may lie in too large a class due to its inherent irregularity or to the dimension of the data; its estimator may not have been chosen carefully enough. Even for well-behaved problems with well-performing estimators, the strong consistency rates required in theory may be hard to reach and impossible to verify: it will often depend on a careful choice of some hyperparameters that ultimately depends on the unknown degree of regularity of the nuisance parameter. This justifies looking at what happens in terms of weak convergence and inference when the consistency rates for the non-parametric first step are relaxed. This leads to what can be labeled as non-regular semi-parametric functional estimation problems, that is, semi-parametric problems of the form described above for which asymptotic linearity subsides and parametric rates of convergence cannot be achieved anymore.

An encompassing theory of non-regular inference holding for all semi-parametric functional estimation problems is inherently hard due to the paucity of general weak convergence results for asymptotically non-linear statistics. For some problems, however, the non-linearity of the statistics is regular enough so that one may hope to invoke some known weak convergence results. Among them, problems where the estimator admits a U-statistics representation are good candidates for tractable analysis. We are not the first one to make this point and explore problems of this form (see [Cattaneo, Crump, and Jansson \(2014b\)](#) and [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#)), but results are still few and sparse. We contribute to this literature by providing new non-regular convergence and inference results for estimating the expected value of a density by an optimal plug-in kernel density estimator. This estimator and its optimal properties were first considered by [Hall and Marron \(1987\)](#) and [Giné and Nickl \(2008a\)](#). Our new non-regular results are of interest for at least three reasons.

1. The expected value of a density or average density, which rewrites as the integrated square of a density under domination, is in itself an important statistical object for which valid inferential rules have often been sought. It is the essential part of the Rényi entropy of order 2 and is used, for instance, in the estimation of the Shannon entropy (see [Laurent \(1996\)](#)) or in the construction of adaptive confidence sets (see [Robins and van der Vaart \(2006\)](#)). The

first appeal of our results is that they extend the range of cases where valid inference for the average density is possible.

2. To obtain these results, we derive, in the terminology of [Cattaneo, Crump, and Jansson \(2014b\)](#), the "small bandwidth asymptotics" of the optimal estimator we consider and we do so without smoothness assumptions on the density. This allows to formally connect this regime with the "low smoothness asymptotics" (under optimal bandwidth sequences) traditionally considered as in [Giné and Nickl \(2008a\)](#) or [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#). This connection provides new insights into the choice of hyper-parameters in two-step semi-parametric problems. Indeed, the results we obtain can be understood as formally justifying under-smoothing as a valid practice to gain robustness for inference even when the nuisance parameter is highly irregular.
3. However, the validity of under-smoothing depends fundamentally on tailored solutions that leverage essential structures of the problem and estimator. In absence of these constructions, standard inferential procedures, such as the plug-in principle for variance estimation and the non-parametric bootstrap, break down as soon as the problem is non-regular. Because average density estimation is one of the simplest semi-parametric problems and, as a consequence, comes with known optimality frontiers and simple optimal estimators as the one we consider, the results we obtain can be fairly interpreted as upper bounds in terms of validity for non-regular inference in more general semi-parametric functional estimation problems. This point is salient given the renewed interest for semi-parametric plug-in estimation by re-substitution using sample splitting (see, for instance, [Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins \(2018\)](#)). One may legitimately conjecture that, at best, similar failures as displayed in this paper obtain in non-regular regimes for more complicated problems where optimal bias-corrected plug-in estimators are not even available.

From a purely technical viewpoint, our inferential results also contribute to the literature on inference based on U-statistics with  $n$ -dependent kernel complementing the results obtained in [Hardle and Mammen \(1993\)](#) and [Cattaneo, Crump, and Jansson \(2014b,a\)](#).

The rest of the paper is constructed as follows. In the remaining parts of the Introduction, we first review previous results in the literature on which we build, then introduce the estimator and the running hypotheses for the problem, and finally review Hoeffding's decomposition for U-statistics. In Section 2, we derive the weak limits of the estimator in two regimes. We first derive the weak limits by varying the rate of convergence for the bandwidth of the kernel estimator while keeping the smoothness class of the true density fixed to some arbitrary level. We then show that these results can be directly used to derive the weak limits of the estimator by varying the degrees of smoothness of the true density while fixing the bandwidth sequence to the optimal one (trading off squared bias against variance). The conclusion is that the estimator remains asymptotically normal when the bandwidth sequence converges to zero very rapidly even if the density is highly irregular. In these cases, however, the rate is no longer parametric and the variance depends on the

density estimator. Given these newly derived weak limits, we then tackle the problem of inference in Section 3. We show that while valid inference remains possible, many standard procedures break down in spite of asymptotic normality. By leveraging the structure of the problem, we first build a simple variance estimator and show its consistency by constructive methods in regular and non-regular cases. We then show that the plug-in variance estimator is inconsistent in the non-regular cases, but that a simple bias-correction restores consistency. These results are then used to show that the non-parametric bootstrap fails in the non-regular cases.

## 1.2 Previous results and related literature

The fact that the estimator we consider has a normal weak limit at the parametric rate in the regular regime (that is, for a smooth enough density and an optimal bandwidth sequence) was proved in [Giné and Nickl \(2008a\)](#). As far as we know, the existence and characterization of the weak limits in non-regular regimes, either under low smoothness or small bandwidth, have not been obtained before. [Giné and Nickl \(2008a\)](#) proved tightness for our problem under low smoothness, but not the existence of a weak limit, nor its nature. [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#) proved a weak limit result under low smoothness but for a Fourier series estimator of the average density (and did not consider inference). [Cattaneo, Crump, and Jansson \(2014b\)](#) proved a weak limit result under small bandwidth but for a different problem in a high-regularity setting (the density-weighted average derivative of a regression function). [Cattaneo, Farrell, Jansson, and Masini \(2024\)](#) recently augmented these results by proving that the "small bandwidth asymptotics" led to smaller higher-order approximation errors compared to the standard asymptotics framework based on linearity. [Cattaneo and Jansson \(2018\)](#) considered another "small bandwidth asymptotics" in high-regularity settings when the slower bandwidth sequences generate a non-negligible bias effect (due to non-linearity or the presence of diagonal elements, and not smoothing "[which their] theory is largely silent about"). [Cattaneo and Jansson \(2018\)](#) considered the average density problem to illustrate their distributional results but only for pedagogical reasons as they purposefully considered "non-optimal" estimators displaying some "nonlinearity" bias or some "leave-in" bias (in their terminology). Their results do not apply to our setting since the estimator we consider is already optimally de-biased by design and thus only presents an unavoidable "smoothing" bias. Our results can thus be alternatively understood as tackling the "smoothing bias" in the case of the average density. Finally, [Cattaneo and Jansson \(2022\)](#) investigated the relationship between efficiency and bootstrap consistency for different estimators of the average density, including the leave-one-out kernel-based one of [Hall and Marron \(1987\)](#) that we consider, but the authors did not consider weak convergence in non-regular regimes (nor inference).

To prove our weak limit result in non-regular regimes, we have to bound higher moments of the estimator we consider. We do this without making use of the smoothness of the density and so resort to arguments that bear more resemblance to the ones used in [Giné and Nickl \(2008a\)](#) and [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#) than to the ones in [Cattaneo, Crump, and Jansson \(2014b\)](#) where the authors resorted to previously derived moment bounds obtained

by leveraging the high smoothness assumptions of their problem (see Remark 1.5 and Remark 2.4 for more on this point). Once the tedious moment bounds are derived in our problem, we resort to more standard arguments. Our proof of asymptotic normality makes direct use of the central limit theorem in [de Jong \(1987\)](#) and its direct extension in [Eubank and Wang \(1999\)](#). A similar line of argument with a kernel-based estimator can be found in [Hardle and Mammen \(1993\)](#). The same idea is used again in [Cattaneo, Crump, and Jansson \(2014b\)](#) for the density-weighted average derivative of a regression function. All these results can be traced back to [Hall \(1984\)](#) who first proved a central limit theorem for  $n$ -dependent kernel U-statistics when deriving the weak limit of the integrated square error of a kernel density estimator. These central limit theorems for quadratic forms seemingly differ from the one developed in [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#). The similarities and differences between these results are briefly explored in Remark 2.3.

Given that the non-regular weak limits we derive are new, our inference results based on them are also new. These results, as they apply to some U-statistics with  $n$ -dependent kernels, fit in a larger literature. However, as far as we know, very few results are available for such statistics: we only know of the results in [Hardle and Mammen \(1993\)](#) and [Cattaneo, Crump, and Jansson \(2014b,a\)](#). It is, however, an interesting question from the standpoint of statistical theory as such statistics exhibit non-standard behaviors and serve as useful counterexamples to some widely held folk results, in particular with respect to the validity of the bootstrap. Their intricacy is not reducible to their non-linearity, but depends fundamentally on the sample size dependence of their higher order terms. This point was duly noted in [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#) in the case of the weak limits. It also holds when it comes to inference as U-statistics with  $n$ -dependent display non-standard patterns that were first exhibited in [Hardle and Mammen \(1993\)](#). Our results confirm what was unearthed in this paper and reaffirmed in [Cattaneo, Crump, and Jansson \(2014a\)](#). We complement these results by introducing a simpler consistent variance estimator whose structure can be easily adapted to other similar problems such as the one in [Cattaneo, Crump, and Jansson \(2014b\)](#). Our results also contribute to the important problem of bootstrap validity in semi-parametric problems (see [Chen, Linton, and Van Keilegom \(2003\)](#), [Kosorok \(2008\)](#), [Cheng and Huang \(2010\)](#)): our result shows that as soon as the regularity conditions are not sufficient to deliver parametric rates, the non-parametric bootstrap fails, in spite of asymptotic normality.

*Remark 1.1.* A terminological clarification is in order given the divergent nomenclatures across fields. The class of semi-parametric procedures we consider in this paper is based on a double plug-in strategy: first, a non-parametric estimator for the nuisance is plugged in the functional defining the low-dimensional parameter; then, an estimator of the distribution is plugged in the functional equation defining the low-dimensional parameter. This procedure is the most widely used for semi-parametric inference of a low-dimensional parameter in presence of an infinite-dimensional nuisance parameter. It has been sometimes referred to as semi-parametric M-estimation or Z-estimation in econometrics (see [Newey \(1994\)](#), [Chen, Linton, and Van Keilegom \(2003\)](#), or [Delsol and Van Keilegom \(2020\)](#)). However, this may be a source of confusion. If it is true that the

procedure in the second stage may be seen as parametric Z-estimation, the whole procedure does not correspond to what is commonly referred to as semi-parametric Z- or M-estimation. This is due to the fact that the first step non-parametric estimator need not be obtained as solution to an equation nor as solution to an optimization problem. This procedure should be contrasted with proper semi-parametric Z- or M-estimation where both the low-dimensional estimator and the nuisance estimator are solutions to a joint equation or a joint optimization problem; standard examples of this procedure include semi-parametric maximum likelihood estimation and semi-parametric least squares estimation (see [Kosorok \(2008\)](#) and [Cheng and Huang \(2010\)](#)).

### 1.3 Estimator and hypotheses

Let  $(X_1, \dots, X_n)$  be i.i.d. real-valued random variables with common distribution  $P_X$  admitting a square integrable density  $f_0$  with respect to the Lebesgue measure on  $\mathbb{R}$ . The parameter of interest is the expectation of  $f_0(X)$  with respect to  $P_X$  and rewrites as the integrated square of  $f_0$  by domination. That is,

$$\theta_0 = \mathbb{E} [f_0(X)] = \int_{\mathbb{R}} f_0^2(x) dx.$$

We consider the estimator of  $\theta_0$  introduced in [Hall and Marron \(1987\)](#) and further studied in [Giné and Nickl \(2008a\)](#) given by

$$U_n = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{1}{h_n} K\left(\frac{X_i - X_j}{h_n}\right),$$

where  $K: \mathbb{R} \rightarrow \mathbb{R}$  is a smoothing kernel with associated bandwidth  $h_n$ . The estimator  $U_n$  is obtained by first plugging a kernel density estimator in the empirical counterpart to the moment condition defining  $\theta_0$  and then removing the diagonal elements in the double sum. The estimator  $U_n$  is directly seen to be a second-order U-statistics with  $n$ -dependent kernel  $k_n$  given by

$$k_n(X_i, X_j) = \frac{1}{h_n} K\left(\frac{X_i - X_j}{h_n}\right).$$

On top of the domination assumption on  $P_X$ , we reuse the standard assumptions considered in [Giné and Nickl \(2008a\)](#) for the kernel function  $K$  and the unknown true density  $f_0$ . We denote these assumptions Assumption K and Assumption D( $s$ ), respectively. Fundamentally, the regularity of the problem is controlled by assuming that  $f_0$  belongs to a Sobolev space and by varying its smoothness parameter  $s > 0$ : the lower the value of  $s$ , the less regular the density  $f_0$ . Before formally stating these assumptions, we introduce some standard notations. For  $1 \leq p < \infty$ , we denote by  $L^p = L^p(\mathbb{R}) = L^p(\mathbb{R}; \lambda)$  the space of  $p$ -integrable functions with respect to the Lebesgue measure and endow it with the  $p$ -norm  $\|\phi\|_p^p = \int_{\mathbb{R}} \phi(x)^p dx$ . For  $\phi \in L^1$ , we define the Fourier transform by  $F\phi(u) = \int_{\mathbb{R}} e^{-iux} \phi(x) dx$  and we extend it by continuity to  $L^2$ .

**Assumption K.** *The kernel  $K: \mathbb{R} \rightarrow \mathbb{R}$  satisfies:*

1.  *$K$  is symmetric and bounded;*

2.  $\int K(u) du = 1$ ;
3.  $\int |K(u)||u| du < \infty$ .

*Remark 1.2.* By assuming  $K$  bounded and integrable, we have  $K \in L^1 \cap L^\infty$ . It follows then that  $K \in L^p$  for  $1 \leq p \leq \infty$ . Indeed, given an arbitrary measure  $\mu$  on a space  $X$ , define  $A = \{x \in X : |f(x)| > 1\}$ , then  $\mu(A) \leq \int_A |f(x)| d\mu \leq \|f\|_1$  and  $|f(x)|^p \leq |f(x)|$  on  $X \setminus A$ , hence  $\int_X |f(x)|^p d\mu \leq \int_A |f(x)|^p d\mu + \int_{X \setminus A} |f(x)|^p d\mu \leq \|f\|_\infty^p \|f\|_1 + \|f\|_1 < \infty$ .

**Assumption D(s).** *The true density  $f_0$  satisfies:*

1.  $f_0$  is bounded;
2.  $f_0 \in H_2^s$ , where  $H_2^s = W^{2,s}(\mathbb{R})$  is the Sobolev space of integrability  $p = 2$  and of order  $s$ , that is,

$$H_2^s = \left\{ \phi \in L^2 : \|\phi\|_{2,s} = \|F\phi(\cdot)(1 + |\cdot|^2)^{s/2}\|_2 < \infty \right\}.$$

*Remark 1.3.* Since  $f_0$  is a density for some random variable, we implicitly assume that  $\int_{\mathbb{R}} f_0(u) du = 1$  and so  $f_0 \in L^1$ . Boundedness of  $f_0$  implies that  $f_0 \in L^\infty$ . Then  $f_0 \in L^1 \cap L^\infty$ , and so, by a similar argument as Remark 1.2, we have  $f \in L^p$  for any  $1 \leq p \leq \infty$ .

*Remark 1.4.* Note that we do not necessarily assume  $f_0$  to be continuous. In particular,  $f \in H_2^s$  can be discontinuous if  $s < 1/2$ , while the Sobolev embedding theorem ensures continuity of  $f \in H_2^s$  for  $s \geq 1/2$ . This is not an issue since continuity is not needed. However, we will make use of the  $L^1$  assumption. In particular, we will make use of  $L^1$ -continuity, that is, the fact that if  $f \in L^1$ , then

$$\lim_{|t| \rightarrow 0} \int |f(x+t) - f(x)| dx = 0.$$

Other results of the sort will be used; they are based on a density argument using the fact that compactly supported continuous functions are dense in  $L^1$ .

*Remark 1.5.* The smoothness and integrability assumptions in D(s) are the one considered by [Giné and Nickl \(2008a\)](#). Under Assumption D(s), the authors showed that the bias  $B_n := \mathbb{E}[U_n] - \theta_0$  of  $U_n$  satisfied  $B_n = O(h^{2s})$  where  $s$  is the smoothness parameter for the density. For completeness, the result and its proof are reproduced in Section D of the Supplementary Material. It is important to note, however, that we are able in Section 2 to derive the weak limit of  $\sigma(U_n)^{-1}(U_n - \mathbb{E}[U_n])$  without using the smoothness assumption in D(s) but only the integrability condition  $f_0 \in L^\infty$ . It is only when looking at the centered quantity  $\sigma(U_n)^{-1}(U_n - \theta_0)$  that the smoothness parameter  $s$  will play a role through the rate of decay of the bias. This has important consequences when compared to previous results in the literature as explained in Remark 2.4 and in the introduction of Section 3.

*Remark 1.6 (On relaxing D(s)).* The smoothness assumption in D(s) is already (much) more general than those used in other results in the literature bearing on sensibly similar problems – see, e.g., [Hall \(1984\)](#), [Hall and Marron \(1987\)](#), [Bickel and Ritov \(1988\)](#), [Hardle and Mammen \(1993\)](#), or [Cattaneo, Crump, and Jansson \(2014b,a\)](#). It is the same smoothness assumption as considered in [Laurent \(1996\)](#). It can be extended at no cost to the slightly more general class considered in

Laurent (2005) – see Section D of the Supplementary Material. The  $L^\infty$  integrability assumption in D(s) is more crucial, both for the bias and the weak limit, and it is unclear if it can be relaxed – see Remark 2 in Giné and Nickl (2008a) for a positive answer for the bias result in the case of the Lipschitz class of Bickel and Ritov (1988).

*Remark 1.7* (On multivariate extensions). We follow Giné and Nickl (2008a) and focus on the one-dimensional case  $d = 1$ . This allows us to reuse their Fourier argument for handling the bias without modification. It also greatly simplifies notations when deriving the tedious moment bounds for the weak limits. This also allows us to work out direct arguments from which we can unearth a simpler variance estimator than in Cattaneo, Crump, and Jansson (2014b). Extending our results to higher dimensions  $d > 1$  is of interest, especially to investigate the effects of the order of the kernel on convergence and inference. We also expect similar higher-order refinements as exhibited in Cattaneo, Crump, and Jansson (2014b) and Cattaneo, Farrell, Jansson, and Masini (2024) to hold in regular cases. These extensions are left for future research.

#### 1.4 A preliminary Hoeffding decomposition

Most of the arguments we will make depend on the Hoeffding decomposition of the second-order U-statistics  $U_n$ . This is a well-known approach that dates back to Hoeffding (1948). Because the decomposition will be used repeatedly, we collect in the next lemma the different terms entering into the decomposition. It is useful to introduce the following notations

$$\begin{aligned} u_n^0 &= \mathbb{E}[U_n] = \mathbb{E}[k_n(X_i, X_j)], \\ u_n^1(X_i) &= \mathbb{E}[k_n(X_i, X_j)|X_i] \\ u_n^2(X_i, X_j) &= k_n(X_i, X_j), \end{aligned}$$

where  $i \neq j$  are any two indexes.

**Lemma 1.1** (Hoeffding Decomposition of  $U_n$ ). *The statistics  $U_n$  admits the following Hoeffding decomposition*

$$U_n = \mathbb{E}[U_n] + 2L_n + W_n \tag{1.1}$$

where

$$\begin{aligned} \mathbb{E}[U_n] &= u_n^0, \\ L_n &= \frac{1}{n} \sum_{i=1}^n \left[ u_n^1(X_i) - u_n^0 \right], \end{aligned}$$

and

$$W_n = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left[ u_n^2(X_i, X_j) - u_n^1(X_i) - u_n^1(X_j) + u_n^0 \right].$$

*Proof.* This follows from Theorem 1 in Section 1.6. in Lee (1990). In this case, the proof is simpler. The equality follows directly by expanding the terms. The fact that  $2L_n$  is a  $L^2$ -projection follows by verifying that  $\mathbb{E}[(U_n - 2L_n) \sum_{i=1}^n g_i(X_i)] = 0$ .  $\square$

**Lemma 1.2** (Variance of  $U_n$ ). *The variance of  $U_n$  is given by*

$$\text{Var } U_n = \frac{4}{n} \text{Var}(u_n^1(X_1)) + \frac{2}{n(n-1)} \text{Var}(u_n^2(X_1, X_2) - u_n^1(X_1) - u_n^1(X_2)).$$

*Proof.* This follows directly from Theorem 4 in Section 1.6. in [Lee \(1990\)](#). For completeness, we rapidly sketch the proof. By construction,  $2L_n$  and  $W_n$  are uncorrelated. In particular, we have

$$\text{Var } U_n = 4\text{Var } L_n + \text{Var } W_n. \quad (1.2)$$

Since the components of  $L_n$  are i.i.d. (as measurable functions of  $X_i$ ), we have

$$\text{Var } L_n = \frac{1}{n} \text{Var}(u_n^1(X_1) - u_n^0) = \frac{1}{n} \text{Var}(u_n^1(X_1)). \quad (1.3)$$

Since the components of  $W_n$  are uncorrelated for any four indexes  $i < j, k < l$  such that at least three are different, we have

$$\begin{aligned} \text{Var } W_n &= \frac{2}{n(n-1)} \text{Var}(u_n^2(X_1, X_2) - u_n^1(X_1) - u_n^1(X_2) + u_n^0) \\ &= \frac{2}{n(n-1)} \text{Var}(u_n^2(X_1, X_2) - u_n^1(X_1) - u_n^1(X_2)) \\ &= \frac{2}{n(n-1)} \left[ \text{Var}(u_n^2(X_1, X_2)) - 2\text{Var}(u_n^1(X_1)) \right], \end{aligned} \quad (1.4)$$

where the last equality follows since  $\text{cov}(u_n^2(X_1, X_2), u_n^1(X_1)) = \text{Var}(u_n^1(X_1))$  and  $\text{cov}(u_n^1(X_1), u_n^1(X_2)) = 0$  (by independence of  $X_1$  and  $X_2$ ).  $\square$

From the proof, we get the alternative expression

$$\text{Var } U_n = \frac{4}{n} \text{Var}(u_n^1(X_1)) + \frac{2}{n(n-1)} \left[ \text{Var}(u_n^2(X_1, X_2)) - 2\text{Var}(u_n^1(X_1)) \right].$$

We also get the two useful inequalities

$$\text{Var } L_n \leq \frac{1}{n} \mathbb{E} [(u_n^1(X_1))^2],$$

and

$$\text{Var } W_n \leq \frac{2}{n(n-1)} \text{Var}(u_n^2(X_1, X_2)) \leq \frac{2}{n(n-1)} \mathbb{E} [(u_n^2(X_1, X_2))^2].$$

## 2 Non-regular weak limits

### 2.1 Varying bandwidth rates with fixed smoothness

In this section, we derive the weak limits of the estimator  $U_n$  for different convergence rates of the bandwidth sequence  $h_n$  to 0 as  $n \rightarrow \infty$  while keeping the smoothness of the density fixed to some arbitrary level  $s > 0$ . This translates into the following assumptions:

1. [Assumption K](#);
2. [Assumption D\(s\)](#) for  $s > 0$ ;
3. we do not fix  $h_n$  but only suppose that  $0 < h_n \rightarrow 0$  as  $n \rightarrow \infty$ .

Under these assumptions, we show in [Corollary 2.3](#) that whenever:

- $nh_n \rightarrow \infty$ : the linear term dominates and we have a standard central limit theorem with variance the semi-parametric lower bound;
- $nh_n \rightarrow (0, \infty)$ : the linear term and the second-order term have the same order and the weak limit is still normal at the standard  $n^{1/2}$  rate but the asymptotic variance depends on the kernel;
- $nh_n \rightarrow 0$ : the quadratic term dominates and the weak limit is still normal but depends on the kernel and convergence happens at a slower rate than  $n^{1/2}$  which depends on the bandwidth sequence.

This result corresponds to the "small bandwidth asymptotics" of the estimator  $U_n$  if we reuse the terminology coined in [Cattaneo, Crump, and Jansson \(2014b\)](#). The authors of this paper derived qualitatively similar results for a different problem based on a weighted regression estimator that leveraged strong regularity conditions<sup>1</sup>. The weak limit in our paper can be derived independently of the smoothness level  $s > 0$ . This is possible due to the delicate moment bounds for  $U_n$  we derive and gather in [Lemma 2.3](#), [Lemma 2.4](#), and [Lemma 2.5](#). This difference has important consequences that are leveraged in [Subsection 2.2](#) and [Section 3](#).

Before leveraging this difference, we have to prove [Corollary 2.3](#). The proof is split into three parts. We start by characterizing the asymptotic variance of  $U_n$  in [Lemma 2.1](#) and [Lemma 2.2](#) (which will be needed to obtain a closed form for the weak limit). We then derive bounds for the moments of  $U_n$ . We finally use these bounds to verify the conditions of the central limit theorem for generalized quadratic forms of [de Jong \(1987\)](#). A similar strategy was used in [Cattaneo, Crump, and Jansson \(2014b\)](#) and before in [Hardle and Mammen \(1993\)](#). This strategy, including the central limit theorem of [de Jong \(1987\)](#), can be directly traced back to [Hall \(1984\)](#) – see [Remark 2.3](#) for more on this point.

**Lemma 2.1.**

$$\lim_{n \rightarrow \infty} n \text{Var } L_n = \int_{\mathbb{R}} f_0(x)^3 dx - \left( \int_{\mathbb{R}} f_0(x)^2 dx \right)^2.$$

*Proof.* For simplicity, we write  $h_n = h$ . From [Equation \(1.3\)](#), we have

$$n \text{Var } L_n = \mathbb{E} [(u_n^1(X_1))^2] - \mathbb{E} [u_n^1(X_1)]^2.$$

---

<sup>1</sup>[Cattaneo, Crump, and Jansson \(2014b\)](#) resorted to moment bounds proved in [Robinson \(1995\)](#) and [Nishiyama and Robinson \(2000\)](#) under strong regularity conditions due to the nature of their problem.

We first have

$$\begin{aligned}\mathbb{E} [(u_n^1(X_1))^2] &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} K_h(x-y) f_0(y) dy \right)^2 f_0(x) dx \\ &= \int_{\mathbb{R}} \left( \int_{\mathbb{R}} K_h(u) f_0(x-u) du \right)^2 f_0(x) dx.\end{aligned}$$

Since  $K \in L^1$  and  $f_0 \in L^2$ , we have by the mollification theorem (see Theorem 8.14. in [Folland \(1999\)](#)) that  $\int_{\mathbb{R}} K_h(u) f_0(x-u) du$  converges in  $L^2$  to  $f_0$  as  $h \rightarrow 0$ . That is,

$$\lim_{h \rightarrow 0} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} K_h(u) f_0(x-u) du - f_0(x) \right)^2 f_0(x) dx = 0.$$

Then, by continuity of the norm, we directly get that

$$\lim_{h \rightarrow 0} \int_{\mathbb{R}} \left( \int_{\mathbb{R}} K_h(u) f_0(x-u) du \right)^2 f_0(x) dx = \int_{\mathbb{R}} f_0(x)^3 dx.$$

For the limit of  $\mathbb{E} [(u_n^1(X_1))]^2$ , we can directly invoke a density argument that extends  $L^1$ -continuity. It is proved in Section B of the Supplementary Material. We first have

$$\begin{aligned}\mathbb{E} [(u_n^1(X_1))] &= \int_{\mathbb{R}} \int_{\mathbb{R}} K_h(x-y) f_0(y) dy f_0(x) dx \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} K(u) f_0(x-uh) du f_0(x) dx \\ &= \int_{\mathbb{R}} K(u) \int_{\mathbb{R}} f_0(x-uh) f_0(x) dx du.\end{aligned}\tag{2.1}$$

Then, as proved in Section B of the Supplementary Material, we have

$$\lim_{h \rightarrow 0} \int_{\mathbb{R}} |f_0(x-uh) f_0(x) - f_0(x)^2| dx = 0.$$

Then, by dominated convergence, we have

$$\begin{aligned}\lim_{h \rightarrow 0} \int_{\mathbb{R}} K(u) \int_{\mathbb{R}} f_0(x-uh) f_0(x) dx du &= \int_{\mathbb{R}} K(u) \int_{\mathbb{R}} f_0(x)^2 dx du \\ &= \int_{\mathbb{R}} f_0(x)^2 dx.\end{aligned}$$

This concludes the proof by composition of limits.  $\square$

*Remark 2.1.* This result was already proved in Theorem 1 in [Giné and Nickl \(2008a\)](#). The proof of Giné and Nickl (which is reproduced in [Cattaneo and Jansson \(2022\)](#)) is based on showing mean squared convergence of  $L_n$  towards a i.i.d. sum with  $Y_i = f_0(X_i) - \int_{\mathbb{R}} f_0(x)^2 dx$  and then obtaining convergence in variance from the triangular inequality and continuity of the norm. The proof above is more direct and is provided because it is based on an argument that will be used repeatedly in this paper. The proof is based on a density argument: the result would be direct

if  $f_0$  was continuous and compactly supported, but we did not assume continuity nor compact support; however, functions in  $L^1$  are "approximately" such in the sense that continuous compactly supported functions are dense in  $L^1$ . This is this approximation that is used in proving the result in Section B of the Supplementary Material as well as the mollification theorem and  $L^1$ -continuity.

**Lemma 2.2.**

$$\lim_{n \rightarrow \infty} \binom{n}{2} h_n \text{Var } W_n = \int_{\mathbb{R}} f_0(x)^2 dx \int_{\mathbb{R}} K(u)^2 du.$$

*Proof.* For simplicity, we write  $h_n = h$ . From Equation (1.4), we have

$$\binom{n}{2} \text{Var } W_n = \text{Var}(u_n^2(X_1, X_2)) - 2\text{Var}(u_n^1(X_1)).$$

From the proof of Lemma 2.1, we know that  $\text{Var}(u_n^1(X_1)) = O(1) = o(h^{-1})$  and  $\mathbb{E}[u_n^2(X_1, X_2)]^2 = \mathbb{E}[u_n^1(X_1)]^2 = O(1) = o(h^{-1})$ . It remains to handle  $\mathbb{E}[(u_n^2(X_1, X_2))^2]$ . We have

$$\begin{aligned} h \mathbb{E}[(u_n^2(X_1, X_2))^2] &= h \int_{\mathbb{R}} \int_{\mathbb{R}} (K_h(x-y))^2 f_0(y) dy f_0(x) dx \\ &= h \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{h} K(u)^2 f_0(x-uh) du f_0(x) dx \\ &= \int_{\mathbb{R}} K(u)^2 \int_{\mathbb{R}} f_0(x-uh) f_0(x) dx du \end{aligned}$$

As in Equation (2.1), we can conclude by using the result proved in Section B of the Supplementary Material and dominated convergence that

$$\lim_{h \rightarrow 0} \int_{\mathbb{R}} K(u)^2 \int_{\mathbb{R}} f_0(x-uh) f_0(x) dx du = \int_{\mathbb{R}} f_0(x)^2 dx \int_{\mathbb{R}} K(u)^2 du.$$

□

From this results, it follows directly that  $\text{Var}(\sqrt{n}L_n) = O(1)$  and  $\text{Var}(\sqrt{n}W_n) = O((nh_n)^{-1})$ , so that the dominating terms in the Hoeffding decomposition depends on  $\lim_{n \rightarrow \infty} nh_n$ . To obtain the weak limit, we have to be more precise and bound higher moments of  $L_n$  and  $W_n$ . This is the objective of Lemma 2.3, Lemma 2.4, and Lemma 2.5 which are proved in Section A.1. of the Supplementary Material. The proofs make use of a recurring density argument that is stated and proved in Section B of the Supplementary Material. This argument is similar to the one used to prove Lemma 2.1 and Lemma 2.2.

**Lemma 2.3.** *Let  $i, j, k \in \{1, 2, \dots, n\}$  with  $i \neq j \neq k$ . Let  $q, r \geq 1$  be integers. Then*

1.

$$\mathbb{E}[|u_n^1(X_i)|^q] = O(1);$$

2.

$$\mathbb{E}[|u_n^2(X_i, X_j)|^q] = O(h^{-q+1});$$

3.

$$\mathbb{E} [|u_n^2(X_i, X_j)|^r |u_n^1(X_i)|^q] = O(h^{-r+1/2});$$

4.

$$\mathbb{E} [|u_n^2(X_i, X_j)|^r |u_n^2(X_i, X_k)|^q] = O(h^{-r-q+2});$$

**Lemma 2.4.** Let  $C_2^n$  denote the set of all pairs  $(i, j)$  with  $i < j$ ,  $1 \leq i, j \leq n$ . For  $(i, j) \in C_2^n$ , let  $r_{i,j} \geq 0$  be non-negative integers. For  $k \in \{1, 2, \dots, n\}$ , let  $s_k \geq 0$  be non-negative integers. Suppose  $r_{i,j} \geq 1$  for at least one pair  $(i, j) \in C_2^n$ . Suppose that  $s_k \geq 1$  for at least one index  $k \in \{1, 2, \dots, n\}$ . Consider the product  $\prod_{(i,j) \in C_2^n} (u_n^2(X_i, X_j))^{r_{i,j}}$ . Denote  $l \in \{1, 2, \dots, n\}$  the number of indexes  $i$  such that  $r_{i,j} \neq 0$  for at least one index  $j \in \{1, 2, \dots, i-1, i+1, \dots, n\}$ . Then

5.

$$\mathbb{E} [\prod_{(i,j) \in C_2^n} |u_n^2(X_i, X_j)|^{r_{i,j}}] = O(h^{-\sum_{(i,j)} r_{i,j} + l - 1});$$

6.

$$\mathbb{E} [\prod_{(i,j) \in C_2^n} |u_n^2(X_i, X_j)|^{r_{i,j}} \prod_{k=1}^n |u_n^1(X_k)|^{s_k}] = O(h^{-\sum_{(i,j)} r_{i,j} + \frac{l-1}{2}});$$

By using two nested density arguments, result (3.) and (6.) of last lemmas can probably be improved to  $O(|h|^{-r+1})$  and  $O(h^{-\sum_{(i,j)} r_{i,j} + l - 1})$ , respectively, but the result is not needed to prove the main result of this section and so we only resort to a cruder bound based on Holder's inequality. As shown in next lemma, the price to pay is a cruder and more cumbersome bound for the higher moments of  $W_n$ , which remains nevertheless sufficient for our purpose.

**Lemma 2.5.** Define

$$\begin{aligned} l(X_i) &= u_n^1(X_i) - u_n^0, \\ w(X_i, X_j) &= u_n^2(X_i, X_j) - u_n^1(X_i) - u_n^1(X_j) + u_n^0 \end{aligned}$$

for  $1 \leq i, j \leq n$ ,  $i \neq j$ . Let  $i, j, k \in \{1, 2, \dots, n\}$  with  $i \neq j \neq k$ . Let  $q, r \geq 1$  be integers. Then

1.

$$\mathbb{E} [|l(X_i)|^q] = O(1);$$

2.

$$\mathbb{E} [|w(X_i, X_j)|^q] = O(h^{-q+1});$$

3.

$$\mathbb{E} [|w(X_i, X_j)|^r |w(X_i, X_k)|^q] = O(h^{-r-q+2});$$

4.

$$\mathbb{E} [\prod_{(i,j) \in C_2^n} |w(X_i, X_j)|^{r_{i,j}}] = O(h^{-\sum_{(i,j)} r_{i,j} + \frac{l+1}{2}} \vee h^{-\sum_{(i,j)} r_{i,j} + l - 1});$$

where the  $C_2^n$  and  $l$  are defined as in Lemma 2.4 for  $w$  instead of  $u_n^2$ .

We are now ready to state the main result of this section, a central limit theorem for the vector with elements properly standardized terms  $L_n$  and  $W_n$ , from which the weak limit of

$(\text{Var } U_n)^{-1/2}(U_n - \mathbb{E}[U_n])$  can be directly derived. The proof is relegated to Section A.2. of the Supplementary Material. The idea is to use the previously derived moment bounds to apply the central limit theorem for quadratic forms of [de Jong \(1987\)](#).

**Proposition 2.1.** *If  $n^2 h_n \rightarrow \infty$ , then the terms  $L_n$  and  $W_n$  in the Hoeffding decomposition (1.1) converges jointly in distribution to a bivariate normal distribution*

$$\begin{pmatrix} \sqrt{n} L_n \\ \sqrt{\binom{n}{2} h_n} W_n \end{pmatrix} \rightsquigarrow \mathcal{N} \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_L^2 & 0 \\ 0 & \sigma_W^2 \end{pmatrix} \right]$$

where

$$\begin{aligned} \sigma_L^2 &= \int_{\mathbb{R}} f_0(x)^3 dx - \left( \int_{\mathbb{R}} f_0(x)^2 dx \right)^2, \\ \sigma_W^2 &= \int_{\mathbb{R}} f_0(x)^2 dx \int_{\mathbb{R}} K(u)^2 du. \end{aligned}$$

From the Hoeffding decomposition for  $U_n$ , Lemma 2.1, and Lemma 2.2, we can directly derive the weak limit of  $(\text{Var } U_n)^{-1/2}(U_n - \mathbb{E}[U_n])$  from the previous result. The proof can be found in Section A.2. of the Supplementary Material.

**Corollary 2.2.** *If  $n^2 h_n \rightarrow \infty$ , then*

$$(\text{Var } U_n)^{-1/2}(U_n - \mathbb{E}[U_n]) \rightsquigarrow N(0, 1).$$

We can now make use of the bias result in [Giné and Nickl \(2008a\)](#) (see Section D of the Supplementary Material) to obtain the weak limits of  $(\text{Var } U_n)^{-1/2}(U_n - \theta_0)$ . The proof is relegated to Section A.2. of the Supplementary Material.

**Corollary 2.3.** *1. If  $nh_n \rightarrow \infty$  and  $nh_n^{4s} \rightarrow 0$ , then*

$$\sqrt{n}(U_n - \theta_0) \rightsquigarrow N(0, 4\sigma_L^2).$$

*2. If  $nh_n \rightarrow C \in (0, \infty)$  and  $nh_n^{4s} \rightarrow 0$ , then*

$$\sqrt{n}(U_n - \theta_0) \rightsquigarrow N\left(0, 4\sigma_L^2 + \frac{2}{C}\sigma_W^2\right).$$

*3. If  $n^2 h_n \rightarrow \infty$ ,  $nh_n \rightarrow 0$ , and  $nh_n^{2s+\frac{1}{2}} \rightarrow 0$ , then*

$$\sqrt{\binom{n}{2} h_n}(U_n - \theta_0) \rightsquigarrow N(0, \sigma_W^2).$$

**Remark 2.2.** Note that conditions in (1.) and (2.) can only hold if  $s > 1/4$ . Note that the last two conditions in (3.) have the following relation: if  $s > 1/4$ , then  $nh_n \rightarrow 0$  implies  $nh_n^{2s+\frac{1}{2}} \rightarrow 0$ ; if  $s < 1/4$ , then  $nh_n^{2s+\frac{1}{2}} \rightarrow 0$  implies  $nh_n \rightarrow 0$ . For (2.), it would be of interest to relax the conditions

to  $nh_n^{4s} \rightarrow C' \in (0, \infty)$ , but it is not clear if such a result holds: the bias result of [Giné and Nickl \(2008a\)](#) only gives  $B_n = O(h^{2s})$  (see Section D of the Supplementary Material).

*Remark 2.3.* (About some CLTs for U-statistics with  $n$ -dependent kernel) In terms of weak convergence, the fundamental result we use is the central limit theorem in [de Jong \(1987\)](#) for generalized quadratic forms so as to handle the quadratic term in the Hoeffding decomposition of the U-statistics with  $n$ -dependent kernel. The result of [Eubank and Wang \(1999\)](#) is a direct extension of [de Jong \(1987\)](#) to handle the case where both the linear and the quadratic terms are of the same order. The result in [de Jong \(1987\)](#) is an extension of the result in [Hall \(1984\)](#), which itself bears similarities to [Beran \(1972\)](#) and [Whittle \(1964\)](#). These results are proved using either a martingale CLT or a Lyapunov CLT. The central limit theorem proved in [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#), which makes use of a Lyapunov condition, bears important similarities to these results. It would be interesting to clarify the connections between them, if possible.

*Remark 2.4.* The weak convergence result of Corollary 2.2 does not make use of the smoothness degree  $s > 0$  but only of the integrability assumption  $f_0 \in L^1 \cap L^\infty$ . It is only when the bias enters the scene that smoothness plays a role as stated in Corollary 2.3. This should be contrasted with the results obtained in [Cattaneo, Crump, and Jansson \(2014b\)](#) for a different kernel-based estimator. For their problem, the moments bounds they borrow from [Robinson \(1995\)](#) and [Nishiyama and Robinson \(2000\)](#) make direct use of strong smoothness assumptions. As a consequence, the weak limit results obtained in [Cattaneo, Crump, and Jansson \(2014b\)](#) depend crucially on the smoothness of the nuisance parameter, even for the quantity  $(\text{Var } \hat{\theta}_n)^{-1/2}(\hat{\theta}_n - \mathbb{E}[\hat{\theta}_n])$  using their notation. This is not the case in our problem as given by Corollary 2.2. This has important consequences that are explored in the following sections. It allows us first to equivalently state the "small bandwidth" weak limit as a "low smoothness" weak limit for an optimal bandwidth sequence in the spirit of [Giné and Nickl \(2008a\)](#) and [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#): this is the object of Corollary 2.4 proved in the next subsection. It has also important consequences in terms of non-regular cases for which inference can be considered – see the introduction of Section 3.

## 2.2 Varying smoothness with optimal bandwidth sequence

In this section, we leverage the fact that not only Corollary 2.3 holds but also the more primitive result Corollary 2.2. This result is possible only due to the assumption-lean bounds we derived in the previous section. It allows us to directly derive the weak limits of  $U_n$  under another asymptotic regime that is more traditional in the statistical literature. This asymptotic regime, which is for instance considered in [Giné and Nickl \(2008a\)](#) or [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#), consists in varying the class of the nuisance parameter, here the degree of smoothness  $s$  of the density, while fixing the bandwidth sequence  $h_n$  to the optimal one (in the sense of trading off bias and variance). In this section, we thus reuse the assumptions:

1. [Assumption K](#);
2. [Assumption D\(s\)](#) for  $s > 0$ ;

but we now fix the bandwidth sequence to the optimal one, which leads to the following assumption:

**Assumption OB.** Given [Assumption D\(s\)](#) for  $s > 0$ , the bandwidth sequence  $h_n$  satisfies

$$0 < h_n = Cn^{-\frac{2}{4s+1}}.$$

for some constant  $C > 0$ .

By directly using [Corollary 2.2](#), we can extend the results in [Giné and Nickl \(2008a\)](#) and show that the estimator  $U_n$  remains asymptotically normal even when the density becomes very irregular as captured by  $s \leq 1/4$ . This leads to the results in [Corollary 2.4](#) which is proved in [Section A.2](#) of the Supplementary Material.

**Corollary 2.4.** 1. If  $s > 1/4$ , then

$$\sqrt{n}(U_n - \theta_0) \rightsquigarrow N(0, 4\sigma_L^2).$$

2. If  $s = 1/4$ , then

$$\sqrt{n}(U_n - \mathbb{E}[U_n]) \rightsquigarrow N(0, 4\sigma_L^2 + \frac{2}{C}\sigma_W^2).$$

3. If  $s < 1/4$ , then

$$\sqrt{\binom{n}{2}}h_n(U_n - \mathbb{E}[U_n]) \rightsquigarrow N(0, \sigma_W^2).$$

*Remark 2.5.* Under [Assumption OB](#), the bias is negligible when  $s > 1/4$ , and so the weak limit can be centered at  $\theta_0$  in this case. When  $s \leq 1/4$ , the estimator is not necessarily unbiased, since then we only have  $\sqrt{n}B_n = O(1)$  and  $\sqrt{\binom{n}{2}}h_nB_n = O(1)$  by the bias result in [Giné and Nickl \(2008a\)](#) (see [Section D](#) of the Supplementary Material). The asymptotic negligibility of the bias when  $s \leq 1/4$  can be obtained from [Corollary 2.2](#) by changing the bandwidth sequence in [Assumption OB](#) to a sub-optimal under-smoothed one. This is made explicit and leveraged in next section when inference is considered.

This result is of particular interest for two main reasons. First, it directly extends the weak limit results in [Giné and Nickl \(2008a\)](#) to the non-regular setting  $s \leq 1/4$ . Secondly, it connects the "small bandwidth asymptotics" of [Cattaneo, Crump, and Jansson \(2014b\)](#) to the results in [Robins and van der Vaart \(2006\)](#). The authors of this latter paper used a different method of proof to obtain a similar result as the one above but for a different estimator of the average density based on Fourier series. Our result makes the connection between their result and the results in [Cattaneo, Crump, and Jansson \(2014b\)](#) explicit, hence bridging a gap whose existence was highlighted in [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#). This also provides additional light on the small bandwidth asymptotic regime introduced in [Cattaneo, Crump, and Jansson \(2014b\)](#). It shows in particular that the small bandwidth result of [Corollary 2.2](#) is more fundamental than the "low smoothness" result above.

*Remark 2.6.* In practice, the smoothness parameter  $s$  is unknown. In this case, it may be estimated from the data to construct an (approximately) optimal bandwidth sequence. But this problem is generally difficult. For estimation, there exists a leaner data-driven procedure using an extension of Lepski's method as in Section 4 in [Giné and Nickl \(2008a\)](#) leading to optimal adaptive rates (with the standard logarithmic cost for the lower regularity cases as in [Efromovich and Low \(1996\)](#)). Inference could then be based on this adaptive estimator. However, given the known limits to adaptive inference (see [Low \(1997\)](#)), we should not expect too many positive results for valid inference when using adaptive estimators. A full investigation of this problem is left for future research. As a second best, we assume the smoothness class to be partially known so as to guarantee asymptotic negligibility of the bias and show that it is then possible to perform inference in the resulting non-regular cases. Solutions to this problem are explored in next section.

### 3 Non-regular inference: Gaussian approximation and bootstrap

Given the newly derived weak limits for  $U_n$ , there is hope to extend the range of cases for which valid inferential rules exist. Indeed, asymptotic normality can now be used not only when  $nh_n \rightarrow \infty$  but also when  $nh_n \rightarrow 0$  provided  $n^2h_n \rightarrow \infty$ . For the limit in Corollary 2.2 to be used for inference, however, we first need to ensure asymptotic negligibility of the bias (see Remark 2.6). From Remark 2.5, this is seen to hold under two different sets of conditions compatible with  $n^2h_n \rightarrow \infty$ , namely

**Assumption NB1.** *Assumption D(s) holds with  $s > 1/4$  and the bandwidth sequence satisfies  $nh_n^{4s} \rightarrow 0$ ; or*

**Assumption NB2.** *Assumption D(s) holds with  $s \leq 1/4$  and the bandwidth sequence satisfies  $nh_n^{2s+\frac{1}{2}} \rightarrow 0$ .*

We can then consider using result 3 in Corollary 2.3 for inference under two new sets of non-regular cases corresponding to Assumption NB1 and Assumption NB2, respectively:

- (A) the density is regular enough ( $s > 1/4$ ) and the estimator belongs to a wide class of under-smoothed estimators (with bandwidths from  $nh_n \rightarrow 0$  to  $n^2h_n \rightarrow \infty$ );
- (B) the density is highly irregular ( $s \leq 1/4$ ) and the estimator belongs to a slightly narrower class of under-smoothed estimators (with bandwidths from  $nh_n^{2s+\frac{1}{2}} \rightarrow 0$  to  $n^2h_n \rightarrow \infty$ ).

For these two cases, the result 3 in Corollary 2.3 can be used since  $nh_n \rightarrow 0$  and  $nh_n^{2s+\frac{1}{2}} \rightarrow 0$  hold. Because Assumption NB1 is compatible with results 1 and 2 of Corollary 2.3 (when  $nh_n \rightarrow C \in (0, +\infty]$ ) while Assumption NB2 is not (see Remark 2.5), we should work under Assumption NB1 in the rest of this section. It is important to note, however, that all the results we derive in this section for the non-regular cases (A) under Assumption NB1 (that is, when  $nh_n \rightarrow 0$ ) hold equivalently for the non-regular cases (B) under NB2.

We then show in this section that:

- whenever  $nh_n \rightarrow 0$ , the plug-in variance estimator is not consistent and the non-parametric bootstrap fails, so they cannot be used for inference in the non-regular cases (A) and (B);
- however, it is possible to construct consistent variance estimators so that valid inference can be performed in the non-regular cases (A) and (B) using result 3 of Corollary 2.3; moreover, these estimators remain consistent when the conditions for results 1 or 2 hold, hence allowing some form of robustness under Assumption NB1.

We will start by constructing a simple leave-one-out estimator and showing its consistency in regular and non-regular cases (Proposition 3.1). We will then prove the inconsistency of the plug-in variance estimator in non-regular cases and show that consistency can be restored by appropriate bias-correction (Proposition 3.2 and Proposition 3.3). We will derive the failure of the non-parametric bootstrap under non-regular cases (Proposition 3.4).

These results in this section can then be used to inform practice. Even without perfect knowledge of the regularity of the density, it is possible to hedge one's decision by under-smoothing more than less. In particular, when the density is not supposed to be so regular, it is still possible to perform valid inference under a decently wide range of under-smoothed bandwidth sequences if one uses the variance estimators we construct in this section. However, one has to be careful since not all inferential solutions that commonly hold in regular cases can be used then. These results can also be seen as a formal way to justify under-smoothing as a decently robust practice for inference provided it is combined with robust problem-specific inferential solutions.

*Remark 3.1.* The results in this section are new but they share some fundamental features with the few results previously obtained for inference with U-statistics with  $n$ -dependent kernels. It is interesting to note, first, that [Robins, Li, Tchetgen Tchetgen, and van der Vaart \(2016\)](#) did not consider the problem of inference. As far as we know, only [Härdle and Mammen \(1993\)](#) and [Cattaneo, Crump, and Jansson \(2014b,a\)](#) directly considered the problem and derived qualitatively similar solutions as ours. The novelty of our results come from our capacity to handle non-regular cases emerging not only from very small bandwidth sequences but also from non-regularity of the nuisance parameter. This follows from our assumption-lean moment bounds as highlighted in Remark 2.4. In a sense, the results in [Härdle and Mammen \(1993\)](#) and [Cattaneo, Crump, and Jansson \(2014b,a\)](#) do not have counterparts to Assumption NB2 and have to work under an assumption that bears resemblance to Assumption NB1. Compared to [Cattaneo, Crump, and Jansson \(2014b,a\)](#), by working in the one-dimensional case, we are also able to construct a simpler variance estimator (that could be adapted to their problem).

*Remark 3.2.* From the standpoint of statistical theory, it is interesting to note that U-statistics with  $n$ -dependent kernels lead to inferential results quite different than for U-statistics with non-dependent kernels. This is a direct outcome of the weak results that obtain for such statistics. In particular, normality under the quadratic term dominating, which does not obtain for second-order U-statistics with non-dependent kernel, allows for solutions valid independently of whether the linear term or the quadratic term dominates. We believe this point has not been appreciated enough

in the literature. As far as we know, [Cattaneo, Crump, and Jansson \(2014b,a\)](#) are the first ones to implicitly unearth such a feature and should be credited for it.

*Remark 3.3.* As traditionally done in statistics and econometrics for problems of the type we consider, inference is performed under asymptotic negligibility of the bias. We could relax the conditions for the bias to be of the same order as the standard deviation. In this case, valid inferential solutions should also exist, either by bootstrapping the debiased estimate and properly correcting as in [Calonico, Cattaneo, and Farrell \(2018\)](#) or by pre pivoting as in [Cavaliere, Gonçalves, Nielsen, and Zanelli \(2024\)](#). The characterization of these solutions are left for future research. They are of interest given the paucity of inferential results for U-statistics with  $n$ -dependent kernels. If one wants to completely drop the assumption that the bias is asymptotically negligible, then data-driven methods could theoretically be considered where the choice of  $s$  is informed by the data. However, there exist fundamental limits to the validity of inference based on these methods as highlighted in [Remark 2.6](#). The solutions in this paper work as a second-best in the sense that approximate knowledge of  $s$  combined with under-smoothing delivers valid inferential results if the variance estimator we construct is used, even when the exact unknown smoothness degree satisfies  $s \leq 1/4$ .

### 3.1 A simple consistent leave-one-out variance estimator

The objective of this subsection is to build from scratch the simplest estimator of  $\text{Var } U_n$  that would be consistent across a wide range of bandwidth sequences so that the Gaussian approximation in [Corollary 2.3](#) can be used to perform (asymptotically valid) inference. The range of bandwidth sequences would cover the non-regular cases (A) and (B) where the quadratic term dominates corresponding to high under-smoothing regimes and low smoothness regimes. The idea behind the estimator in this section is similar to the one used to build a consistent variance estimator in [Cattaneo, Crump, and Jansson \(2014b\)](#). The construction and the proof here are more direct as we directly unpack all elements that could be plugged-in, hence leading to a simpler estimator that could be adapted to their problem. The cost of this simpler estimator is a more complicated proof.

Based on [Equation \(1.3\)](#), [Equation \(1.4\)](#), and the results in [Lemma 2.1](#) and [Lemma 2.2](#), it is natural to consider the estimators

$$\begin{aligned} \widetilde{\sigma}_L^2 = & n^{-1} \sum_{i=1}^n \binom{n-1}{2}^{-1} \sum_{\substack{j=1 \\ j \neq i}}^{n-1} \sum_{\substack{k=j+1 \\ k \neq i}}^n k_n(X_i, X_j) k_n(X_i, X_k) \\ & - \left( \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=i+1}^n k_n(X_i, X_j) \right)^2 \end{aligned}$$

and

$$\widetilde{\sigma}_W^2 = h_n \binom{n}{2}^{-1} \sum_{i=1}^n \sum_{j=i+1}^n k_n(X_i, X_j)^2.$$

Then, based on Equation (1.2), it is natural to consider the estimator of  $\text{Var } U_n$  given by

$$\tilde{V}_n = 4n^{-1} \widetilde{\sigma_L^2} + \left(\frac{n}{2}\right)^{-1} h_n^{-1} \widetilde{\sigma_W^2}.$$

We show that this estimator is consistent for the asymptotic variance. Given the characterization of  $\text{Var } U_n$

$$\text{Var } U_n = 4n^{-1} \left( \sigma_L^2 + o(1) \right) + \left(\frac{n}{2}\right)^{-1} h_n^{-1} \left( \sigma_W^2 + o(1) \right),$$

the proof of the consistency of  $\tilde{V}_n$  follows directly from the lemma stated below. The proof of this lemma is long and is relegated to Section A.3. of the Supplementary Material.

**Lemma 3.1.** *If  $n^2 h_n \rightarrow \infty$ , then:*

1.

$$\widetilde{\sigma_L^2} = \sigma_L^2 + o_P(1);$$

2.

$$\widetilde{\sigma_W^2} = \sigma_W^2 + o_P(1).$$

**Proposition 3.1.** *If  $n^2 h_n \rightarrow \infty$ , then*

$$\tilde{V}_n^{-1/2} (U_n - \theta_0) \rightsquigarrow N(0, 1)$$

with

$$\tilde{V}_n = 4n^{-1} \widetilde{\sigma_L^2} + \left(\frac{n}{2}\right)^{-1} h_n^{-1} \widetilde{\sigma_W^2}.$$

*Proof.* This follows directly from Lemma 3.1, the characterization of  $\text{Var } U_n$  in Equation (A.1), and an application of Slutsky's theorem in Corollary 2.3.  $\square$

*Remark 3.4.* The form of  $\widetilde{\sigma_L^2}$  stems from the moment characterization

$$n \text{Var } L_n = \mathbb{E} [u_n^1(X_1)^2] - \mathbb{E} [u_n^1(X_1)]^2.$$

The first term in the estimator  $\widetilde{\sigma_L^2}$  is then a leave-one-out bias-corrected estimator of the plug-in estimator for  $\mathbb{E} [u_n(X_1)^2]$ . Indeed, we have that

$$\begin{aligned} n^{-1} \sum_{i=1}^n \left( (n-1)^{-1} \sum_{\substack{j=1 \\ j \neq i}}^n k_n(X_i, X_j) \right)^2 &= n^{-1} \sum_{i=1}^n (n-1)^{-2} \sum_{\substack{j=1 \\ j \neq i}}^n k_n(X_i, X_j)^2 \\ &\quad + 2n^{-1} \sum_{i=1}^n (n-1)^{-2} \sum_{\substack{j=1 \\ j \neq i}}^{n-1} \sum_{\substack{k=j+1 \\ k \neq i}}^n k_n(X_i, X_j) k_n(X_i, X_k). \end{aligned}$$

The first term is the one contributing to  $\sigma_W^2$  under  $nh_n \rightarrow 0$ , while the second term is the one contributing to  $\sigma_L^2$  under  $nh_n \rightarrow \infty$ . This justifies the form of the first term in  $\widetilde{\sigma_L^2}$ . For

completeness, we show that the first term

$$\widehat{\sigma_{LW}^2} = n^{-1} \sum_{i=1}^n (n-1)^{-2} \sum_{j=1, j \neq i}^n k_n(X_i, X_j)^2$$

is the one contributing to  $\sigma_W^2$  when  $nh_n \rightarrow 0$ . To see this, note that  $\mathbb{E} [n^{-1} \sum_{i=1}^n (n-1)^{-2} \sum_{j=1, j \neq i}^n k_n(X_i, X_j)^2] = (n-1)^{-1} \mathbb{E} [k_n(X_i, X_j)^2]$ . Then

$$\begin{aligned} & \mathbb{E} \left[ \left| nh_n n^{-1} \sum_{i=1}^n (n-1)^{-2} \sum_{j=1, j \neq i}^n k_n(X_i, X_j)^2 - nh_n (n-1)^{-1} \mathbb{E} [k_n(X_i, X_j)^2] \right|^2 \right] \\ &= h_n^2 n^2 (n-1)^{-2} \text{Var} \left( \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)^2 \right) \\ &= O(n^{-2} h_n^{-1} + n^{-1}), \end{aligned}$$

where the last equality follows from the bounds derived in Section C of the Supplementary Material. We conclude by  $L^2$ -convergence that

$$nh_n n^{-1} \sum_{i=1}^n (n-1)^{-2} \sum_{j=1, j \neq i}^n k_n(X_i, X_j)^2 = \sigma_W^2 + o_P(1).$$

### 3.2 Inconsistency of the plug-in variance estimator

We show in this subsection that the plug-in variance estimator is inconsistent whenever asymptotic linearity subsides, but that a simple bias-correction can restore consistency. This is in line with the results in [Hardle and Mammen \(1993\)](#) and [Cattaneo, Crump, and Jansson \(2014a\)](#), completing the results available for plug-in inference for U-statistics with  $n$ -dependent kernels.

Consider then the following plug-in estimators

$$\widehat{\sigma_L^2} = n^{-1} \sum_{i=1}^n \widehat{l}_{n,i}^2, \quad (3.1)$$

$$\widehat{\sigma_W^2} = h_n \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \widehat{w}_{n,i,j}^2, \quad (3.2)$$

with

$$\widehat{u}_n^0 = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j), \quad (3.3)$$

$$\widehat{l}_{n,i} = (n-1)^{-1} \sum_{\substack{j=1 \\ j \neq i}}^n k_n(X_i, X_j) - \widehat{u}_n^0, \quad (3.4)$$

$$\widehat{w}_{n,i,j} = k_n(X_i, X_j) - \widehat{l}_{n,i} - \widehat{l}_{n,j} + \widehat{u}_n^0. \quad (3.5)$$

We first collect in the following lemmas the limits in probability of the estimators  $\widehat{\sigma_L^2}$  and  $\widehat{\sigma_W^2}$  for a whole range of bandwidth sequence rates. These limits are then directly used to show consistency and inconsistency of different (rescaled) plug-in variance estimators. The proofs of these lemmas are relegated to Section A.4. of the Supplementary Material.

**Lemma 3.2.**

1. If  $nh_n \rightarrow \infty$ , then

$$\widehat{\sigma_L^2} = \sigma_L^2 + o_P(1).$$

2. If  $nh_n \rightarrow C \in (0, \infty)$ , then

$$n^{-1}\widehat{\sigma_L^2} = n^{-1}\left(\sigma_L^2 + \frac{2}{C}\sigma_W^2 + o_P(1)\right).$$

3. If  $n^2h_n \rightarrow \infty$  and  $nh_n \rightarrow 0$ , then

$$n^{-1}\widehat{\sigma_L^2} = \left(\frac{n}{2}\right)^{-1} h_n^{-1}\left(\sigma_W^2 + o_P(1)\right)$$

**Lemma 3.3.** If  $n^2h_n \rightarrow \infty$ , then

$$\widehat{\sigma_W^2} = \sigma_W^2 + o_P(1).$$

Recall again the characterization of  $\text{Var } U_n$  given by

$$\text{Var } U_n = 4n^{-1}\left(\sigma_L^2 + o(1)\right) + \left(\frac{n}{2}\right)^{-1} h_n^{-1}\left(\sigma_W^2 + o(1)\right).$$

Then it follows directly from the two previous lemmas that the plug-in estimator

$$\hat{V}_{n,p} = 4n^{-1}\widehat{\sigma_L^2} + \left(\frac{n}{2}\right)^{-1} h_n^{-1}\widehat{\sigma_W^2}$$

is consistent when the linear term dominates but inconsistent otherwise. However, it is possible to construct a rescaled version

$$\hat{V}_{n,r} = 4n^{-1}\widehat{\sigma_L^2} - 3\left(\frac{n}{2}\right)^{-1} h_n^{-1}\widehat{\sigma_W^2}$$

which is directly seen to be consistent in all cases. These results are summarized in the next propositions.

**Proposition 3.2.** If  $nh_n \rightarrow 0$  and  $n^2h_n \rightarrow \infty$ , then

$$\hat{V}_{n,p} - 3\text{Var } U_n = o_P(1)$$

with

$$\hat{V}_{n,p} = 4n^{-1}\widehat{\sigma_L^2} + \left(\frac{n}{2}\right)^{-1} h_n^{-1}\widehat{\sigma_W^2}.$$

*Proof.* This follows directly from Lemma 3.2 and Lemma 3.3 and the characterization of  $\text{Var } U_n$  in Equation (A.1).  $\square$

**Proposition 3.3.** *If  $n^2 h_n \rightarrow \infty$ , then*

$$\hat{V}_{n,r}^{-1/2}(U_n - \theta_0) \rightsquigarrow N(0, 1)$$

with

$$\hat{V}_{n,r} = 4n^{-1} \widehat{\sigma_L^2} - 3 \binom{n}{2}^{-1} h_n^{-1} \widehat{\sigma_W^2}.$$

*Proof.* This follows directly from Lemma 3.2 and Lemma 3.3, the characterization of  $\text{Var } U_n$  in Equation (A.1), and an application of Slutsky's theorem in Corollary 2.3.  $\square$

*Remark 3.5.* There exists another method to restore consistency of the plug-in variance estimator that is also considered in Cattaneo, Crump, and Jansson (2014b,a). It consists in estimating the variance with a bandwidth sequence  $H_n$  converging at a different rate than the bandwidth sequence  $h_n$  used to estimate  $\theta_0$ . The validity of the method follows directly from Lemma 3.2 by taking  $H_n$  in the estimation of  $\widehat{\sigma_L^2}$  such that  $nH_n \rightarrow \infty$ . In this case, the plug-in estimator with double bandwidth sequences is directly seen to be consistent without rescaling. The method is, however, difficult to implement in practice as it requires the choice of two bandwidth sequences.

### 3.3 Inconsistency of the non-parametric bootstrap

We now show that the non-parametric bootstrap fails to reproduce the underlying distribution across the whole range of bandwidth sequences for which Corollary 2.3 holds. Failure happens whenever the linear terms does not dominate asymptotically. A similar result for a different problem was first obtained by Hardle and Mammen (1993). The same logic underlies the result in Cattaneo, Crump, and Jansson (2014a). The underlying reason can already be seen from the previous results where we had to "manually" rescale the variance of the quadratic term to obtain a consistent estimator of  $\text{Var } U_n$ . Sensibly similar issues and solutions were already reported on jackknife estimate of variance for U-statistics – see Efron and Stein (1981).

Let  $\mathcal{X}_n = \{X_1, X_2, \dots, X_n\}$  be an i.i.d. sample. Then take  $\mathcal{X}_n^* = \{X_1^*, X_2^*, \dots, X_n^*\}$  an i.i.d. sample from the empirical distribution  $\mathbb{P}_n$  based on  $\mathcal{X}_n$ . Equivalently,  $\mathcal{X}^*$  can be obtained by uniformly sampling  $n$  times from  $\mathcal{X}_n$  with replacement. Denote by  $P^*, \mathbb{E}^*, \text{Var}^*, \text{cov}^*$ , the probability, expectation, variance, and covariance taken with respect to the empirical distribution conditional on  $\mathcal{X}_n$ . We introduce the bootstrap analogue to the estimator previously introduced

$$U_n^* = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{1}{h_n} K\left(\frac{X_i^* - X_j^*}{h_n}\right) = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i^*, X_j^*).$$

The statistics  $U_n^*$  is the same second-order U-statistics with  $n$ -dependent kernel  $k_n$  as  $U_n$  but computed over the random sample  $\mathcal{X}_n^*$  instead of  $\mathcal{X}_n$ . Note that conditional on  $\mathcal{X}_n$ , the empirical

distribution is a discrete (non-random) distribution, namely, multinomial with uniform weights  $1/n$ . It follows that the statistics  $U_n^*$  admits a Hoeffding decomposition with respect to the empirical distribution conditional on  $\mathcal{X}_n$ . In virtue of Lemma 1.1, we have

$$U_n^* = \mathbb{E}^*[U_n^*] + 2L_n^* + W_n^*$$

with

$$L_n^* = \frac{1}{n} \sum_{i=1}^n \left[ u_n^{1*}(X_i^*) - u_n^{0*} \right]$$

$$W_n^* = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left[ u_n^{2*}(X_i^*, X_j^*) - u_n^{1*}(X_i^*) - u_n^{1*}(X_j^*) + u_n^{0*} \right]$$

and

$$u_n^{0*} = \mathbb{E}^*[U_n^*]$$

$$u_n^{1*}(X_i^*) = \mathbb{E}^*[k_n(X_i^*, X_j^*) | X_i^*]$$

$$u_n^{2*}(X_i^*, X_j^*) = k_n(X_i^*, X_j^*)$$

where  $i \neq j$  are any two indexes.

By Lemma 1.2, we have

$$\text{Var}^* U_n^* = \frac{4}{n} \text{Var}^*(u_n^{1*}(X_1)) + \frac{2}{n(n-1)} \text{Var}^*(u_n^{2*}(X_1, X_2) - u_n^{1*}(X_1) - u_n^{1*}(X_2)).$$

Moreover, for the same reason as in the proof of Lemma 1.2, we also have

$$\text{Var}^* U_n^* = \frac{4}{n} \text{Var}^*(u_n^{1*}(X_1)) + \frac{2}{n(n-1)} \left[ \text{Var}^*(u_n^{2*}(X_1, X_2)) - \text{Var}^*(u_n^{1*}(X_1)) \right].$$

Then to compute  $\text{Var}^* U_n^*$ , we make use of the multinomial representation of the empirical measure conditional on the observed sample. Similar computations have been used repeatedly when bootstrapping U-statistics (be they with standard kernels or  $n$ -dependent kernels), see, for instance, [Dehling and Mikosch \(1994\)](#) or [Cattaneo, Crump, and Jansson \(2014a\)](#). In particular, note that  $u_n^{1*}(X_i^*)$  can be rewritten as

$$u_n^{1*}(X_i^*) = \mathbb{E}_{\Xi}[k_n(\xi_i(X_1, \dots, X_n), \xi_j(X_1, \dots, X_n)) | \xi_i, X_1, X_2, \dots, X_n]$$

where  $\Xi$  is the multimomial distribution with uniform weights and  $\xi_1, \xi_2, \dots, \xi_n$  is an i.i.d. sample

from this distribution, and so it follows that

$$u_n^{1*}(X_i^*) = \frac{1}{n} \sum_{j=1}^n k_n(X_i^*, X_j).$$

Then we have

$$u_n^{0*} = \mathbb{E}^*[u_n^{1*}(X_i^*)] = \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n k_n(X_i, X_j) = \frac{n-1}{n} U_n.$$

Additional moment calculations then lead to the following result, which yields inconsistency of the bootstrap variance whenever linearity subsides. The proof can be found in Section A.5. of the Supplementary material.

**Proposition 3.4.** *If  $n^2 h_n \rightarrow \infty$ , then*

$$\text{Var}^* U_n^* - 4n^{-1} \sigma_L^2 - 3 \binom{n}{2}^{-1} h_n^{-1} \sigma_W^2 = o_P(1).$$

*In particular, if  $nh_n \rightarrow 0$ , then*

$$\text{Var}^* U_n^* - 3 \text{Var} U_n = o_P(1).$$

This proves the inconsistency of the bootstrap variance. However, inconsistency of the bootstrap variance is not generally sufficient for inconsistency of the bootstrap distribution. To see that it holds in this case, suppose by contradiction that  $\binom{n}{2}^{1/2} h_n^{1/2} (U_n^* - \frac{n-1}{n} U_n) \rightsquigarrow N(0, \sigma_W^2)$  in probability. Then the fact that  $\text{Var}^* U_n^* - 4n^{-1} \sigma_L^2 - 3 \binom{n}{2}^{-1} h_n^{-1} \sigma_W^2 = o_P(1)$  is enough to ensure uniform integrability and convergence of second moments, as in Lemma 2.1. in [Kato \(2011\)](#) which extends Theorem 4.5.2. in [Chung \(2001\)](#) to conditional distributions. Then  $\lim_{n \rightarrow \infty} \binom{n}{2} h_n \text{Var}^* U_n^* = \sigma_W^2$  in probability, a contradiction.

*Remark 3.6.* Given the nature of bootstrap failure in this problem, there is a number of natural potential candidates to restore consistency. The first ones are those in [Cattaneo, Crump, and Jansson \(2014a\)](#), respectively subsampling and bootstrapping the studentized statistics for the consistent variance estimator. A second set of solutions is based on recentering the kernel of the U-statistics in the bootstrap world, or equivalently adjusting the random sampling weights (see [Arcones and Gine \(1992\)](#) and [Dehling and Mikosch \(1994\)](#)). Other reweighing solutions based on a martingale representation of the estimator in the spirit of [Otsu and Rai \(2017\)](#), extending the wild bootstrap in [Hardle and Mammen \(1993\)](#), can also be investigated. A last set of candidates is based on the smoothed bootstrap where resampling is not based on a (conditional) discrete distribution, but a continuous one. Assessing the validity of and comparing these bootstrap methods for the average density estimator considered here form an interesting problem that is left for future research.

## 4 Conclusion

Estimating the expected value of a density is a foundational statistical problem: it is not only a recurrently studied example due to its inherent simplicity among semi-parametric problems but also a fundamental object due to its intrinsic connection to entropy estimation. The derivation of valid inferential results for the average density beyond regular cases remains an important challenge. We provide new solutions to this problem by first showing that the optimal kernel-based plug-in estimator considered in [Giné and Nickl \(2008a\)](#) remains asymptotically normal both when the bandwidth sequence converges to zero very rapidly and when the density is highly irregular. This happens, however, at a cost, namely slower-than-parametric (but minimax-optimal) rates of convergence and a convoluted asymptotic variance that depends on the density estimator. Because of that, many standard inferential rules break down in spite of asymptotic normality: both the plug-in principle for variance estimation and the non-parametric bootstrap fail. We show that the problem can still be solved by fully leveraging the representation of the estimation as a U-statistics with  $n$ -dependent kernel. We borrow techniques from and confirm features of previously studied problems with similar structures – notably [Hardle and Mammen \(1993\)](#) and [Cattaneo, Crump, and Jansson \(2014b,a\)](#). However, as we are able to work under much weaker regularity assumptions, we can highlight new important features. We first connect the "small bandwidth asymptotics" introduced in [Cattaneo, Crump, and Jansson \(2014b\)](#) for kernel-based estimators with the "low smoothness asymptotics" traditionally considered in the literature. We then provide formal guarantees that under-smoothing can produce robust rules even when the nuisance parameter is very irregular. However, all these results depend fundamentally on the structure of the problem at hand. Because average density estimation can be viewed as one of the simplest semi-parametric problems and the estimator is optimal (in the sense of being semi-parametric efficient in regular cases and minimax rate-optimal in non-regular cases), our results (both positive and negative) show the inherent difficulty of semi-parametric problems beyond regular cases and the high-level results available in [Newey \(1994\)](#), [Andrews \(1994\)](#), or [Chen, Linton, and Van Keilegom \(2003\)](#).

## References

- ANDREWS, D. W. (1994): "Asymptotics for semiparametric econometric models via stochastic equicontinuity," *Econometrica*, pp. 43–72.
- ARCONES, M. A., AND E. GINÉ (1992): "On the bootstrap of U and V statistics," *The Annals of Statistics*, pp. 655–674.
- BERAN, R. (1972): "Rank spectral processes and tests for serial dependence," *The Annals of Mathematical Statistics*, pp. 1749–1766.
- BICKEL, P. J., AND Y. RITOV (1988): "Estimating integrated squared density derivatives: sharp best order of convergence estimates," *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 381–393.
- CALONICO, S., M. D. CATTANEO, AND M. H. FARRELL (2018): "On the effect of bias estimation on

- coverage accuracy in nonparametric inference,” *Journal of the American Statistical Association*, 113(522), 767–779.
- CATTANEO, M. D., R. K. CRUMP, AND M. JANSSON (2014a): “Bootstrapping density-weighted average derivatives,” *Econometric Theory*, 30(6), 1135–1164.
- (2014b): “Small bandwidth asymptotics for density-weighted average derivatives,” *Econometric Theory*, 30(1), 176–200.
- CATTANEO, M. D., M. H. FARRELL, M. JANSSON, AND R. P. MASINI (2024): “Higher-Order Refinements of Small Bandwidth Asymptotics for Density-Weighted Average Derivative Estimators,” *Journal of Econometrics*, p. 105855.
- CATTANEO, M. D., AND M. JANSSON (2018): “Kernel-based semiparametric estimators: Small bandwidth asymptotics and bootstrap consistency,” *Econometrica*, 86(3), 955–995.
- (2022): “Average density estimators: Efficiency and bootstrap consistency,” *Econometric Theory*, 38(6), 1140–1174.
- CAVALIERE, G., S. GONÇALVES, M. Ø. NIELSEN, AND E. ZANELLI (2024): “Bootstrap inference in the presence of bias,” *Journal of the American Statistical Association*, pp. 1–11.
- CHEN, X., O. LINTON, AND I. VAN KEILEGOM (2003): “Estimation of semiparametric models when the criterion function is not smooth,” *Econometrica*, 71(5), 1591–1608.
- CHENG, G., AND J. Z. HUANG (2010): “Bootstrap Consistency for General Semiparametric M-Estimation,” .
- CHERNOZHUKOV, V., D. CHETVERIKOV, M. DEMIRER, E. DUFLO, C. HANSEN, W. NEWEY, AND J. ROBINS (2018): “Double/Debiased Machine Learning for Treatment and Structural Parameters,” *The Econometrics Journal*, 21(1), C1–C68.
- CHUNG, K. (2001): *A Course in Probability Theory*. Elsevier Science.
- DE JONG, P. (1987): “A central limit theorem for generalized quadratic forms,” *Probability Theory and Related Fields*, 75(2), 261–277.
- DEHLING, H., AND T. MIKOSCH (1994): “Random quadratic forms and the bootstrap for U-statistics,” *Journal of Multivariate Analysis*, 51(2), 392–413.
- DELSOL, L., AND I. VAN KEILEGOM (2020): “Semiparametric M-estimation with Non-Smooth Criterion Functions,” *Annals of the Institute of Statistical Mathematics*, 72(2), 577–605.
- EFROMOVICH, S., AND M. LOW (1996): “On optimal adaptive estimation of a quadratic functional,” *The Annals of Statistics*, 24(3), 1106–1125.
- EFRON, B., AND C. STEIN (1981): “The jackknife estimate of variance,” *The Annals of Statistics*, pp. 586–596.
- EUBANK, R., AND S. WANG (1999): “A central limit theorem for the sum of generalized linear and quadratic forms,” *Statistics: A Journal of Theoretical and Applied Statistics*, 33(1), 85–91.
- FOLLAND, G. B. (1999): *Real analysis: modern techniques and their applications*, vol. 40. John Wiley & Sons.

- GINÉ, E., AND R. NICKL (2008a): “A Simple Adaptive Estimator of the Integrated Square of a Density,” *Bernoulli*, pp. 47–61.
- (2008b): “Uniform central limit theorems for kernel density estimators,” *Probability Theory and Related Fields*, 141(3), 333–387.
- HALL, P. (1984): “Central limit theorem for integrated square error of multivariate nonparametric density estimators,” *Journal of multivariate analysis*, 14(1), 1–16.
- HALL, P., AND J. S. MARRON (1987): “Estimation of integrated squared density derivatives,” *Statistics & Probability Letters*, 6(2), 109–115.
- HARDLE, W., AND E. MAMMEN (1993): “Comparing nonparametric versus parametric regression fits,” *The Annals of Statistics*, pp. 1926–1947.
- HOEFFDING, W. (1948): “A Class of Statistics with Asymptotically Normal Distribution,” *The Annals of Mathematical Statistics*, pp. 293–325.
- KATO, K. (2011): “A note on moment convergence of bootstrap M-estimators,” *Statistics & Decisions*, 28(1), 51–61.
- KOSOROK, M. R. (2008): *Introduction to Empirical Processes and Semiparametric Inference*, vol. 61. Springer.
- LAURENT, B. (1996): “Efficient estimation of integral functionals of a density,” *The Annals of Statistics*, 24(2), 659–681.
- (2005): “Adaptive estimation of a quadratic functional of a density by model selection,” *ESAIM: Probability and Statistics*, 9, 1–18.
- LEE, A. J. (1990): *U-statistics: Theory and Practice*. CRC Press.
- LOW, M. G. (1997): “On nonparametric confidence intervals,” *The Annals of Statistics*, 25(6), 2547–2554.
- NEWBY, W. K. (1994): “The asymptotic variance of semiparametric estimators,” *Econometrica*, pp. 1349–1382.
- NISHIYAMA, Y., AND P. M. ROBINSON (2000): “Edgeworth expansions for semiparametric averaged derivatives,” *Econometrica*, 68(4), 931–979.
- OTSU, T., AND Y. RAI (2017): “Bootstrap inference of matching estimators for average treatment effects,” *Journal of the American Statistical Association*, 112(520), 1720–1732.
- ROBINS, J., AND A. VAN DER VAART (2006): “Adaptive Nonparametric Confidence Sets,” *The Annals of Statistics*, pp. 229–253.
- ROBINS, J. M., L. LI, E. TCHETGEN TCHETGEN, AND A. VAN DER VAART (2016): “Asymptotic normality of quadratic estimators,” *Stochastic processes and their applications*, 126(12), 3733–3759.
- ROBINSON, P. M. (1995): “The normal approximation for semiparametric averaged derivatives,” *Econometrica*, pp. 667–680.
- TSYBAKOV, A. B. (2009): *Introduction to Nonparametric Estimation*. Springer.
- WHITTLE, P. (1964): “On the convergence to normality of quadratic forms in independent variables,” *Theory of Probability & Its Applications*, 9(1), 103–108.

# Supplement to "Average Density: Weak Limits and Inference in Non-Regular Semi-Parametric Problems"

PAUL DELATTE

The Supplementary Material to "Average Density: Weak Limits and Inference in Non-Regular Semi-Parametric Problems" contains four sections. The first section (A) contains most of the proofs of the results found in the main text. The second section (B) states and proves a result based on a density argument extending  $L^1$ -continuity. The third section (C) contains additional bounds on the moments of some U-statistics that appear in Section A. The last section (D) reproduces and comments on the bias result in [Giné and Nickl \(2008a\)](#).

## A Proofs

### A.1 Proof of the moment bounds of $U_n$

*Proof of Lemma 2.3.* 1. By change of variable, we have

$$\mathbb{E} [|u_n^1(X_i)|^q] = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} K_h(u) f_0(x-u) du \right|^q f_0(x) dx.$$

Since  $f \in L^q$ , we conclude using the mollification theorem and continuity of the norm.

2. By change of variable and Fubini's theorem, we have

$$\mathbb{E} [|u_n^2(X_i, X_j)|^q] = \frac{1}{h^{q-1}} \int_{\mathbb{R}} |K(u)|^q \int_{\mathbb{R}} f_0(x-uh) f_0(x) dx du,$$

and since  $K \in L^q$ , we conclude by using Lemma B.1 and dominated convergence.

3. By Holder's inequality, we have

$$\mathbb{E} [|u_n^2(X_i, X_j)|^r |u_n^1(X_i)|^q] \leq (\mathbb{E} [|u_n^2(X_i, X_j)|^{2r}])^{1/2} (\mathbb{E} [|u_n^1(X_i)|^{2q}])^{1/2}$$

and we conclude directly by using (1.) and (2.).

4. By change of variable and Fubini's theorem, we have

$$\begin{aligned} \mathbb{E} [|u_n^2(X_i, X_j)|^r |u_n^2(X_i, X_k)|^q] &= \int_{\mathbb{R}} \int_{\mathbb{R}} |K_h(x-y)|^r f_0(y) dy \int_{\mathbb{R}} |K_h(x-z)|^q f_0(z) dz f_0(x) dx \\ &= \frac{1}{h^{r+q-2}} \int_{\mathbb{R}} \int_{\mathbb{R}} |K(u)|^r |K(v)|^q \int_{\mathbb{R}} f_0(x-uh) f_0(x-wh) f_0(x) dx du dv \end{aligned}$$

and we conclude by using the extension of Lemma B.1 and dominated convergence.  $\square$

*Proof of Lemma 2.4.* 5. It suffices to show that

$$\mathbb{E} [|u_n^2(X_i, X_j)|^r |u_n^2(X_i, X_k)|^q |u_n^2(X_j, X_k)|^t] = O(h^{-r-q-t+2}).$$

The result then obtains by induction on  $l$ , using either independence, result (4.), or this result. By using the same change of variable as in (4.), we directly obtain that

$$\begin{aligned} & \mathbb{E} [|u_n^2(X_i, X_j)|^r |u_n^2(X_i, X_k)|^q |u_n^2(X_j, X_k)|^t] \\ &= \frac{1}{h^{r+q-2}} \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{h^t} |K(u-v)|^t |K(u)|^r |K(v)|^q \int_{\mathbb{R}} f_0(x-uh) f_0(x-wh) f_0(x) dx du dv \\ &\leq \frac{\|K\|_{\infty}}{h^{r+q+t-2}} \int_{\mathbb{R}} \int_{\mathbb{R}} |K(u)|^r |K(v)|^q \int_{\mathbb{R}} f_0(x-uh) f_0(x-wh) f_0(x) dx du dv \end{aligned}$$

and we can conclude as in (4.).

6. This follows directly from Holder's inequality, independence, (5.) and (1.) as in (3.).  $\square$

*Proof of Lemma 2.5.* 1. We have  $l(X_i) = u_n^1(X_i) - u_n^0 = u_n^1(X_i) - \mathbb{E}[u_n^1(X_i)]$ , hence  $\mathbb{E}[|l(X_i)|^q] \leq C(q) \mathbb{E}[|u_n^1(X_i)|^q]$ , and we conclude by Lemma 2.3.

2. If  $q = 1$ , the result follows from the triangle inequality and Lemma 2.2. Suppose now  $q \geq 2$ . By the multinomial theorem and Lemma 2.3, the term that dominates asymptotically is  $\mathbb{E}[|u_n^2(X_i, X_j)|^q] = O(h^{-q+1})$ , since  $\mathbb{E}[|u_n^2(X_i, X_j)|^{q-1} |u_n^1(X_i)|] = O(h^{-q+3/2})$  and all other terms are of lower order. This concludes the proof.

3. If  $r = q = 1$ , the result follows from the triangle inequality and Lemma 2.3. Suppose now w.l.o.g. that  $r > 1$ . By the multinomial theorem and Lemma 2.4 with  $l = 3$ , the only two terms that can dominate asymptotically are  $\mathbb{E}[|u_n^2(X_i, X_j)|^r |u_n^2(X_i, X_k)|^q] = O(h^{-r-q+2})$  and  $\mathbb{E}[|u_n^2(X_i, X_j)|^{r-1} |u_n^2(X_i, X_k)|^q |u_n^1(X_i)|] = O(h^{-r+1-q+\frac{3-1}{2}}) = O(h^{-r-q+2})$ , since all the other terms are of lower order. This concludes the proof.

4. The same argument generalizes by induction on  $l$ , as there are always only two terms in the multinomial expansions that can dominate asymptotically.  $\square$

## A.2 Proof of the weak limits

*Proof of Proposition 2.1.* For simplicity, we write  $h = h_n$ . To directly apply the results of [de Jong \(1987\)](#) and [Eubank and Wang \(1999\)](#), we recall and introduce some notations

$$\begin{aligned} l(X_i) &= u_n^1(X_i) - u_n^0, \\ w(X_i, X_j) &= u_n^2(X_i, X_j) - u_n^1(X_i) - u_n^1(X_j) + u_n^0, \end{aligned}$$

and

$$L_i = n^{-1/2} l(X_i) \quad \text{and} \quad W_{i,j} = \binom{n}{2}^{-1/2} h^{1/2} w(X_i, X_j)$$

for  $1 \leq i, j \leq n, i \neq j$ . From there, it follows that

$$\sqrt{n}L_n = \sum_{i=1}^n L_i =: L(n) \quad \text{and} \quad \sqrt{\binom{n}{2}}hW_n = \sum_{i=1}^{n-1} \sum_{j=i+1}^n W_{i,j} =: W(n).$$

In particular,  $\text{Var}(\sqrt{n}L_n) = O(1)$  and  $\text{Var}(\sqrt{\binom{n}{2}}hW_n) = O(1)$ , so

$$\text{Var}\left(\sqrt{n}L_n + \sqrt{\binom{n}{2}}hW_n\right) = O(1),$$

and so, in all the Lyapunov-type conditions, the normalizing variances can be taken to be 1. The conditions (1.3) to (1.6) in [Eubank and Wang \(1999\)](#) then rewrite as

$$\binom{n}{2}^{-1} h \max_{1 \leq i \leq n} \sum_{j=1}^n \text{Var}(w(X_i, X_j)) \rightarrow 0, \quad (\text{EW1.3})$$

$$\mathbb{E}[W(n)^4]/(\text{Var } W(n))^2 \rightarrow 3, \quad (\text{EW1.4})$$

$$n^{-2} \sum_{i=1}^n \mathbb{E}[l(X_i)^4] \rightarrow 0, \quad (\text{EW1.5})$$

$$\binom{n}{2}^{-1} n^{-1} h \mathbb{E}\left[\left(\sum_{i=2}^n \sum_{j=1}^{i-1} \mathbb{E}[w(X_i, X_j)l(X_i)|X_1, \dots, X_{i-1}]\right)^2\right] \rightarrow 0. \quad (\text{EW1.6})$$

The i.i.d. assumption allows us to considerably simplify those expressions. In particular, the conditions above are equivalent to

$$n^{-1} h \text{Var}(w(X_1, X_2)) \rightarrow 0, \quad (\text{EW1.3bis})$$

$$\mathbb{E}[W(n)^4]/(\text{Var } W(n))^2 \rightarrow 3, \quad (\text{EW1.4})$$

$$n^{-1} \mathbb{E}[l(X_1)^4] \rightarrow 0, \quad (\text{EW1.5bis})$$

$$n^{-1} h \text{Var}(\mathbb{E}[w(X_2, X_1)l(X_2)|X_1]) \rightarrow 0. \quad (\text{EW1.6bis})$$

The last equivalence follows from  $\mathbb{E}[X^2] = \text{Var}(X) + \mathbb{E}[X]^2$  and  $\mathbb{E}[w(X_i, X_j)l(X_i)] = 0$ . To handle (EW1.4), we make use of the expansion of  $\mathbb{E}[W(n)^4]$  in Table 1 in [de Jong \(1987\)](#). In particular, using the notations of [de Jong \(1987\)](#), it follows from our normalization that (EW1.4) holds whenever the terms  $G_I, G_{II}, G_{III}, G_{IV}$  tend to zero and the term  $G_V$  is asymptotically equivalent to  $(\text{Var}(W(n))^2)/2$ . Using the i.i.d. assumption and our notations, this reduces to the following

conditions

$$n^{-2}h^2\mathbb{E}[w(X_1, X_2)^4] \rightarrow 0 \quad (\text{dJ}.G_I)$$

$$n^{-1}h^2\mathbb{E}[w(X_1, X_2)^2w(X_1, X_3)^2] \rightarrow 0 \quad (\text{dJ}.G_{II})$$

$$n^{-1}h^2\mathbb{E}[w(X_1, X_2)^2w(X_1, X_3)w(X_3, X_2)] \rightarrow 0 \quad (\text{dJ}.G_{III})$$

$$h^2\mathbb{E}[w(X_1, X_2)w(X_1, X_3)w(X_4, X_2)w(X_4, X_3)] \rightarrow 0 \quad (\text{dJ}.G_{IV})$$

$$h\mathbb{E}[w(X_1, X_2)^2]/\text{Var}(W(n)) \rightarrow 1 \quad (\text{dJ}.G_V)$$

where the last equivalence follows from  $3\binom{n}{4}\binom{n}{2}^{-2} \sim 1/2$ . We now use Lemma 2.5 to prove that all limits are as given. For (EW1.3bis), we have

$$\text{Var } w(X_1, X_2) \leq E[w(X_1, X_2)^2] = O(h^{-1}),$$

hence the result. For (dJ.G<sub>I</sub>), we have

$$E[w(X_1, X_2)^4] = O(h^{-3}),$$

and so the result follows since  $n^2h \rightarrow \infty$ . For (dJ.G<sub>II</sub>),

$$\mathbb{E}[w(X_1, X_2)^2w(X_1, X_3)^2] = O(h^{-2}),$$

so the result follows. For (dJ.G<sub>III</sub>), we have

$$\mathbb{E}[w(X_1, X_2)^2w(X_1, X_3)w(X_3, X_2)] = O(h^{-2}),$$

so the result follows. For (dJ.G<sub>IV</sub>), we have

$$\mathbb{E}[w(X_1, X_2)w(X_1, X_3)w(X_4, X_2)w(X_4, X_3)] = O(h^{-3/2}),$$

so the result follows since  $h \rightarrow 0$ . For (dJ.G<sub>V</sub>), the result follows immediately from Lemma 2.2 and Lemma 2.5. For (EW1.5bis), we have

$$\mathbb{E}[l(X_1)^4] = O(1),$$

so the result follows. For (EW1.6bis), we have

$$\text{Var}(\mathbb{E}[w(X_2, X_1)l(X_2)|X_1]) \leq \mathbb{E}((\mathbb{E}[w(X_2, X_1)l(X_2)|X_1])^2).$$

By monotonicity, conditional Holder's inequality, and independence,

$$\text{Var}(\mathbb{E}[w(X_2, X_1)l(X_2)|X_1]) \leq \mathbb{E}[w(X_2, X_1)](\mathbb{E}[l(X_2^2)])^{1/2} = O(1),$$

and so the result follows. This concludes the proof.  $\square$

*Proof of Corollary 2.2.* From Lemma 2.1 and Lemma 2.2 and Equation (1.2), we have

$$\text{Var } U_n = 4n^{-1}(\sigma_L^2 + o(1)) + \left(\frac{n}{2}\right)^{-1} h_n^{-1}(\sigma_W^2 + o(1)). \quad (\text{A.1})$$

In particular,  $(\text{Var } U_n)^{-1/2} = O(n^{1/2} \wedge nh_n^{1/2})$ . By distinguishing three cases if necessary, the result then follows immediately from Equation (1.1), Slutsky's theorem, Proposition 2.1, and the normality of the marginals of bivariate normals.  $\square$

*Proof of Corollary 2.3.* 1. If  $nh_n \rightarrow \infty$ , then  $(\text{Var } U_n)^{-1/2} \sim (4\sigma_L^2)^{-1/2} n^{1/2}$ . By Lemma D.1,  $n^{1/2}(\mathbb{E}[U_n] - \theta_0) = O(n^{1/2}h^{2s})$  and so  $n^{1/2}(\mathbb{E}[U_n] - \theta_0) = o(1)$  since  $nh_n^{4s} \rightarrow 0$ . The result then follows from Slutsky's theorem and Corollary 2.2.

2. If  $nh_n \rightarrow C \in (0, \infty)$ , then  $(\text{Var } U_n)^{-1/2} \sim (4\sigma_L^2 + \frac{2}{C}\sigma_W^2)^{-1/2} n^{1/2}$ . As in (1.), since  $nh_n^{4s} \rightarrow 0$ ,  $n^{1/2}(\mathbb{E}[U_n] - \theta_0) = o(1)$ . The result follows again from Slutsky's theorem and Corollary 2.2.

3. If  $nh_n \rightarrow 0$ , then  $(\text{Var } U_n)^{-1/2} \sim (\sigma_W^2)^{-1/2} h^{1/2} \left(\frac{n}{2}\right)^{1/2}$ . By Lemma D.1,  $h^{1/2} \left(\frac{n}{2}\right)^{1/2}(\mathbb{E}[U_n] - \theta_0) = O(nh^{2s+1/2})$  and so  $n^{1/2}(\mathbb{E}[U_n] - \theta_0) = o(1)$  since  $nh^{2s+1/2} \rightarrow 0$ . The result follows again from Slutsky's theorem and Corollary 2.2.  $\square$

*Proof of Corollary 2.4.* 1. If  $s > 1/4$ , then  $nh_n = Cn^{\frac{4s-1}{4s+1}} \rightarrow \infty$  and  $nh^{4s} = Cn^{\frac{-4s+1}{4s+1}} \rightarrow 0$ .

2. If  $s = 1/4$ , then  $nh_n \rightarrow C \in (0, \infty)$ .

3. If  $s < 1/4$ , then  $nh_n = Cn^{\frac{4s-1}{4s+1}} \rightarrow 0$ .

The results then follow directly from Corollary 2.2 as in the proof of Corollary 2.3.  $\square$

### A.3 Proofs of the consistency of the simple variance estimator

*Proof of Lemma 3.1.* 1. We start with  $\left(\frac{n}{2}\right)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)$ . It is seen from an i.i.d. argument that

$$\mathbb{E} \left[ \left(\frac{n}{2}\right)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j) \right] = \mathbb{E} [k_n(X_1, X_2)].$$

Then, again by i.i.d. and Lemma 2.3,

$$\begin{aligned} \mathbb{E} \left[ \left| \left(\frac{n}{2}\right)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j) - \mathbb{E} [k_n(X_1, X_2)] \right|^2 \right] &= \text{Var} \left( \left(\frac{n}{2}\right)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j) \right) \\ &= \text{Var } U_n = O(n^{-1} + n^{-2}h^{-1}). \end{aligned}$$

By  $L^2$ -convergence, it follows that

$$\left(\frac{n}{2}\right)^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j) = \int_{\mathbb{R}} f_0(x) dx + o_P(1).$$

We now consider  $n^{-1} \sum_{i=1}^n \binom{n-1}{2}^{-1} \sum_{j=1, j \neq i}^{n-1} \sum_{k=j+1, k \neq i}^n k_n(X_i, X_j) k_n(X_i, X_k)$ . It is seen from an i.i.d. argument that

$$\mathbb{E} \left[ n^{-1} \sum_{i=1}^n \binom{n-1}{2}^{-1} \sum_{\substack{j=1, \\ j \neq i}}^{n-1} \sum_{\substack{k=j+1, \\ k \neq i}}^n k_n(X_i, X_j) k_n(X_i, X_k) \right] = \mathbb{E} [k_n(X_1, X_2) k_n(X_1, X_3)].$$

Note then by i.i.d. and the law of iterated expectation that

$$\begin{aligned} \mathbb{E} [\mathbb{E} [k_n(X_1, X_2) | X_1]^2] &= \mathbb{E} [\mathbb{E} [k_n(X_1, X_2) | X_1] \mathbb{E} [k_n(X_1, X_3) | X_1]] \\ &= \mathbb{E} [\mathbb{E} [k_n(X_1, X_2) k_n(X_1, X_3) | X_1]] \\ &= \mathbb{E} [k_n(X_1, X_2) k_n(X_1, X_3)]. \end{aligned}$$

Then

$$\begin{aligned} &\mathbb{E} \left[ \left| n^{-1} \sum_{i=1}^n \binom{n-1}{2}^{-1} \sum_{\substack{j=1, \\ j \neq i}}^{n-1} \sum_{\substack{k=j+1, \\ k \neq i}}^n k_n(X_i, X_j) k_n(X_i, X_k) - \mathbb{E} [\mathbb{E} [k_n(X_1, X_2) | X_1]^2] \right|^2 \right] \\ &= \text{Var} \left( n^{-1} \sum_{i=1}^n \binom{n-1}{2}^{-1} \sum_{j=1, j \neq i}^{n-1} \sum_{k=j+1, k \neq i}^n k_n(X_i, X_j) k_n(X_i, X_k) \right) \\ &= n^{-1} \text{Var} \left( \binom{n-1}{2}^{-1} \sum_{j=2}^{n-1} \sum_{k=j+1}^n k_n(X_1, X_j) k_n(X_1, X_k) \right) \\ &\quad + n^{-2} n(n-1) \binom{n-1}{2}^{-2} \text{cov} \left( \sum_{i=2}^{n-1} \sum_{j=i+1}^n k_n(X_1, X_i) k_n(X_1, X_j), \sum_{\substack{k=1 \\ k \neq 2}}^{n-1} \sum_{\substack{l=k+1 \\ l \neq 2}}^n k_n(X_2, X_k) k_n(X_2, X_l) \right) \\ &= n^{-1} R_1 + 2n^{-2} \binom{n-1}{2}^{-1} R_2, \end{aligned}$$

with

$$R_1 = \text{Var} \left( \binom{n-1}{2}^{-1} \sum_{j=2}^{n-1} \sum_{k=j+1}^n k_n(X_1, X_j) k_n(X_1, X_k) \right),$$

and

$$R_2 = \text{cov} \left( \sum_{i=2}^{n-1} \sum_{j=i+1}^n k_n(X_1, X_i) k_n(X_1, X_j), \sum_{\substack{k=1 \\ k \neq 2}}^{n-1} \sum_{\substack{l=k+1 \\ l \neq 2}}^n k_n(X_2, X_k) k_n(X_2, X_l) \right),$$

where the penultimate equality follows from expanding the variance around the sum and identical distributions across  $i$ . In Lemma C.3, it is shown that

$$R_2 = O(n^3 + n^2 h_n^{-1} + n h_n^{-2}).$$

For  $R_1$ , note that by the law of total variance, we have

$$R_1 = A + B$$

where

$$A = \mathbb{E} \left[ \text{Var} \left( \binom{n-1}{2}^{-1} \sum_{j=2}^{n-1} \sum_{k=j+1}^n k_n(X_1, X_j) k_n(X_1, X_k) \middle| X_1 \right) \right],$$

$$B = \text{Var} \left( \binom{n-1}{2}^{-1} \sum_{j=2}^{n-1} \sum_{k=j+1}^n \mathbb{E} [k_n(X_1, X_j) k_n(X_1, X_k) | X_1] \right).$$

Note that conditional on  $X_1$ , the term within the variance in  $A$  is a second-order U-statistics with kernel  $k_n(X_1, X_2)k_n(X_1, X_3)$ , hence it admits a Hoeffding decomposition and its variance can be bounded by standard argument. The quadratic term can be shown to be  $O(n^{-2}h_n^{-2})$ . The linear term can be shown to be  $O(n^{-1})$ . This is proved in Lemma C.1. Now, we analyze the term  $B$ . We have

$$\begin{aligned} B &= \text{Var} \left( \binom{n-1}{2}^{-1} \sum_{j=2}^{n-1} \sum_{k=j+1}^n \mathbb{E} [k_n(X_1, X_j) k_n(X_1, X_k) | X_1] \right) \\ &= \text{Var} \left( \binom{n-1}{2}^{-1} \sum_{j=2}^{n-1} \sum_{k=j+1}^n \mathbb{E} [k_n(X_1, X_2) | X_1]^2 \right) \\ &\leq \mathbb{E} [\mathbb{E} [k_n(X_1, X_2) | X_1]^4] = O(1), \end{aligned}$$

where the last equality follows from independence, the law of iterated expectation, and Lemma 2.4.

It follows that

$$R_1 = O(1 + n^{-2}h_n^{-2} + n^{-1}),$$

and so

$$n^{-1}R_1 + 2n^{-2} \binom{n-1}{2}^{-1} R_2 = O(n^{-3}h_n^{-2} + n^{-1} + n^{-2} + n^{-2}h_n^{-1})$$

By  $L^2$ -convergence, it follows that, whenever  $nh_n \rightarrow C \in (0, \infty]$ , we have

$$n^{-1} \sum_{i=1}^n \binom{n-1}{2}^{-1} \sum_{j=1, j \neq i}^{n-1} \sum_{k=j+1, k \neq i}^n k_n(X_i, X_j) k_n(X_i, X_k) = \int_{\mathbb{R}} f_0(x)^3 dx + o_P(1).$$

If  $nh_n \rightarrow 0$ , the same argument shows that

$$nh_n n^{-1} \sum_{i=1}^n \binom{n-1}{2}^{-1} \sum_{j=1, j \neq i}^{n-1} \sum_{k=j+1, k \neq i}^n k_n(X_i, X_j) k_n(X_i, X_k) = o_P(1).$$

A double application of the continuous mapping theorem then yields the result.

2. The proof operates with similar arguments. Note first that

$$\mathbb{E} \left[ h_n \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)^2 \right] = h_n \mathbb{E} [k_n(X_i, X_j)^2].$$

Then

$$\mathbb{E} \left[ \left| h_n \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)^2 - h_n \mathbb{E} [k_n(X_i, X_j)^2] \right|^2 \right] = h_n^2 \text{Var} \left( \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)^2 \right).$$

Note that  $\binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)^2$  is a second-order U-statistics with kernel  $k_n(X_i, X_j)^2$ . By Hoeffding decomposition and density arguments, its variance can be bounded in the same way as the variance of  $U_n$ . As proved in Lemma C.2, we can show that

$$\text{Var} \left( \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)^2 \right) = O(n^{-2} h_n^{-3}) + O(n^{-1} h_n^{-2}).$$

We then have

$$\mathbb{E} \left[ \left| h_n \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)^2 - h_n \mathbb{E} [k_n(X_i, X_j)^2] \right|^2 \right] = O(n^{-2} h_n^{-1} + n^{-1}).$$

By  $L^2$ -convergence, it follows that

$$h_n \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)^2 = \sigma_W^2 + o_P(1).$$

□

#### A.4 Proofs of the inconsistency of the plug-in variance estimator

*Proof.* We start by expanding  $\widehat{\sigma_L^2}$ . We have

$$\begin{aligned} \widehat{\sigma_L^2} &= n^{-1} \sum_{i=1}^n \left( (n-1)^{-1} \sum_{\substack{j=1 \\ j \neq i}}^n k_n(X_i, X_j) - \widehat{u_n^0} \right)^2 \\ &= n^{-1} (n-1)^{-2} \sum_{i=1}^n \left( \sum_{\substack{j=1 \\ j \neq i}}^n k_n(X_i, X_j) \right)^2 - \widehat{u_n^0} \binom{n}{2}^{-1} \sum_{\substack{j=1 \\ j \neq i}}^n k_n(X_i, X_j) + \widehat{u_n^0}^2 \\ &= n^{-1} (n-1)^{-2} \sum_{i=1}^n \left( \sum_{\substack{j=1 \\ j \neq i}}^n k_n(X_i, X_j) \right)^2 - \widehat{u_n^0}^2. \end{aligned}$$

We now expand the first term on the right-end side

$$\begin{aligned}
& n^{-1}(n-1)^{-2} \sum_{i=1}^n \left( \sum_{\substack{j=1 \\ j \neq i}}^n k_n(X_i, X_j) \right)^2 \\
&= n^{-1}(n-1)^{-2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n k_n(X_i, X_j)^2 + 2n^{-1}(n-1)^{-2} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^{n-1} \sum_{\substack{k=j+1 \\ k \neq i}}^n k_n(X_i, X_j) k_n(X_i, X_k) \\
&= \widetilde{\sigma_{LW}^2} + (n-2)(n-1)^{-1} \widetilde{\sigma_L^2}.
\end{aligned}$$

The result then follows directly from Lemma 3.1 and Remark 3.4.  $\square$

*Proof.* We start by expanding  $\widehat{\sigma_W^2}$ . We have

$$\begin{aligned}
\widehat{\sigma_W^2} &= h_n \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left[ k_n(X_i, X_j)^2 + (\widehat{l}_{n,i} + \widehat{l}_{n,j})^2 + \widehat{u}_n^0{}^2 \right. \\
&\quad \left. - 2k_n(X_i, X_j)(\widehat{l}_{n,i} + \widehat{l}_{n,j}) + 2k_n(X_i, X_j)\widehat{u}_n^0 - 2(\widehat{l}_{n,i} + \widehat{l}_{n,j})\widehat{u}_n^0 \right] \\
&= h_n \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \left[ k_n(X_i, X_j)^2 + \widehat{l}_{n,i}^2 + \widehat{l}_{n,j}^2 + 3\widehat{u}_n^0{}^2 \right. \\
&\quad + 2(n-1)^{-2} \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{l=1 \\ l \neq j}}^n k_n(X_i, X_k) k_n(X_j, X_l) \\
&\quad - 2(n-1)^{-1} k_n(X_i, X_j) \left( \sum_{\substack{k=1 \\ k \neq i}}^n k_n(X_i, X_k) + \sum_{\substack{l=1 \\ l \neq j}}^n k_n(X_j, X_l) \right) \\
&\quad \left. - 2k_n(X_i, X_j)\widehat{u}_n^0 - 4(\widehat{l}_{n,i} + \widehat{l}_{n,j})\widehat{u}_n^0 \right]
\end{aligned}$$

The only terms that are not directly covered by the previous results are the summands with cross-terms, namely

$$k_n(X_i, X_j) \left( \sum_{\substack{k=1 \\ k \neq i}}^n k_n(X_i, X_k) + \sum_{\substack{l=1 \\ l \neq j}}^n k_n(X_j, X_l) \right)$$

and

$$\sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{l=1 \\ l \neq j}}^n k_n(X_i, X_k) k_n(X_j, X_l).$$

We show that all these terms are  $o_p(1)$  by using previous results. We first have

$$\left| \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j) \sum_{\substack{k=1 \\ k \neq i}}^n k_n(X_i, X_k) \right| \leq \sum_{i=1}^n \left( \sum_{\substack{j=1 \\ j \neq i}}^n |k_n(X_i, X_j)| \right)^2$$

and

$$\left| \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j) \sum_{\substack{l=1 \\ l \neq j}}^n k_n(X_j, X_l) \right| \leq \sum_{i=1}^n \left( \sum_{\substack{j=1 \\ j \neq i}}^n |k_n(X_i, X_j)| \right)^2.$$

Similarly, we have

$$\left| \sum_{i=1}^{n-1} \sum_{j=i+1}^n \sum_{\substack{k=1 \\ k \neq i}}^n \sum_{\substack{l=1 \\ l \neq j}}^n k_n(X_i, X_k) k_n(X_j, X_l) \right| \leq 3 \sum_{i=1}^n \left( \sum_{\substack{j=1 \\ j \neq i}}^n |k_n(X_i, X_j)| \right)^2.$$

From the proof of Lemma 3.2, we directly get that

$$(n-1)^{-1} h_n \binom{n}{2}^{-1} \sum_{i=1}^n \left( \sum_{\substack{j=1 \\ j \neq i}}^n |k_n(X_i, X_j)| \right)^2 = o_P(1).$$

Then by Lemma 3.1 and Lemma 3.2, we get that all terms in the expansion are  $o_P(1)$ , except

$$\widetilde{\sigma_W^2} = h_n \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)^2 = \sigma_W^2 + o_P(1).$$

This concludes the proof.  $\square$

## A.5 Proof of the inconsistency of the non-parametric bootstrap

*Proof of Proposition 3.4.* From the multinomial representation, we have

$$\begin{aligned} \text{Var}^*(u_n^{1*}(X_i^*)) &= \mathbb{E}^*[(u_n^{1*}(X_i^*))^2] - (\mathbb{E}^*[u_n^{1*}(X_i^*)])^2 \\ &= \frac{1}{n^3} \sum_{i=1}^n \left( \sum_{j=1}^n k_n(X_i, X_j) \right)^2 - \left( \frac{n-1}{n} \right)^2 U_n^2 \\ &= \left( \frac{n-1}{n} \right)^2 \widehat{\sigma_L^2} \end{aligned}$$

and

$$\begin{aligned} \text{Var}^*(u_n^{2*}(X_i^*, X_j^*)) &= \mathbb{E}^*[k_n(X_i^*, X_j^*)^2] - (\mathbb{E}^*[k_n(X_i^*, X_j^*)])^2 \\ &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n k_n(X_i, X_j)^2 - \left( \frac{n-1}{n} \right)^2 U_n^2 \\ &= h_n^{-1} \frac{n-1}{n} \widetilde{\sigma_W^2} - \left( \frac{n-1}{n} \right)^2 U_n^2. \end{aligned}$$

It follows that

$$\text{Var}^* U_n^* = \frac{4}{n} \left( \frac{n-1}{n} \right)^2 \widehat{\sigma_L^2} + \binom{n}{2}^{-1} \left[ h_n^{-1} \frac{n-1}{n} \widetilde{\sigma_W^2} - \left( \frac{n-1}{n} \right)^2 U_n^2 - 2 \left( \frac{n-1}{n} \right)^2 \widehat{\sigma_L^2} \right].$$

Then the result follows directly from Lemma 3.2 and Lemma 3.1.  $\square$

## B A density argument

**Lemma B.1.** *Let  $f \in L^1 \cap L^\infty$  and  $a, b \in \mathbb{R}$ . Then*

$$\lim_{h \rightarrow 0} \int |f(x+ah)f(x+bh) - f^2(x)| dx = 0.$$

*Proof.* We make use of a density argument. Suppose  $g$  is continuous and compactly supported. Then

$$\begin{aligned} \int |g(x+ah)g(x+bh) - g^2(x)| dx &\leq \lambda(K) \sup_x |g(x+ah)g(x+bh) - g^2(x)| \\ &\rightarrow 0 \text{ as } h \rightarrow 0, \end{aligned} \tag{B.1}$$

since

$$|ab - cd| \leq |b||a - c| + |c||b - d|$$

which yields

$$\begin{aligned} &\sup_x |g(x+ah)g(x+bh) - g^2(x)| \\ &\leq \sup_x |g(x+ah)||g(x+bh) - g(x)| + \sup_x |g(x)||g(x+ah) - g(x)| \\ &\leq c_1 \sup_x |g(x+bh) - g(x)| + c_2 \sup_x |g(x+ah) - g(x)| \\ &\rightarrow 0 \text{ as } h \rightarrow 0, \text{ by uniform continuity of } g. \end{aligned}$$

Now, by density of  $C_c$  in  $L^1$ , for  $\varepsilon > 0$ , there is  $g \in C_c$  satisfying  $\int |f(x) - g(x)| dx \leq \varepsilon$ . Then

$$\begin{aligned} &\int |f(x+ah)f(x+bh) - g(x+ah)g(x+bh)| dx \\ &\leq \|f\|_\infty \int |f(x+ah) - g(x+ah)| dx + \|g\|_\infty \int |f(x+bh) - g(x+bh)| dx \\ &\leq 2(\|f\|_\infty + \|g\|_\infty)\varepsilon \end{aligned}$$

by density and translation invariance. Moreover,

$$\begin{aligned} \int |f^2(x) - g^2(x)| dx &\leq \|f\|_\infty \int |f(x) - g(x)| dx + \|g\|_\infty \int |f(x) - g(x)| dx \\ &\leq 2(\|f\|_\infty + \|g\|_\infty)\varepsilon \end{aligned}$$

by density. Finally, by (B.1),

$$\int |g(x+ah)g(x+bh) - g^2(x)| \leq \varepsilon$$

for  $h$  sufficiently close to 0. It follows by the triangular inequality and the previous inequalities that

$$\begin{aligned}
& \int |f(x+ah)f(x+bh) - f^2(x)| dx \\
& \leq \int |f(x+ah)f(x+bh) - g(x+ah)g(x+bh)| dx \\
& + \int |g(x+ah)g(x+bh) - g^2(x)| + \int |f^2(x) - g^2(x)| \\
& \leq (1 + 4(\|f\|_\infty + \|g\|_\infty))\varepsilon.
\end{aligned}$$

□

The same argument can be extended by extending the absolute inequality for higher order products, e.g., for  $k = 3$ ,

$$|abc - def| \leq |b||c||a - d| + |c||d||b - e| + |d||e||c - f|.$$

By the exact same density argument, we obtain the following extension.

**Corollary B.1.** *Let  $f \in L^1 \cap L^\infty$  and  $A \subset \mathbb{R}$  a finite subset. Then*

$$\lim_{h \rightarrow 0} \int |\Pi_{i \in A} f(x + ih) - f^{\#A}(x)| dx = 0.$$

## C Additional bounds on the variance of some U-statistics

When proving  $L^2$  convergence for estimators of the variance, a number of other second-order U-statistics appear whose variance need to be bounded. In this section, we collect some of these bounds. We also bound the covariance between two related U-statistics.

First, consider

$$A_1 = \binom{n-1}{2}^{-1} \sum_{j=2}^{n-1} \sum_{k=j+1}^n k_n(X_1, X_j) k_n(X_1, X_k).$$

Conditional on  $X_1$ , this is a second-order U-statistics with kernel  $k_n(X_1, X_j)k_n(X_1, X_k)$ . By the Hoeffding decomposition and the general variance bounds for second-order U-statistics, we know that, a.s.,

$$\text{Var}(A_1|X_1) \leq \kappa_1(n^{-2}\mathbb{E}[k_n(X_1, X_2)^2 k_n(X_1, X_3)^2|X_1] + n^{-1}\mathbb{E}[\mathbb{E}[k_n(X_1, X_2)k_n(X_1, X_3)|X_2]^2|X_1])$$

where  $\kappa_1 > 0$ , and so, by monotonicity and tower property of the expectation, we have

$$\mathbb{E}[\text{Var}(A_1|X_1)] \leq \kappa_1(n^{-2}\mathbb{E}[k_n(X_1, X_2)^2 k_n(X_1, X_3)^2] + n^{-1}\mathbb{E}[\mathbb{E}[k_n(X_1, X_2)k_n(X_1, X_3)|X_2]^2]).$$

**Lemma C.1.** *We have that*

$$\mathbb{E}[\text{Var}(A_1|X_1)] = O(n^{-2}h_n^{-2} + n^{-1})$$

*Proof.* The bound for the quadratic term follows directly from Lemma 2.3, that is,

$$\mathbb{E} [k_n(X_1, X_2)^2 k_n(X_1, X_3)^2] = O(h_n^{-2}).$$

For the linear term, a subtle application of the mollification theorem in two dimensions can deliver the result. However, a simpler argument using the properties of conditional expectations is presented. Note that

$$\begin{aligned} \mathbb{E} [\mathbb{E} [k_n(X_1, X_2) k_n(X_1, X_3) | X_2]^2] \\ &= \mathbb{E} [\mathbb{E} [k_n(X_1, X_2) k_n(X_1, X_3) | X_2] \mathbb{E} [k_n(X_4, X_2) k_n(X_4, X_5) | X_2]] \\ &= \mathbb{E} [\mathbb{E} [k_n(X_1, X_2) k_n(X_1, X_3) k_n(X_4, X_2) k_n(X_4, X_5) | X_2]] \\ &= \mathbb{E} [k_n(X_1, X_2) k_n(X_1, X_3) k_n(X_4, X_2) k_n(X_4, X_5)]. \end{aligned}$$

Then, by Lemma 2.4, it follows that

$$\mathbb{E} [\mathbb{E} [k_n(X_1, X_2) k_n(X_1, X_3) | X_2]^2] = O(1).$$

□

Consider now

$$A_2 = \binom{n}{2}^{-1} \sum_{i=1}^{n-1} \sum_{j=i+1}^n k_n(X_i, X_j)^2.$$

This is a second-order U-statistics with kernel  $k_n(X_i, X_j)^2$ . By the Hoeffding decomposition and the general variance bounds for second-order U-statistics, we know that

$$\text{Var } A_2 \leq \kappa_2(n^{-2} \mathbb{E} [k_n(X_1, X_2)^4] + n^{-1} \mathbb{E} [\mathbb{E} [k_n(X_1, X_2)^2 | X_2]^2])$$

for some  $\kappa_2 > 0$ .

**Lemma C.2.** *We have that*

$$\text{Var } A_2 = O(n^{-2} h_n^{-3} + n^{-1} h_n^{-2}).$$

*Proof.* The bound for the quadratic term follows directly from Lemma 2.3, that is,

$$\mathbb{E} [k_n(X_1, X_2)^4] = O(h_n^{-3}).$$

For the linear term, we use the properties of the conditional expectation. We have

$$\mathbb{E} [\mathbb{E} [k_n(X_1, X_2)^2 | X_2]^2] = \mathbb{E} [k_n(X_1, X_2)^2 k_n(X_3, X_2)^2]$$

and again by Lemma 2.3, we have

$$\mathbb{E} [\mathbb{E} [k_n(X_1, X_2)^2 | X_2]^2] = O(h_n^{-2}).$$

□

Now, consider the covariance

$$R_2 = \text{cov} \left( \sum_{i=2}^{n-1} \sum_{j=i+1}^n k_n(X_1, X_i) k_n(X_1, X_j), \sum_{\substack{k=1 \\ k \neq 2}}^{n-1} \sum_{\substack{l=k+1 \\ l \neq 2}}^n k_n(X_2, X_k) k_n(X_2, X_l) \right).$$

Modulo the scaling factors, this is the covariance between two U-statistics (with random kernels anchored at  $X_1$  and  $X_2$ , respectively). By finitude of all moments, a rough bound is directly given by  $O(n^4 h_n^{-2})$ . However, because of the i.i.d. assumption, many of summands are 0, namely all those such that  $1, i, j$  are all different from  $2, k, l$ . This allows us to drastically refine the bound. We have that

**Lemma C.3.** *We have that*

$$R_2 = O(n^3 + n^2 h_n^{-1} + n h_n^{-2}).$$

*Proof.* Note first that the following expansion holds

$$\begin{aligned} R_2 = & \sum_{i=2}^{n-1} \sum_{j=i+1}^n \sum_{l=3}^n \text{cov}(k_n(X_1, X_i) k_n(X_1, X_j), k_n(X_2, X_l) k_n(X_2, X_l)) \\ & + \sum_{j=3}^n \sum_{\substack{k=1 \\ k \neq 2}}^{n-1} \sum_{\substack{l=k+1 \\ l \neq 2}}^n \text{cov}(k_n(X_1, X_2) k_n(X_1, X_j), k_n(X_2, X_k) k_n(X_2, X_l)) \\ & + \sum_{i=3}^{n-2} \sum_{j=i+1}^{n-1} \sum_{l=j+1}^n \text{cov}(k_n(X_1, X_i) k_n(X_1, X_j), k_n(X_2, X_j) k_n(X_2, X_l)) \\ & + \sum_{i=3}^{n-1} \sum_{j=i+1}^n \sum_{l=i+1}^n \text{cov}(k_n(X_1, X_i) k_n(X_1, X_j), k_n(X_2, X_i) k_n(X_2, X_l)) \\ & + \sum_{i=4}^{n-1} \sum_{j=i+1}^n \sum_{k=3}^{i-1} \text{cov}(k_n(X_1, X_i) k_n(X_1, X_j), k_n(X_2, X_k) k_n(X_2, X_i)), \end{aligned}$$

which we rewrite as

$$R_2 = S_1 + S_2 + S_3 + S_4 + S_5.$$

Note now that, for all  $m \in \{1, 2, 3, 4, 5\}$ ,

$$S_m = O(n^3 + n^2 h_n^{-1} + n h_n^{-2}).$$

This follows from Lemma 2.4 and the fact that there are 4 different indexes in each summand, except when one of the free indexes is exactly equal to one of the other indexes. This is shown for

$S_1$  for illustration, but the same argument applies to the other terms. We have

$$\begin{aligned}
S_1 = & \sum_{i=3}^{n-1} \sum_{j=i+1}^n \sum_{\substack{l=3 \\ l \neq i, l \neq j}}^n \text{cov}(k_n(X_1, X_i)k_n(X_1, X_j), k_n(X_2, X_1)k_n(X_2, X_l)) \\
& + \sum_{i=3}^{n-1} \sum_{j=i+1}^n \text{cov}(k_n(X_1, X_i)k_n(X_1, X_j), k_n(X_2, X_1)k_n(X_2, X_i)) \\
& + \sum_{i=3}^{n-1} \sum_{j=i+1}^n \text{cov}(k_n(X_1, X_i)k_n(X_1, X_j), k_n(X_2, X_1)k_n(X_2, X_j)) \\
& + \sum_{j=3}^n \text{cov}(k_n(X_1, X_2)k_n(X_1, X_j), k_n(X_2, X_1)k_n(X_2, X_j)) \\
& + \sum_{j=3}^n \sum_{\substack{l=3 \\ l \neq j}}^n \text{cov}(k_n(X_1, X_2)k_n(X_1, X_j), k_n(X_2, X_1)k_n(X_2, X_l))
\end{aligned}$$

By the i.i.d. assumption, each summand in the five terms of the expansion is equal, respectively, to

$$\begin{aligned}
& \text{cov}(k_n(X_1, X_3)k_n(X_1, X_4), k_n(X_2, X_1)k_n(X_2, X_5)) \\
& \text{cov}(k_n(X_1, X_3)k_n(X_1, X_4), k_n(X_2, X_1)k_n(X_2, X_3)) \\
& \text{cov}(k_n(X_1, X_3)k_n(X_1, X_4), k_n(X_2, X_1)k_n(X_2, X_4)) \\
& \text{cov}(k_n(X_1, X_2)k_n(X_1, X_3), k_n(X_2, X_1)k_n(X_2, X_3)) \\
& \text{cov}(k_n(X_1, X_2)k_n(X_1, X_3), k_n(X_2, X_1)k_n(X_2, X_4))
\end{aligned}$$

From Lemma 2.4, it follows that the first term in the expansion is  $O(n^3)$ , the second, third, and fifth terms are  $O(n^2 h_n^{-1})$ , while the fourth one is  $O(n h_n^{-2})$ .  $\square$

## D Bias results in Giné and Nickl (2008a)

**Lemma D.1** (Part 1 of Theorem 1 in Giné and Nickl (2008a)). *If  $K$  satisfies Assumption K and  $f_0$  satisfies Assumption D(s) with  $s \in (0, 1/2]$ . Then the bias of  $U_n$  satisfies*

$$\mathbb{E}[U_n] - \theta_0 = O(h_n^{2s}).$$

*Proof.* Write  $K_{h_n}(x) = h_n^{-1} K_n(h_n^{-1}x)$ . We have

$$\begin{aligned}
\mathbb{E}[U_n] - \theta_0 &= \int_{\mathbb{R}} \int_{\mathbb{R}} K_{h_n}(x-y) f_0(y) dy f_0(x) dx - \int_{\mathbb{R}} f_0(x) f_0(x) dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} K_{h_n}(x-y) (f_0(y) - f_0(x)) f_0(x) dy dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} K(u) (f_0(x - uh_n) - f_0(x)) f_0(x) du dx \\
&= \int_{\mathbb{R}} K(u) \left( \int_{\mathbb{R}} \bar{f}_0(uh_n - x) f_0(x) dx - \int_{\mathbb{R}} \bar{f}_0(0 - x) f_0(x) dx \right) du \\
&= \int_{\mathbb{R}} K(u) ((\bar{f}_0 * f_0)(uh_n)) - (\bar{f}_0 * f_0)(0) du,
\end{aligned}$$

where  $*$  denotes convolution and  $\bar{f}_0(x) = f_0(-x)$ . The second equality follows from the fact that the kernel integrates to one, the third from the change of variable  $y = x - uh_n$ , and the fourth from Fubini's theorem. Then by applying Lemma D.2 since  $f_0 \in L^1$  as a probability density function, we get

$$\begin{aligned}
|\mathbb{E}[U_n] - \theta_0| &\leq \int_{\mathbb{R}} C \|f_0\|_{2,s}^2 |K(u)| |uh_n|^{2s} du \\
&= c_1 h_n^{2s}
\end{aligned}$$

where  $c_1 = C \|f_0\|_{2,s}^2 \int_{\mathbb{R}} |K(u)| |u|^{2s} du$  and  $0 < C < \infty$  is a constant independent of  $f_0$  and  $h_n$ .  $\square$

**Lemma D.2** (Lemma 1 in Giné and Nickl (2008a)). *If  $f, g \in L^1$  satisfy Assumption D(s) with  $0 < s \leq 1/2$ , then for any  $x \in \mathbb{R}$  and  $t \neq 0$ ,*

$$\frac{|(f * g)(x+t) - (f * g)(x)|}{|t|^{2s}} \leq C \|f\|_{2,s} \|g\|_{2,s}$$

where  $0 < C < \infty$  is a fixed constant independent of  $f, g, x$  and  $t$ .

*Proof.* Denote  $F$  the Fourier transform. Since  $f, g \in L^1$  and  $g$  bounded,  $f * g \in L^1$  and continuous, and since  $f, g \in L^2$ , we have  $F(f * g) \in L^1$ . We then have

$$\begin{aligned}
\frac{|(f * g)(x+t) - (f * g)(x)|}{|t|^{2s}} &\leq |t|^{-2s} \|F^{-1} F[(f * g)(\cdot + t) - (f * g)(\cdot)]\|_{\infty} \\
&\leq (2\pi)^{-1} |t|^{-2s} \|F[(f * g)(\cdot + t) - (f * g)(\cdot)]\|_1 \\
&= (2\pi)^{-1} |t|^{-2s} \int_{\mathbb{R}} |F(f * g)(u)(e^{-iut} - 1)| du \\
&= (2\pi)^{-1} \int_{\mathbb{R}} |Ff(u)| |u|^s |Fg(u)| |u|^s \frac{|(e^{-iut} - e^{-i0})|}{|u|^{2s} |t|^{2s}} du \\
&\leq C \|f\|_{2,s} \|g\|_{2,s}.
\end{aligned}$$

The first inequality follows from the definition of the  $L^{\infty}$  norm and the Fourier inversion theorem, the second from the inequality  $\|f\|_{\infty} \leq \|Ff\|_1$  (which also follows from the Fourier inversion

theorem). The first equality follows from the definition of the  $L^1$  norm and the second from the convolution theorem. The last inequality follows from Hölder's inequality and the fact that  $e^{-i(\cdot)}$  is bounded Lipschitz.  $\square$

*Remark D.1.* The assumption  $s \leq 1/2$  is needed for  $c_1$  to be finite under assumption K. This can be relaxed if  $\int |K(u)| |u|^{2s} du < \infty$ . If the kernel  $K$  is non-negative (and so is a density function), then this condition is equivalent to the random variable with density  $K$  has finite  $2s$  order moments. This is often the case for kernel of order 1, and so the result can be naturally generalized to  $s \geq 1/2$ .

*Remark D.2.* The same proof is reproduced in [Cattaneo and Jansson \(2022\)](#). They use the following trick to avoid introducing any Fourier analysis arguments. The last equality in the first display of the proof of Lemma D.1. can be interpreted as the bias of density estimation where the density is  $\tilde{f}_0 * f_0(\cdot)h_n$ . A similar argument as Lemma D.2 can be used (namely Lemma 12 in [Giné and Nickl \(2008b\)](#)) to show that convolution of functions in Besov spaces belong to some Hölder spaces. Then results on the bias for density estimation in Hölder spaces can be invoked directly (e.g., Proposition 1.2. in [Tsybakov \(2009\)](#)). A consequence of this argument is that the bias result Lemma D.1 generalizes to a slightly larger smoothness class, the same as [Laurent \(2005\)](#), namely a Besov class with smoothness parameter  $s > 1/4$ .