INTEGRATED SQUARE OF A DENSITY: SEMIPARAMETRIC INFERENCE BEYOND THE $o_p(n^{-1/4})$ RULE

Paul Delatte

University of Southern California, delatte@usc.edu

We tackle the problem of quantifying uncertainty in an optimal estimator of $\int_{\mathbb{R}} f_0^2(x)\,dx$ when f_0 belongs to some class of bounded densities with smoothness $s \in (0, 1/2)$. The functional in this case is paradigmatic of semiparametric models where the nuisance parameter cannot be estimated with $o_p(n^{-1/4})$ errors and optimal semiparametric estimation requires more than Neyman-orthogonalization. We assume that s is known to evade the limits of adaptive uncertainty quantification and estimate $\int_{\mathbb{R}} f_0^2(x)\,dx$ optimally using a simple U-statistic kernel-based estimator. We first prove that the asymptotic normality of the estimator, which is known to hold in the parametric regime with $s \in (1/4, 1/2)$, extends to the nonparametric regime with $s \in (0, 1/4)$. We then show that it is possible to build unbiased variance estimators from higher-order U-statistics that converge to the limiting variances at the best possible rates in each regime. By contrast, we prove that variance estimators obtained from the plug-in principle and the nonparametric bootstrap are biased, converge at suboptimal rates in the parametric regime, and diverge in the nonparametric regime. Subsampling can be used to restore consistency in the nonparametric regime, but does not provide improvements in the parametric regime. We use the variance estimators to build Wald-type confidence intervals whose finite-sample coverage and length are estimated through Monte Carlo simulations at different smoothness levels $s \in (0, 1/2)$. The results we obtain for this problem are a first step in delineating the applicability of semiparametric models beyond high-regularity conditions.

Keywords: U-statistic; kernel density estimator; semiparametric inference; semiparametric bootstrap

1 Introduction

1.1 Motivation and results overview

Semiparametric models, characterized by a low-dimensional parameter of interest and an infinite-dimensional nuisance parameter, have been at the forefront of statistical and econometric theory for decades. This central position can be easily explained by the benefits these models provide in fields where specification is a major challenge but the relevant information for

---

This version: 19 May, 2026. We thank Timothy Armstrong, Yuehao Bai, Stéphane Bonhomme, Kengo Kato, Hyungsik Roger Moon, Peter C. B. Phillips, Geert Ridder, and Angela Zhou for helpful comments. Some of the results in this paper were previously circulated in a different manuscript titled "Average density: weak limits and inference in non-regular semi-parametric problems".

decision-making can be summarized in a few dimensions. These features have led to uncountable applications of semiparametric models in fields such as economics, finance, or medicine from the 1980s to the present day. Recent advances in nonparametric estimation (under the marketable term of “machine learning") and new applications in causal inference have led a new generation of statisticians and econometricians to pay close attention to these models – rediscovering in the process many facts previously established. One main advantage of semiparametric models is that they come with a simple unified procedure for estimation and inference with parametric guarantees under regularity conditions. The idea is to plug nonparametric estimators of the nuisance parameters in the population characterization of the low-dimensional parameter of interest (be it through a functional equation, a moment condition, or the minimization of a loss function). It is well known that this naive plug-in approach can generate a substantial bias that may be corrected by estimating the influence function(s). In the context of M-estimation, these types of corrections have come to be known as “Neyman-orthogonalization” of the objective function. A major result of semiparametric theory is that problems that are Neyman-orthogonal (by nature or by correction) lead to plug-in estimators that are asymptotically unbiased at the parametric rate $\sqrt{n}$ provided the nuisance parameter can be estimated with $o_p(n^{-1/4})$ errors. In this case, procedures for the construction of confidence intervals that work in the parametric world transpose (almost) without any issue to the semiparametric world – including the nonparametric bootstrap and other resampling schemes. The regularity conditions under which these results hold have been formalized over the last forty years – see, for instance, Andrews (1994), Newey (1994), Chen, Linton, and Van Keilegom (2003), Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey, and Robins (2018) on the econometric side. These results now serve as a basis for inference in most applications of semiparametric models with little (if any) discussion of the validity of the regularity conditions in each particular instantiation. Nonetheless, the rate conditions for the nuisance parameter are far from trivial. Estimation in functional spaces is inherently hard, and so independently of the chosen estimation routine and the available computational power. Moreover, the difficulty of the task increases prohibitively with the dimension of the domain space in what is famously known in nonparametric statistics as the curse of dimensionality. Consequently, any empirical result obtained from semiparametric models where non-trivial rate conditions on the nuisance are used without strong a priori justifications shall be treated with some reservations.

However, the profusion of hard-to-estimate nuisance parameters shall not condemn a priori the use of semiparametric models. Another essential result of semiparametric theory is that, even when the nuisance parameters cannot be estimated with $o_p(n^{-1/4})$ errors, it may still be possible to construct estimators of the parameter of interest that converge at parametric rate. The possibility to do so and the best rates that can be achieved when $\sqrt{n}$-estimation is unfeasible vary from problem to problem. No general theory yet exists to directly characterize the optimal rates under lower regularity and no unified procedure is yet available to easily construct optimal estimators in this case. Nonetheless, it is known through examples that plug-in estimators after Neyman-orthogonalization may not be sufficient and further refinements may be needed to obtain estimators that attain the best possible rates. Results of this nature were first obtained in simple semiparametric models defined by integral functionals such as $\int_{\mathbb{R}} f_0^2(x)\,dx$. For this paradigmatic problem of semiparametric theory, Bickel and Ritov (1988) obtained tight bounds on the convergence rates of any estimator of $\int_{\mathbb{R}} f_0^2(x)\,dx$ when f_0 belongs to some class of densities with smoothness $s \in (0, 1/2)$ preventing its estimation with $o_p(n^{-1/4})$ errors. The authors unearthed the famed phase transition at $s = 1/4$ from the parametric rate $\sqrt{n}$ to the nonparametric rates $n^{4s/(4s+1)}$ and built a corrected plug-in kernel-based estimator attaining them. These results served as a basis for many subsequent extensions that led to a more complete picture of semiparametric estimation beyond Neyman-orthogonalization and high-regularity conditions. A natural requirement for the construction of optimal estimators in this context is that the regularity level of the nuisance (i.e., the value of s for the functional class containing f_0) is known or can be approximated to some degree. The possibility to do so from the data in view of estimation has led to a number of famed results and associated techniques building on Lepski's method – see Lepskii (1991), Lepski and Spokoiny (1997). Unfortunately, these techniques do not translate so well for the quantification of uncertainty as exemplified by a large literature on adaptive confidence sets in purely nonparametric problems – see, for instance, Robins and van der Vaart (2006) or Section 8.3 in Giné and Nickl (2021). Because of these limitations, the question of uncertainty quantification in optimal estimators for semiparametric models under lower regularity has not been explored as thoroughly. However, the problem remains of central importance for the applicability of semiparametric models beyond high-regularity conditions. In this context, one may still want to assume that the regularity level of the nuisance is known and use an optimal estimator to perform inference (provided it exists). Even in this simpler setting, little is known on how to properly quantify uncertainty in optimal semiparametric estimators.

To make progress in this direction, we focus on the simple and paradigmatic problem of estimating $\int_{\mathbb{R}} f_0^2(x)\,dx$ when f_0 belongs to some class of bounded densities with smoothness $s \in (0, 1/2)$. In this case, the density cannot be estimated with $o_p(n^{-1/4})$ errors. However, the "bias problem" is well-understood and optimal estimators of $\int_{\mathbb{R}} f_0^2(x)\,dx$ are directly available – all rely on refinements beyond Neyman-orthogonalization. Among them, the plug-in kernel-based U-statistic estimator considered by Giné and Nickl (2008) stands out by virtue of its simplicity while capturing many of the essential features of any optimal estimator for the problem. This point is made clear in Section 2 and motivates its use to tackle the problem of uncertainty quantification when $s \in (0, 1/2)$. It is already known from Giné and Nickl (2008) that the estimator is asymptotically normal in the parametric regime $s \in (1/4, 1/2)$ with limiting variance $4(\int_{\mathbb{R}} f_0^3(x)\,dx - (\int_{\mathbb{R}} f_0^2(x)\,dx)^2)$. Our first result is to show that asymptotic normality extends to the nonparametric regime $s \in (0, 1/4)$ with limiting variance a scaled version of the parameter itself. For this purpose, we leverage the U-statistic structure of the estimator and verify that the conditions of the central limit theorem of de Jong (1987) apply. In view of inference, we then extend the bias bound of Giné and Nickl (2008) from the Sobolev spaces $W^{2,s}(\mathbb{R})$ to the Besov spaces $B^s_{2,\infty}(\mathbb{R})$ as defined in Section 2.1. This allows us to make the bounds on the best possible rates bind both uniformly and pointwise for some densities in the space. Based on these results, we then tackle the problem of variance estimation. We show that it is possible to build an optimal variance estimator from higher-order U-statistics. This variance estimator is unbiased in finite samples and converges at the best possible rates to the respective limiting variances. The estimator is adaptive in the sense that it converges to the right limit at the optimal rates in each regime and so without any modification. Moreover, it is optimal under the same bandwidth sequence as used for the optimal estimation of $\int_{\mathbb{R}} f_0^2(x)\, dx$. In stark contrast to these results, we exhibit a number of pathologies that affect all standard methods for variance estimation typically used under high-regularity conditions with non-optimal estimators. We start by showing that plug-in variance estimators based on the empirical measure are biased, consistent but sub-optimal in the parametric regime, and inconsistent in the nonparametric regime. The best bandwidth sequence for these estimators is also seen to differ from the optimal bandwidth sequence for the original estimator. If the latter is used for plug-in variance estimation, the convergence rates become arbitrarily slow even in the parametric regime near the boundary. A subtle correction by decoupling the bandwidth sequences restores consistency in the nonparametric regime but does not provide further rate improvements. We then show that the nonparametric bootstrap variance estimator is analytically equivalent to the plug-in variance estimator (without decoupling) and so inherits all the pathologies of the latter. While we do not directly address the problem of bootstrap (distributional) consistency in this paper, we can still use the variance inconsistency result to point to some distributional failure in the nonparametric regime. We complete our results on variance estimation by considering variance estimators obtained by subsampling without replacement. We show that these estimators are still biased in finite samples but that subsampling restores consistency in the nonparametric regime. No further rate improvements are obtained in the parametric regime, but the convergence rates for these estimators are seen to exhibit their own phase transition in the nonparametric regime and match (quite surprisingly) the optimal ones only under very low regularity. The problem of subsampling (distributional) consistency is beyond the scope of this paper and is addressed in another manuscript. To complete and support these theoretical results, we implement a number of Monte Carlo simulations. We use Weierstrass perturbations of regular densities to instantiate bounded densities in $B^s_{2,\infty}$ and construct a simple sampler based on rejection sampling leveraging the analytical expression of these densities. The different variance estimators are then used to build Wald-type confidence intervals based on the normal limit (taking into account the unavoidable bias in the nonparametric regime). For each sample size $n \in \{100, 500, 1000, 2000\}$ and each smoothness level $s \in \{0.05, 0.20, 0.30, 0.45\}$, the average length and coverage of the respective confidence intervals are approximated over 2,000 simulations.

Our results for uncertainty quantification in an optimal estimator of $\int_{\mathbb{R}} f_0^2(x)\, dx$ can be summarized as follows: asymptotic normality holds both in the parametric and nonparametric regimes but the limiting variance undergoes a phase transition; unbiased optimal variance estimation is possible but requires problem-specific corrections; standard methods for variance estimation fail smoothly as the regularity drops with a discontinuity at the phase transition; subsampling can help prevent some of these failures but not all. The possibility to generalize these findings beyond the functional $\int_{\mathbb{R}} f_0^2(x)\, dx$ is an open question. The problem is challenging in the same way that optimality beyond Neyman-orthogonolization does not fall under a unified procedure. However, the integrated square of a density is so paradigmatic of semiparametric models that our results should be at least indicative of similar challenges and remedies in more complicated models. In particular, the optimal variance estimator we build may be directly adapted to models where the optimal estimator of the functional is a second-order n-dependent U-statistic. The problem of uncertainty quantification in adaptive estimators of $\int_{\mathbb{R}} f_0^2(x)\,dx$ is not considered in this paper. The problem is of independent interest, especially at the transition $s = 1/4$, and is left for future research.

1.2 Literature overview

The integrated square of a density is a foundational model of semiparametric statistics. This functional and its direct extensions have been at the forefront of the field and its theoretical development since its inception – see, notably, Levit (1978), Hasminskii and Ibragimov (1979), Pfanzagl (1982), Hall and Marron (1987), Bickel and Ritov (1988), Donoho and Nussbaum (1990), Birgé and Massart (1995), Efromovich and Low (1996), Laurent (1996), and Giné and Nickl (2008). The explanation for its centrality is two-fold. The functional is first of independent interest in itself: it is notably central in information theory since it corresponds exactly to the quadratic Rényi entropy after a log transform – see, e.g., Leonenko, Pronzato, and Savani (2008). Secondly, and more importantly, the functional is one of the simplest semiparametric models when these models are defined by the fact that $\sqrt{n}$-estimation is possible but only if the nuisance parameter satisfies some local regularity conditions. The simplicity of the model can be seen from the identity $\int_{\mathbb{R}} f_0^2(x)\,dx = \int_{\mathbb{R}}\int_{\mathbb{R}} \delta(x-y)\,dP_0(x)\,dP_0(y)$ where δ is the Dirac delta distribution. The model is thus one of the simplest integral functionals with a non-smooth kernel that still allows estimation at the parametric rate. The seemingly simpler model defined by the functional $f_0(z) = \int_{\mathbb{R}} \delta(x-z)\,dP_0(x)$, and naturally known as the density at a point, is actually harder and purely nonparametric in the sense that it completely evades $\sqrt{n}$-estimation even under strong smoothness assumptions on f_0. The consequence is that countless paradigmatic results in semiparametric statistics were first obtained through the study of $\int_{\mathbb{R}} f_0^2(x)\,dx$.

The problem of optimally estimating $\int_{\mathbb{R}} f_0^2(x)\,dx$ has led to a large literature built around the central results of Bickel and Ritov (1988) in which the optimal rates for the model are characterized. The challenge of constructing optimal estimators for the model is reviewed in Section 2.1. Two standard plug-in constructions are commonly considered – one based on kernel density estimators as in Bickel and Ritov (1988) or Giné and Nickl (2008) and one based on orthogonal series estimators as in Laurent (1996). The problem of adaptive estimation for the model has also received significant attention with positive results by Giné and Nickl (2008) and Laurent (2005) under a standard logarithmic cost for the respective constructions. The problem of uncertainty quantification in optimal estimators of $\int_{\mathbb{R}} f_0^2(x)\,dx$, with and without adaptation, has received much less attention. The only available results we are aware of are Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016) and Cattaneo and Jansson (2022): the first paper provides asymptotic normality results for the estimator of Laurent (1996) under low regularity $s \in (0, 1/4)$ and the second paper characterizes the bias of the nonparametric bootstrap distribution for the estimator of Giné and Nickl (2008). We comment on these results at greater length in Remark 3.2 and Remark 4.10, respectively. Beyond the particular problem $\int_{\mathbb{R}} f_0^2(x)\, dx$, very few results are available for uncertainty quantification in semiparametric estimators under low regularity conditions. The only two papers we know of that directly face the challenge are Robins, Li, Tchetgen, van der Vaart, et al. (2008) and Cattaneo and Jansson (2018). Both papers focus on issues created by the bias: the first paper leverages higher-order influence functions to correct for it; the second paper focuses exclusively on non-optimal estimators. Our paper differs in that we tackle the problem of normal approximation and variance estimation for an optimal estimator under low regularity. As far as we know, no other semiparametric model has been characterized at this level.

By virtue of the structure of the optimal estimator we consider, our results also relate to a larger literature on inference based on second-order U-statistics with n-dependent kernels – see, Hall (1984), de Jong (1987), Hardle and Mammen (1993), Laurent (1996), Giné and Nickl (2008), Cattaneo, Crump, and Jansson (2014b), Cattaneo, Crump, and Jansson (2014a), and Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016). The contribution of our normal approximation results to this literature is made clear in Remark 3.2 – of particular interest is the possibility to apply the central limit theorem of de Jong (1987) without local regularity conditions on the density. Our results on variance estimation illustrate a trove of new phenomena for uncertainty quantification in U-statistics with n-dependent kernels. For the nonparametric bootstrap variance estimators, we do not recover the typical inconsistency under quadratic dominance by a factor of 3 as in Hardle and Mammen (1993) or Cattaneo, Crump, and Jansson (2014a), but by an exploding factor corresponding to the quadratic diagonal term. The main difference comes from the fact that in these papers the quadratic diagonal term either cancels automatically or is asymptotically negligible. For variance estimators obtained from subsampling without replacement, our results seem to be new in this literature and highlight some specific challenges of the scheme – of particular interest are the constraints on the subsample choice, the different scaling in each regime, the impossibility to tackle the phase transition, and the optimality of the rates in the very-low-regularity regime.

1.3 Notations

We denote by λ the Lebesgue measure on $\mathbb{R}$ and we write $\int_{\mathbb{R}} g\, d\lambda = \int_{\mathbb{R}} g(x)\, d\lambda(x) = \int_{\mathbb{R}} g(x)\, dx$ for the Lebesgue integral of a measurable function $g\colon \mathbb{R} \to \mathbb{R}$. For $p \in [1, \infty]$, we denote by $L^p = L^p(\mathbb{R}; \lambda)$ the Lebesgue spaces of measurable functions $g\colon \mathbb{R} \to \mathbb{R}$ such that $\|g\|_p < \infty$ where $\|g\|_p^p = \int_{\mathbb{R}} |g|^p\, d\lambda = \int_{\mathbb{R}} |g(x)|^p\, dx$ if $p \in [1, \infty)$ and $\|g\|_\infty = \inf\{C \geq 0 : |g(x)| \leq C \text{ for a.e. } x\}$. For $g \in L^1$, we define the Fourier transform by $F(g)(u) = \int_{\mathbb{R}} e^{-iux} g(x)\, dx$ and we extend it by continuity to L^2. We denote by $\rightsquigarrow$ weak convergence of random variables as $n \to \infty$. We use the Landau notations o, O as $n \to \infty$ and their stochastic variants o_p, O_p as $n \to \infty$. If $a_n = O(b_n)$ and $b_n = O(a_n)$, then we write $a_n \asymp b_n$. We say that a sequence $\{A_n\}$ is stochastically bounded if $A_n = O_p(1)$. Rates of convergence are expressed with positive exponents, e.g., $c_n = n^{1/2}$ or $c_n = \sqrt{n}$ instead of $c_n = n^{-1/2}$. The density and the cumulative distribution function of a standard Gaussian random variable $Z \sim N(0, 1)$ are denoted by $\phi(\cdot)$ and $\Phi(\cdot)$, respectively.

2 Local regularity conditions and optimal estimation

2.1 Regularity, plug-in estimators, and bias corrections

Let $X_1, \ldots, X_n$ be i.i.d. real-valued random variables according to a distribution P_0 admitting a density f_0 with respect to the Lebesgue measure on $\mathbb{R}$. The object of interest is the functional given by

$$T(P_0) := \int_{\mathbb{R}} f_0^2(x)\, dx$$

and known as the integrated square of the density f_0. No restriction is put on the support of f_0, but f_0 is assumed to be (essentially) bounded, that is, $f_0 \in L^\infty$. Since f_0 is a density for some finite measure P_0, it also holds that $f_0 \in L^1$ with $\|f_0\|_1 = 1$. It follows then that $f_0 \in L^p$ for any $p \in [1, \infty]$ and so the functional $T(\cdot)$ is well-defined. When considering inference, the local regularity of the density f_0 has to be controlled. For this purpose, we assume that $f_0 \in B^s_{2,\infty}(\mathbb{R})$ for some $s \in (0, 1/2)$ where $B^s_{2,\infty}(\mathbb{R}) = B^s_{2,\infty}$ is the Besov space defined as

$$B^s_{2,\infty} := \left\{ f \in L^2 : \|f\|_{B^s_{2,\infty}} = \|f\|_2 + \sup_{z \in \mathbb{R}\setminus\{0\}} |z|^{-s} \|f(\cdot + z) - f(\cdot)\|_2 < \infty \right\}.$$

The regularity of the problem is then dictated by the value of s. The lower s is, the larger $B^s_{2,\infty}$ is by inclusion of more irregular functions as measured through their L^2 modulus of continuity. These functional restrictions are standard but variations exist – for a clarification, see Remark 2.1.

Remark 2.1. The local regularity hypotheses on the density are not always identical in the literature. The differences are small, but not always immaterial as made clear in Section 3.3. Bickel and Ritov (1988) considers the spaces defined by $S(s) = \{f \in L^1 : |f(x+z) - f(x)| \le g(x)|z|^s \text{ for all } x \in \mathbb{R},\ |z| < 1\}$ for some function $g \in L^2 \cap L^\infty$. Laurent (1996) and Giné and Nickl (2008) consider the Sobolev spaces $W^{2,s}$ defined as $W^{2,s} := \{f \in L^2 : \|f\|_{2,s} = \|F(f)(\cdot)(1 + |\cdot|^2)^{s/2}\|_2 < \infty\}$. Laurent (2005) considers the spaces $B^s_{2,\infty}$. The following inequalities, where all the inclusions are strict are seen, to hold: $S(s) \subset B^s_{2,\infty}$ and $W^{2,s} \subset B^s_{2,\infty}$; for any $\varepsilon > 0$, $B^{s+\varepsilon}_{2,\infty} \subset W^{2,s} \subset B^s_{2,\infty}$ and $S(s+\varepsilon) \subset W^{2,s}$. The proof of these inclusions follows from the results in Section 4.3.3 in Giné and Nickl (2021) where other useful properties of these spaces can be found.

Remark 2.2. We restrict attention to the one-dimensional problem $\int_{\mathbb{R}} f_0^2(x)\, dx$ as in Giné and Nickl (2008). This restriction is sufficient to illustrate the challenges and remedies to uncertainty quantification in U_n. The results we obtain can be generalized to densities in $\mathbb{R}^d$ with $d > 1$ only at the cost of notational burden. In spite of our focus on the one-dimensional problem, it is worth noting that a central motivation for studying semiparametric problems under low regularity stems from the curse of dimensionality. In the case of $\int_{\mathbb{R}^d} f_0^2(x)\, dx$, the threshold for s under which the density cannot be estimated with $o_p(n^{-1/4})$ errors scales linearly with d.

Remark 2.3. To avoid distinguishing degenerate cases, we shall restrict attention to distributions with densities that are not constant on their support. This condition is often omitted in the literature, but is needed to avoid introducing cases when $\int_{\mathbb{R}} f^3(x)\, dx - (\int_{\mathbb{R}} f^2(x)\, dx)^2 = 0$.

Under these conditions, it follows from Bickel and Ritov (1988) and the inclusions of Remark 2.1 that for any estimator sequence T_n, the rates c_n at which the sequence $c_n(T_n - \int_{\mathbb{R}} f^2(x)\,dx)$ is stochastically bounded uniformly in f over any closed ball $B(f_0, r) \subseteq B^s_{2,\infty}$ with $r > 0$ are upper bounded by $\sqrt{n}$ if $s \in (1/4, 1/2)$ and by $n^{4s/(4s+1)}$ if $s \in (0, 1/4]$. The simplest candidates to achieve these rates are plug-in estimators of the form $T_{n,1} := \int_{\mathbb{R}} \hat{f}_n^2(x)\,dx$ where $\hat{f}_n$ is an estimator of f_0. However, it is seen that estimators of this form converge at rates upper bounded by $n^{2s/(2s+1)}$ and so cannot reach the optimal rates for any value of $s \in (0, 1/2)$. A simple refinement of central importance in semiparametric theory consists in a first-order correction based on the influence function $\mathrm{IF}(x_0, \hat{f}_n) = 2\hat{f}_n(x_0) - 2\int_{\mathbb{R}} \hat{f}_n^2(x)\,dx$. By taking the empirical average of the influence function, this leads to the corrected plug-in estimator $T_{n,3} := \int_{\mathbb{R}} \hat{f}_n^2(x)\,dx + n^{-1}\sum_{i=1}^n (2\hat{f}_n(X_i) - 2\int_{\mathbb{R}} \hat{f}_n^2(x)\,dx) = 2T_{n,2} - T_{n,1}$ where $T_{n,2} := n^{-1}\sum_{i=1}^n \hat{f}_n(X_i)$. However, this correction is of no help to achieve the optimal rates since the estimator sequence $T_{n,3}$ can also be shown to converge at rates that are upper bounded by $n^{2s/(2s+1)}$. In both cases, the limitation comes from the fact that the nuisance density f_0 cannot be estimated with $o_p(n^{-1/4})$ errors in the L^2 sense as soon as $s \in (0, 1/2)$. Indeed, direct and exact expansions in each case lead to $T_{n,1} - T(P_0) = 2P_0(\hat{f}_n - f_0) + \|f_0 - \hat{f}_n\|_2^2$ and $T_{n,3} - T(P_0) = 2(\mathbb{P}_n - P_0)f_0 - 2(\mathbb{P}_n - P_0)(f_0 - \hat{f}_n) - \|f_0 - \hat{f}_n\|_2^2$ where $P_0 f_0 = \int_{\mathbb{R}} f_0\,dP_0 = T(P_0)$ and $\mathbb{P}_n f_0 = \int f_0\,d\mathbb{P}_n = T_{n,2}$. Cross-fitting the first-order corrected estimator $T_{n,3}$ by evaluating $\mathbb{P}_n$ and $\hat{f}_n$ on two independent samples can be used to handle the empirical process remainder $2(\mathbb{P}_n - P_0)(f_0 - \hat{f}_n)$ without invoking Donsker conditions, but this leaves the estimation error $\|f_0 - \hat{f}_n\|_2^2$ as a binding constraint for the convergence rates. In both cases, higher-order corrections are needed to go beyond the binding constraint imposed by the regularity of f_0. However, this correction cannot be based on higher-order influence functions since the second-order influence function for f_0 does not exist – the functional is purely quadratic.

For plug-in estimators $T_{n,1}$ based on a kernel density estimator $\tilde{f}_n$, a simple second-order algebraic correction based on deleting the diagonal terms is possible. This correction turns out to be sufficient to achieve the optimal rates. This is exactly the construction used in Bickel and Ritov (1988). To see it, consider a kernel estimator $\tilde{f}_n := n^{-1}\sum_{i=1}^n K_n(x - X_i)$ for some standard kernel function K_n. By expanding the square, this leads to the plug-in estimator $T_{n,1}(\tilde{f}_n) = n^{-2}\sum_{i=1}^n\sum_{j=1}^n H_n(X_i, X_j)$ where $H_n(X_i, X_j) = \int_{\mathbb{R}} K_n(x - X_i)K_n(x - X_j)\,dx$. It is then seen that the diagonal term satisfies $n^{-2}\sum_{i=1}^n \int_{\mathbb{R}} K_n^2(x - X_i)\,dx = n^{-1}\int_{\mathbb{R}} K_n^2(x)\,dx$ and so is independent of the data. It thus forms a pure bias term that can be deleted. This second-order bias correction (with the superfluous first-order correction) using a two-fold split leads exactly to the estimator of Bickel and Ritov (1988). Following the idea of Hall and Marron (1987), one can use n folds for cross-fitting and directly consider the kernel $K_n(X_i - X_j)$ instead of the convolved one $H(X_i, X_j) = \int_{\mathbb{R}} K_n(x - X_i)K_n(x - X_j)\,dx$. This leads to a satisfying simplification that corresponds exactly to the simple U-statistic estimator considered by Giné and Nickl (2008) and given by $U_n := \binom{n}{2}^{-1}\sum_{i=1}^{n-1}\sum_{j=i+1}^n K_n(X_i - X_j)$. As the estimator of Bickel and Ritov (1988), this estimator can be shown to attain the optimal rates under the conditions above. We will use this estimator in the rest of the paper as it is both simple in the class of rate-optimal estimators and archetypal of cross-fitted bias-corrected estimators beyond first-order corrections. Additional assumptions and known results for U_n are provided in the next section. The same idea behind the construction of U_n can be implemented with estimators of f_0 different from the kernel-based ones $\tilde{f}_n$ provided that they are linear smoothers and satisfy similar invariance properties. The construction was successfully tackled in Laurent (1996) for orthogonal series estimators of f_0.

Remark 2.4. Plug-in estimators such as $T_{n,2} = n^{-1}\sum_{i=1}^n \hat{f}_n(X_i)$ are commonly used in semiparametric problems. In this context, they are sometimes called plug-in estimators by resubstitution, but they more generally correspond to M-estimators based on plug-in estimators for the nuisance. Under the conditions above, estimators in the class of $T_{n,2}$ share the same upper bound for convergence rates as the classes $T_{n,1}$ and $T_{n,3}$. However, the bottleneck with $T_{n,2}$ does not come from $\|f_0 - \hat{f}_n\|_2^2$ but from the empirical process remainder $(\mathbb{P}_n - P_0)(\hat{f}_n - f_0)$. Indeed, a simple and exact expansion gives $T_{n,2} - T(P_0) = (\mathbb{P}_n - P_0)f_0 + \mathbb{P}_n(\hat{f}_n - f_0) = (\mathbb{P}_n - P_0)f_0 + (\mathbb{P}_n - P_0)(\hat{f}_n - f_0) + P_0(\hat{f}_n - f_0)$. This bottleneck can be handled directly by cross-validation and this gives another justification for the form and optimality of estimators of the form of U_n. This derivation also makes clear that it is possible to extend the construction beyond linear smoothers $\tilde{f}_n$.

2.2 Optimal U-statistic and Hoeffding decomposition

To tackle the problem of uncertainty quantification, we consider the simple U-statistic kernel-based estimator U_n of $\int_{\mathbb{R}} f_0^2(x)\,dx$ introduced in Hall and Marron (1987) and further analyzed in Giné and Nickl (2008). This choice is motivated by the simplicity and exemplary nature of U_n among all (known) optimal estimators of $\int_{\mathbb{R}} f_0^2(x)\,dx$. In particular, U_n is seen to capture the essential debiasing correction shared by all optimal estimators of $\int_{\mathbb{R}} f_0^2(x)\,dx$ without any unnecessary complications. We review in this section the main known properties of the estimator, including its Hoeffding decomposition and rate-optimality. The estimator rewrites as

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

where $K\colon \mathbb{R} \to \mathbb{R}$ is a smoothing kernel with associated bandwidth sequence h_n satisfying $0 < h_n$ and $\lim_{n\to\infty} h_n = 0$. Throughout the paper, we assume that K is symmetric and bounded, $\int K(u)\,du = 1$, $\int |K(u)||u|\,du < \infty$, and $K(0) \neq 0$. These conditions imply, in particular, that $K \in L^1 \cap L^\infty$, and hence $K \in L^p$ for any $1 \leq p \leq \infty$. The assumption that $K(0) \neq 0$ is rarely stated explicitly but is needed and is satisfied by all smoothing kernels in practice. When there is a risk of confusion, we make explicit the dependence of U_n on h_n and write $U_n = U_n(h_n)$.

It is seen that U_n is a second-order U-statistic with n-dependent kernel k_n given by

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

for any $1 \leq i < j \leq n$. The n-dependence of the kernel stems from the bandwidth sequence h_n. Since U_n is a second-order U-statistic, it is possible to leverage its Hoeffding decomposition –

Hoeffding (1948) and Section 1.6 in Lee (1990). Let us define

$$L_n := \frac{1}{n}\sum_{i=1}^{n}\Big[k_n(X_i) - \mathbb{E}\,[U_n]\Big],$$

$$W_n := \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\Big[k_n(X_i,X_j) - k_n(X_i) - k_n(X_j) + \mathbb{E}\,[U_n]\Big],$$

where we abuse notation by writing the first-order projection as $k_n(X_i) := \mathbb{E}\,[k_n(X_i,X_j)\,|\,X_i]$ for any $1 \le i < j \le n$. Then the Hoeffding decomposition of U_n reads as

$$U_n = \mathbb{E}\,[U_n] + 2L_n + W_n\,. \tag{2.1}$$

This decomposition leads to a well-known equality for the variance of U_n given by

$$\operatorname{Var} U_n = \frac{4(n-2)}{n(n-1)}\operatorname{Var}\left(k_n(X_1)\right) + \frac{2}{n(n-1)}\operatorname{Var}(k_n(X_1,X_2))\,. \tag{2.2}$$

From a change of variables, the boundedness of f_0, and the integrability of K, it is seen that $\operatorname{Var}\left(k_n(X_1)\right) = O(1)$ and $\operatorname{Var}(k_n(X_1,X_2)) = O(h_n^{-1})$. The variance equality then implies that $\operatorname{Var} U_n = O(n^{-1} \vee n^{-2}h_n^{-1})$, and so the term that asymptotically dominates the sequence $(\operatorname{Var} U_n)^{-1/2}(U_n - \mathbb{E}\,[U_n])$ depends directly on the choice of the bandwidth sequence h_n. If $nh_n \to \infty$, then the linear term L_n dominates. If $nh_n \to 0$, then the quadratic term W_n dominates. If $nh_n \to c > 0$, then L_n and W_n are equivalent. This structure is common to second-order U-statistics with n-dependent kernels – see Hall (1984), Hardle and Mammen (1993), Cattaneo, Crump, and Jansson (2014b), Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016).

For this problem, the choice of h_n can be linked exactly to the regularity of the functional class to which f_0 belongs. Under the assumption that $f_0 \in L^\infty \cap W^{2,s}$ for some $s \in (0,1/2)$ where $W^{2,s}$ is defined in Remark 2.1, Giné and Nickl (2008) proved that the bias of U_n is uniformly bounded by h_n^{2s} and some numerical constant. Using the previous moment bounds, Giné and Nickl (2008) proved that any bandwidth sequence $h_n \asymp n^{-2/(4s+1)}$ ensures that U_n attains the optimal rates under the Sobolev assumption $W^{2,s}$. More precisely, the authors obtained in this case that:

- if $s \in (1/4,1/2)$, then $\sqrt{n}\big(U_n - \int_{\mathbb{R}} f_0^2(x)\,dx\big) \rightsquigarrow N\big(0, 4\big[\int_{\mathbb{R}} f_0^3(x)\,dx - (\int_{\mathbb{R}} f_0^2(x)\,dx)^2\big]\big)$;
- if $s \in (0,1/4]$, then $n^{4s/(4s+1)}\big(U_n - \int_{\mathbb{R}} f_0^2(x)\,dx\big) = O_p(1)$.

In preparation for uncertainty quantification, we will extend these results in two directions. In Section 3.2, we first generalize the asymptotic normality of the sequence to the case $s \in (0,1/4]$. More precisely, we obtain that the stochastic term $O_p(1)$ when $s \in (0,1/4]$ can be exactly characterized as a normal random variable centered at some (unavoidable) constant bias. We then show in Section 3.3 that the bias bounds of Giné and Nickl (2008) can be obtained under the slight relaxation from $W^{2,s}$ to $B^s_{2,\infty}$. The purpose of the relaxation is made clear in the same section. These newly derived limits set the stage for uncertainty quantification in U_n and naturally lead to the open problem of variance estimation for the estimator. The problem is tackled in Section 4. The theoretical results are supported in Section 5 by Monte Carlo simulations for Wald-type confidence intervals built from the normal limits and the different variance estimators.

Remark 2.5. Another important contribution of Giné and Nickl (2008) is to show that adaptive estimation of $\int_{\mathbb{R}} f_0^2(x)\,dx$ using U_n and Lepski's method is possible at a minimal logarithmic cost. We shall not consider the problem of uncertainty quantification in these adaptive estimators. The problem is difficult by virtue of known general impossibility results that affect purely nonparametric problems – see, for instance, Section 8.3 in Giné and Nickl (2021). To evade these difficulties and make progress on the problem of variance estimation in semiparametric models under low regularity, we shall assume that s is known. However, our focus on uncertainty quantification without adaptation does not mean that the problem with adaptation is solved and without interest. Au contraire, the problem is of major importance given the paucity of adaptation results for proper semiparametric models (with likely challenges at the phase transition) and is left for future research.

3 Moment bounds and normal approximations

3.1 Moment bounds for bounded densities

We start by deriving general moment bounds that will be used throughout the proofs. These bounds are obtained using standard properties of the Lebesgue integral and the fact that $f_0 \in L^\infty$. They do not require any local regularity assumption on f_0. The only difficulty is to properly account for the contribution of each cross-term based on their shared indices. We start by bounding the pure terms in Lemma 3.1 before bounding the cross-terms in Lemma 3.2 and Lemma 3.3. The proofs of all results in this section are collected in Section A.1 of the Supplementary Material.

Lemma 3.1. *Let $i, j \in \{1, 2, \ldots, n\}$ with $i \neq j$. Let $r \geq 1$ be any integer. If $f_0 \in L^\infty$, then $\mathbb{E}\left[|k_n(X_i)|^r\right] = O(1)$ and $\mathbb{E}\left[|k_n(X_i, X_j)|^r\right] = O(h_n^{-r+1})$.*

To bound the cross-terms, we introduce some graph notations. We denote by C_2^n the set of all pairs (i, j) with $1 \leq i < j \leq n$. For any pair $(i, j) \in C_2^n$, we let $r_{i,j} \geq 0$ be non-negative integers and denote their sum by $R = \sum_{(i,j)\in C_2^n} r_{i,j}$. We consider the graph with vertices the indices $\{1, 2, \ldots, n\}$. If $r_{i,j} \geq 1$, we draw an edge between i and j. We denote by p the number of connected components formed by these edges. We denote by l the number of active vertices with at least one edge. Using these notations, the cross-terms are bounded as follows.

Lemma 3.2. *If $f_0 \in L^\infty$, then $\mathbb{E}\left[\Pi_{(i,j)\in C_2^n} |k_n(X_i, X_j)|^{r_{i,j}}\right] = O(h_n^{-R+l-p})$.*

For any $k \in \{1, 2, \ldots, n\}$, we denote by $s_k \geq 0$ some non-negative integers. It is possible to bound the cross-terms $\mathbb{E}\left[\Pi_{(i,j)\in C_2^n} |k_n(X_i, X_j)|^{r_{i,j}} \Pi_{k=1}^n |k_n(X_k)|^{s_k}\right]$ directly from the previous lemma by noting that the first-order projections $|k_n(X_k)|^{s_k}$ only act as weights. This leads to the exact same bound which is collected in the next lemma.

Lemma 3.3. *If $f_0 \in L^\infty$, then $\mathbb{E}\left[\Pi_{(i,j)\in C_2^n} |k_n(X_i, X_j)|^{r_{i,j}} \Pi_{k=1}^n |k_n(X_k)|^{s_k}\right] = O(h_n^{-R+l-p})$.*

From these results, we can directly bound any moment of U_n and other statistics based on $k_n(\cdot,\cdot)$ and powers of it thereof. For ease of reference, we collect a direct application of these bounds to the summands of L_n and W_n respectively defined for any $i, j \in \{1, 2, \ldots, n\}$ with $i \neq j$ by

$$\begin{aligned} l(X_i) &:= k_n(X_i) - \mathbb{E}[U_n] , \\ w(X_i, X_j) &:= k_n(X_i, X_j) - k_n(X_i) - k_n(X_j) + \mathbb{E}[U_n] . \end{aligned}$$

The bounds are immediate applications of the previous results and collected in the lemma below.

Lemma 3.4. *If $f_0 \in L^\infty$, then $\mathbb{E}[\Pi_{(i,j)\in C_2^n} |w(X_i, X_j)|^{r_{i,j}} \Pi_{k=1}^n |l(X_k)|^{s_k}] = O(h_n^{-R+l-p})$.*

Remark 3.1. The hypothesis $f_0 \in L^\infty$ can be relaxed to finite integrability conditions on f_0 by using Hölder's inequality. This leads to looser bounds for the cross-terms involving the first-order and second-order projections, but these bounds are still tight enough to obtain all subsequent results in the paper. For simplicity, we follow Giné and Nickl (2008) and directly assume that $f_0 \in L^\infty$.

3.2 Normal approximations for bounded densities

The asymptotic normality of U_n in Giné and Nickl (2008) was obtained under the condition that the linear term in the Hoeffding decomposition of U_n dominates asymptotically. We show that the asymptotic normality of U_n still holds when the quadratic term dominates asymptotically. We obtain this result only using the fact that $f_0 \in L^\infty$. We leverage the general moment bounds of Section 3.1 to verify the conditions of the central limit theorem for generalized quadratic forms of de Jong (1987). This generalization opens the way for rate-optimal uncertainty quantification in U_n with $f_0 \in B^s_{2,\infty}$ even when $s \in (0, 1/4]$ as made clear in the next section. The proofs of all results in this section are collected in Section A.2 of the Supplementary Material.

Proposition 3.1. *Let $f_0 \in L^\infty$. If $n^2 h_n \to \infty$, then*

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

As an intermediate step before inference, we can make explicit the dependence of the normal limits in Proposition 3.1 on the choice of the bandwidth sequence h_n, or equivalently, on the dominating term in the Hoeffding decomposition of U_n. These results will become of more direct interest in the next section when the choice of the bandwidth sequence is constrained by the bias. To state the normal limits in each case, let us define

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

and simply write $\sigma_L^2(f_0) = \sigma_L^2$ and $\sigma_W^2(f_0) = \sigma_W^2$ when there is no risk of confusion. Without loss of generality, we take $h_n = cn^\beta$ for $\beta \in (-2, 0)$ and some constant $c > 0$. It follows then from Proposition 3.1 that: if $\beta \in (-1, 0)$, then $\sqrt{n}\left(U_n - \mathbb{E}[U_n]\right) \rightsquigarrow Z_1 \sim N(0, 4\sigma_L^2)$; if $\beta = -1$, then $\sqrt{n}\left(U_n - \mathbb{E}[U_n]\right) \rightsquigarrow Z_2 \sim N(0, 4\sigma_L^2 + 2\sigma_W^2/c)$; and if $\beta \in (-2, -1)$, then $n^{1+\beta/2}\left(U_n - \mathbb{E}[U_n]\right) \rightsquigarrow Z_3 \sim N(0, 2\sigma_W^2/c)$. The rates of convergence are parametric when $\beta \in [-1, 0)$ and nonparametric when $\beta \in (-2, -1)$. The asymptotic variance depends neither on the kernel K nor on c when $\beta \in (-1, 0)$. It is also seen to be equal to the semiparametric information lower bound for the model in this case. This invariance does not hold when $\beta \in (-2, -1]$.

Remark 3.2. The result of Proposition 3.1 may be appropriately contrasted with a similar result obtained by Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016) for the series estimator of Laurent (1996) – see, in particular, Section 3.1 (pp. 3737-3738) and Proposition 4.2 (p. 3742). To obtain this result, the authors derived a quite general conditional central limit theorem based on contracting conditions for the kernels (of the U-statistics) – see Theorem 2.1 (p. 3737) – and proved the conditions of this theorem to be verified by a variety of kernels including the Fourier kernel used in Laurent (1996) and the convolution kernel used for U_n – see Proposition 4.3 (p. 3742). However, to verify these conditions, the authors effectively required the densities to be compactly supported. This compactness restriction prevents an application of their result under the assumption that $f_0 \in L^\infty$. Our proof is different and is based on applying the central limit theorem of de Jong (1987) which directly extends that of Hall (1984). This approach should be of independent interest for the theory of second-order U-statistics with n-dependent kernels. As far as we know, all previous applications of de Jong's theorem in density estimation leveraged local regularity assumptions on the density – see, for instance, Hall (1984), Hardle and Mammen (1993), or Cattaneo, Crump, and Jansson (2014b). Our derivation without local regularity assumptions suggests, in particular, that the central limit theorems of Hall (1984) or de Jong (1987) are more primitive tools than the results of Robins, Li, Tchetgen Tchetgen, and van der Vaart (2016) and can be used to derive (at least the unconditional version of) their Theorem 2.1. A formal statement of this fact is, nonetheless, beyond the object of this paper and is left for future research.

Our use of de Jong's theorem in the proof of Proposition 3.1 allows us to handle the corner case (when the linear term and the quadratic term are asymptotically equivalent) at no extra cost using the extension of de Jong (1987) in Eubank and Wang (1999). We can similarly access a number of results building on de Jong's theorem and notably Berry–Esseen bounds quantifying the rate of convergence – see, for instance, Döbler (2024). This can be used to prove that the normal convergence in Proposition 3.1 holds uniformly over closed balls in L^∞.

Proposition 3.2. *Let $M > 0$ and $\delta > 0$. Define $\mathcal{F}(M, \delta) = \{f \in L^\infty : \|f\|_\infty \leq M,\ \sigma_L^2(f) \wedge \sigma_W^2(f) \geq \delta\}$. If $n^2 h_n \to \infty$, then $U_n^N = (\operatorname{Var} U_n)^{-1/2}(U_n - \mathbb{E}[U_n])$ converges weakly to $Z \sim N(0, 1)$ uniformly over $\mathcal{F}(M, \delta)$, that is,*

$$\lim_{n\to\infty} \sup_{f \in \mathcal{F}(M,\delta)} \sup_{x \in \mathbb{R}} \left| P(U_n^N \leq x) - \Phi(x) \right| = 0.$$

The result above can be refined by characterizing the rates of convergence of the errors for the normal approximation, that is, the Berry–Esseen rates for U_n^N. The proof in Section A.2 of the Supplementary Material already provides the individual rates for the components L_n and W_n when properly rescaled. To obtain the limit above, we used a crude inequality allowing us to combine the errors for both terms. It is possible, however, to directly expand the cross terms in each regime and obtain that the normal approximation for U_n^N has errors of the order $O(n^{-1/2} \vee n^{-1}h_n^{-1/2})$. In particular, if $h_n \asymp n^{-2/(4s+1)}$, we recover the bound $O(n^{-1/2} \vee n^{-4/(4s+1)})$. As the density gets rougher, the normal approximation gets poorer with the now standard regime change at $s = 1/4$. The closer s is to 0, the slower the convergence to the normal limit.

3.3 Bias bounds for bounded densities in $B^s_{2,\infty}(\mathbb{R})$

To use the generalized normal limit of Proposition 3.1 for uncertainty quantification in U_n, the bias of U_n needs to be bounded. It is well known, however, that it is impossible to do so only assuming that $f_0 \in L^\infty$ – see, for instance, Ritov and Bickel (1990). A natural solution is to control the local regularity of the density. Giné and Nickl (2008) assumed that f_0 belonged to the Sobolev space $W^{2,s}$ with $s \in (0, 1/2)$ and proved that the bias of U_n scaled (uniformly) as h_n^{2s}. To be able to instantiate a density for which the bias scales (pointwise) as h_n^{2s}, we have to consider the slightly more general Besov spaces $B^s_{2,\infty}$. This leads to Proposition 3.3 whose proof is collected in Section A.3 of the Supplementary Material.

Proposition 3.3. *Let $s \in (0, 1/2)$. For any $B \geq 0$, there is a constant $C(B, K) > 0$ depending only on B and K such that for all densities $f \in L^\infty \cap B^s_{2,\infty}$ with $\|f\|_{B^s_{2,\infty}} \leq B$,*

$$\left| \mathbb{E}\left[U_n\right] - \int_{\mathbb{R}} f^2(x)\, dx \right| \leq C(B, K) h_n^{2s}.$$

Moreover, if the kernel function satisfies $K(u) \geq 0$ for all $u \in \mathbb{R}$, then there exists a density $\bar{f} \in L^\infty \cap B^s_{2,\infty}$ and a constant $C_f > 0$ such that for any $n \in \mathbb{N}$ large enough,

$$\left| \mathbb{E}\left[U_n\right] - \int_{\mathbb{R}} \bar{f}^2(x)\, dx \right| > C_f h_n^{2s}.$$

Remark 3.3. The first part of the proposition parallels the result in Giné and Nickl (2008) for $W^{2,s}$ and leads directly to a uniform bound over the ball in $B^s_{2,\infty}$ with radius B. The proof differs and directly leverages the definition of $B^s_{2,\infty}$. The second part of the proposition does not hold in $W^{2,s}$.

With the variance bound $\operatorname{Var} U_n = O(n^{-1} \vee n^{-2}h_n^{-1})$ of Section 2.2, Proposition 3.3 leads to the "elbow phenomenon" of Bickel and Ritov (1988): the bias becomes a binding constraint as soon as $s \in (0, 1/4]$. In this case, in order to get $(\operatorname{Var} U_n)^{-1/2}(U_n - \int_{\mathbb{R}} f_0^2(x)\, dx) = O_p(1)$, we necessarily have to select h_n such that $\operatorname{Var} U_n = O(n^{-2}h_n^{-1})$. Because the rates of convergence then depend on h_n, it is optimal to select $h_n \asymp n^{-2/(4s+1)}$. By Proposition 3.1 and Proposition 3.3, the next result follows. As shown in Section 5, these limits can serve as a natural basis for uncertainty quantification in U_n when s is known.

Corollary 3.4. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$ with $\|f_0\|_{B^s_{2,\infty}} \le B$ for some constant $B > 0$. Suppose that $h_n = cn^{-2/(4s+1)}$ for some constant $c > 0$.*

1. *If $s \in (1/4, 1/2)$, then*

$$\sqrt{n}\left(U_n - \int_{\mathbb{R}} f_0^2(x)\,dx\right) \rightsquigarrow Z_1 \sim N\big(0, 4\sigma_L^2\big)\,.$$

2. *If $s \in (0, 1/4]$, then*

$$n^{4s/(4s+1)}\Big(U_n - \mathbb{E}\left[U_n\right]\Big) \rightsquigarrow \begin{cases} Z_2 \sim N\big(0, 4\sigma_L^2 + 2\sigma_W^2/c\big) & \text{if } s = 1/4\,, \\ Z_3 \sim N\big(0, 2\sigma_W^2/c\big) & \text{if } s < 1/4\,, \end{cases}$$

and there is a constant $C(B, K) > 0$ that only depends on B and K such that

$$n^{4s/(4s+1)}\left|\mathbb{E}\left[U_n\right] - \int_{\mathbb{R}} f_0^2(x)\,dx\right| \le C(B, K)\,.$$

Remark 3.4. The limits in Corollary 3.4 are stated pointwise for simplicity. The uniformity of the bias bounds as made clear in Remark 3.3 and the uniformity of the general normal limits as obtained in Proposition 3.2 lead directly to uniform counterparts to the limits above. In particular, we have that $(\sqrt{n} \wedge n^{4s/(4s+1)})(U_n - \int_{\mathbb{R}} f_0^2(x)\,dx)$ is stochastically bounded uniformly over the intersection of closed balls in L^∞ and $B^s_{2,\infty}$ when $s \in (0, 1/2)$. The estimator U_n with $h_n \asymp n^{-2/(4s+1)}$ is thus rate-optimal over $B^s_{2,\infty}$ with $s \in (0, 1/2)$ as it attains the rate bounds of Bickel and Ritov (1988) (keeping the inclusions of Remark 2.1 in mind). We recover, moreover, that the estimator attains the semiparametric efficiency bound when $s \in (1/4, 1/2)$.

Remark 3.5. The problem is purely nonparametric as soon as $s \in (0, 1/4)$ with all the classical implications for optimality and inference. It is seen, in particular, from the construction of Birgé and Massart (1995) and the inclusion $S(s) \subset B^s_{2,\infty}$ that the Hellinger modulus of continuity for the functional $T(P_0) = \int_{\mathbb{R}} f_0^2(x)\,dx$ is $b(\varepsilon) = A\varepsilon^{8s/(4s+1)} + o(\varepsilon^{8s/(4s+1)})$ when $f \in L^\infty \cap B^s_{2,\infty}$. This implies, for instance, that the impossibility results of Liu and Brown (1993) apply when $s \in (0, 1/4)$, so that there is no estimator that can be (locally uniformly) asymptotically unbiased at the rates $n^{4s/(4s+1)}$ with finite asymptotic variance. In other words, when $s \in (0, 1/4)$, uniform inference at the optimal rates $n^{4s/(4s+1)}$ necessarily involves dealing with a non-zero asymptotic bias. The case $s = 1/4$ is pathological and of independent interest for semiparametric theory, but its complete analysis is beyond the scope of this paper and is deferred to another manuscript.

4 Variance estimation: finite-sample bias and rate-optimality

If we know the value of $s \in (0, 1/2)$, we can estimate $\int_{\mathbb{R}} f_0^2(x)\,dx$ optimally using U_n with any bandwidth sequence $h_n \asymp n^{-2/(4s+1)}$. To assess the accuracy of the point estimate provided by U_n, it is natural to look for an estimator of the variance $\operatorname{Var} U_n$ of U_n. Provided these variance estimators are consistent, they can also be used in combination with the limits and bounds of

Section 3 to construct asymptotically valid Wald-type confidence intervals for $\int_{\mathbb{R}} f_0^2(x)\,dx$. The estimation of $\operatorname{Var} U_n$ presents two main challenges: the problem is semiparametric in nature in the sense that the rescaled limits of $\operatorname{Var} U_n$ are integral transforms of the unknown f_0 as seen from the values of σ_L^2 and σ_W^2; the discontinuous regime change at $s = 1/4$ implies two different finite-sample and limiting behaviors of U_n that any variance estimator has to adapt to even when s is known. The purpose of this section is two-fold. We first show in Section 4.1 that the problem can be optimally solved: we use U-statistic kernel-based estimators to build a variance estimator that is unbiased, consistent, and rate-optimal for any $s \in (0, 1/2)$ using the same bandwidth sequence $h_n \asymp n^{-2/(4s+1)}$ as for the optimal estimation of $\int_{\mathbb{R}} f_0^2(x)\,dx$ with U_n. We then contrast this optimal variance estimator with those obtained from automatic methods commonly used in practice for semiparametric inference. In Section 4.2, Section 4.3, and Section 4.4, we characterize the finite-sample bias, the consistency, and the rates of convergence of the variance estimators respectively obtained by the plug-in principle, the nonparametric bootstrap, and subsampling.

4.1 Rate-optimal unbiased variance estimator

Consider the following three U-statistics respectively given by

$$
\begin{aligned}
\upsilon_{1,n}(h_n) &:= \frac{1}{n(n-1)(n-2)} \sum_{i=1}^{n} \sum_{\substack{j=1\\ j\neq i}}^{n} \sum_{\substack{k=1\\ k\neq i,j}}^{n} k_n(X_i, X_j) k_n(X_i, X_k)\,, \\
\upsilon_{2,n}(h_n) &:= \frac{1}{n(n-1)(n-2)(n-3)} \sum_{i=1}^{n} \sum_{\substack{j=1\\ j\neq i}}^{n} \sum_{\substack{k=1\\ k\neq i,j}}^{n} \sum_{\substack{l=1\\ l\neq i,j,k}}^{n} k_n(X_i, X_j) k_n(X_k, X_l)\,, \\
\upsilon_{3,n}(h_n) &:= \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} k_n^2(X_i, X_j)\,.
\end{aligned}
$$

These U-statistics can be used to build estimators of $\operatorname{Var} k_n(X_1)$, $h_n \operatorname{Var} k_n(X_1, X_2)$, and $\operatorname{Var} U_n$ that satisfy strong optimality properties at a moderate computational cost. As estimators of $\operatorname{Var} k_n(X_1)$ and $h_n \operatorname{Var} k_n(X_1, X_2)$, we define, respectively,

$$
\begin{aligned}
\upsilon_{L,n}(h_n) &:= \upsilon_{1,n}(h_n) - \upsilon_{2,n}(h_n)\,, \\
\upsilon_{W,n}(h_n) &:= h_n(\upsilon_{3,n}(h_n) - \upsilon_{2,n}(h_n))\,,
\end{aligned}
$$

and, as an estimator of $\operatorname{Var} U_n$, we define

$$
\begin{aligned}
V_{U,n}(h_n) :&= \frac{4(n-2)}{n(n-1)} \upsilon_{L,n}(h_n) + \frac{2}{n(n-1)} \frac{1}{h_n} \upsilon_{W,n}(h_n) \\
&= \frac{1}{n(n-1)} \Big[4(n-2)\upsilon_{1,n}(h_n) + 2\upsilon_{3,n}(h_n) - (4n-6)\upsilon_{2,n}(h_n) \Big]\,.
\end{aligned}
$$

The estimators $\upsilon_{L,n}$ and $\upsilon_{W,n}$ are also estimators of σ_L^2 and σ_W^2. The following results, which are proved in Section A.4 of the Supplementary Material, collect the most important properties of

$\upsilon_{L,n}$, $\upsilon_{W,n}$, and $V_{U,n}$ for uncertainty quantification in U_n. It is seen, in particular, that $V_{U,n}$ is an unbiased and consistent estimator of $\operatorname{Var} U_n$ with the same bandwidth sequence $h_n \asymp n^{-2/(4s+1)}$ used for the optimal estimation of $\int_{\mathbb{R}} f_0^2(x)\,dx$ with U_n. Moreover, $V_{U,n}$ converges at the best possible rates to the limiting variances in each regime. Beyond these optimal properties, $V_{U,n}$ displays one major feature of practical interest: it is adaptive in the sense that it works both in the parametric regime and in the nonparametric regime without any modifications. Additional properties of the estimators can be found in Section A.4 of the Supplementary Material.

Proposition 4.1. *Let $f_0 \in L^\infty$. For any bandwidth sequence $h_n \to$ and any $n \geq 4$, it holds that*

$$\mathbb{E}\left[V_{U,n}(h_n)\right] = \operatorname{Var} U_n(h_n)\,.$$

Moreover, if $n^2 h_n \to 0$, then it holds that

$$V_{U,n}(h_n) = \frac{4}{n}\sigma_L^2 + \frac{2}{n^2 h_n}\sigma_W^2 + o_P\left(\frac{1}{n} \vee \frac{1}{n^2 h_n}\right).$$

Lemma 4.1. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$. If $h_n \asymp n^{-2/(4s+1)}$, then*

$$\upsilon_{L,n}(h_n) = \sigma_L^2 + O_P\left(n^{-1/2} \vee n^{(1-12s)/(8s+2)}\right).$$

Remark 4.1. A bandwidth $h_n \asymp n^{-2/(4s+1)}$ is not optimal for $\upsilon_{L,n}$. A choice of $h_n \asymp n^{-3/(4s+2)}$ yields the improved rates $\sqrt{n} \wedge n^{3s/(2s+1)}$. This improvement only bites when $s \in (0, 1/4)$ and is thus of second-order for our purpose as made clear in Proposition 4.2.

Lemma 4.2. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$. If $h_n \asymp n^{-2/(4s+1)}$, then*

$$\upsilon_{W,n}(h_n) = \sigma_W^2 + O_P\left(n^{-1/2} \vee n^{-4s/(4s+1)}\right).$$

Proposition 4.2. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$. If $h_n \asymp n^{-2/(4s+1)}$, then*

$$\frac{V_{U,n}(h_n)}{\operatorname{Var} U_n(h_n)} = 1 + O_P\left(n^{-1/2} \vee n^{-4s/(4s+1)}\right).$$

Remark 4.2. The third-order U-statistic $\upsilon_{1,n}$ is an estimator of $\int_{\mathbb{R}} f_0^3(x)\,dx$. Even with the optimal bandwidth $h_n \asymp n^{-3/(4s+2)}$, the estimator $\upsilon_{1,n}$ cannot reach the rates $n^{4s/(4s+1)}$ when $s \in (0, 1/4)$ in view of the cubic term in the variance that dominates when $nh_n \to 0$. In an elegant and surprising result, Kerkyacharian and Picard (1996) showed that the optimal rates for the estimation of $\int_{\mathbb{R}} f_0^3(x)\,dx$ are the same as for $\int_{\mathbb{R}} f_0^2(x)\,dx$. The ingenious solution of Kerkyacharian and Picard (1996) is based on a multiscale approach where, informally, the bias problem and the variance problem are tackled separately using different bandwidth sequences. This approach is not directly reproducible using kernel estimators. As made clear in Proposition 4.2, the effect of the cubic term of $\upsilon_{1,n}$ when $s \in (0, 1/4)$ does not affect the convergence rates of $V_{U,n}$ so that further refinements of $\upsilon_{1,n}$ should not be of direct concern to us in this problem.

Remark 4.3. The estimators $\upsilon_{L,n}$, $\upsilon_{W,n}$, and $V_{U,n}$ are unbiased, consistent, and rate-optimal. It is possible to sacrifice one (or several) of these properties to obtain some practical advantages. Indeed, these estimators suffer from a few drawbacks including the fact that they can take negative values (especially in very small samples) and that they are not trivial from a computational viewpoint (especially in very large samples). A solution is to consider appropriately biased estimators that help mitigate these issues. As an example, the estimator $\upsilon_{W,n}$ can be replaced by $U_n \int_{\mathbb{R}} K^2(u)\,du$ whose rate-optimality is direct from the properties of U_n and whose finite-sample bias is of second order when used in the construction of $V_{U,n}$. In practice, the computation of confidence intervals forces us to consider the biased estimator $\max(0, V_{U,n})$.

Remark 4.4. The estimator $V_{U,n}$ may also be contrasted with the estimator $\upsilon_{L,n}/n$. This estimator is also consistent for σ_L^2 and rate-optimal in the parametric regime $s \in (1/4, 1/2)$. The difference between the two estimators stems from their finite-sample bias: $V_{U,n}$ captures the quadratic term in $\operatorname{Var} U_n$ while $\upsilon_{L,n}/n$ completely ignores it. As one may expect, constructing Wald-type confidence intervals using $\upsilon_{L,n}/n$ leads to significant finite-sample undercoverage in the parametric regime $s \in (1/4, 1/2)$ that worsens the closer s is to $1/4$. The deficiency of $\upsilon_{L,n}/n$ is explicit and so this estimator is not included in the simulation exercise of Section 5.

4.2 Plug-in variance estimator and correction by bandwidth decoupling

To obtain the optimal variance estimator $V_{U,n}$, we leveraged the particular structure of the problem using higher-order corrections in a similar way to how U_n was constructed. In practice, it is not uncommon for variance estimators to be obtained in a more direct fashion (and so even when corrections were applied to the original estimator). A natural approach is given by the plug-in principle. For our problem, the idea translates into plugging the empirical measure in the moments $\operatorname{Var}(k_n(X_1))$ and $h_n \operatorname{Var}(k_n(X_1, X_2))$. This leads to estimators of $\operatorname{Var}(k_n(X_1))$ and $h_n \operatorname{Var}(k_n(X_1, X_2))$ respectively given by

$$s_{L,n}^2(h_n) := \frac{1}{n}\sum_{i=1}^n \Big(\frac{1}{n}\sum_{j=1}^n k_n(X_i, X_j)\Big)^2 - \Big(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n k_n(X_i, X_j)\Big)^2,$$

$$s_{W,n}^2(h_n) := h_n\left[\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n k_n^2(X_i, X_j) - \Big(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n k_n(X_i, X_j)\Big)^2\right].$$

Naturally, these statistics are also estimators of σ_L^2 and σ_W^2, respectively. If we plug these estimators in the variance formula for $\operatorname{Var} U_n$, we directly obtain a plug-in variance estimator given by

$$S_{U,n}^2(h_n, h_n) := \frac{4(n-2)}{n(n-1)} s_{L,n}^2(h_n) + \frac{2}{n(n-1)}\frac{1}{h_n} s_{W,n}^2(h_n).$$

We disassociate the bandwidth sequence used for the estimation of $s_{W,n}^2$ from the one used in the reconstruction of $S_{U,n}^2$ as decoupling them allows us to improve the original plug-in variance estimator in an astute way. This leads to the decoupled plug-in variance estimator given by

$$S^2_{U,n}(h_n, h'_n) := \frac{4(n-2)}{n(n-1)} s^2_{L,n}(h'_n) + \frac{2}{n(n-1)} \frac{1}{h_n} s^2_{W,n}(h'_n) \, .$$

The following results, which are all proved in Section A.5 of the Supplementary Material, make clear the properties of $s^2_{L,n}(h_n)$, $s^2_{W,n}(h_n)$, $S^2_{U,n}(h_n, h_n)$, and $S^2_{U,n}(h_n, h'_n)$ for uncertainty quantification in U_n. Proposition 4.3 characterizes the asymptotic bias and the probability limit of the plug-in variance estimator $S^2_{U,n}(h_n, h_n)$ without decoupling. It implies, in particular, that if $h_n \asymp n^{-2/(4s+1)}$, the estimator is consistent when $s \in (1/4, 1/2)$ but inconsistent when $s \in (0, 1/4]$. The inconsistency is not restricted to the choice of $h_n \asymp n^{-2/(4s+1)}$, but obtains as soon as $nh_n \not\to \infty$. The individual statistics $s^2_{L,n}(h_n)$ and $s^2_{W,n}(h_n)$ can still be used to consistently estimate σ^2_L and σ^2_W in each regime (for different choices of h_n in each case), but cannot be combined without decoupling the bandwidth sequences to consistently estimate $\operatorname{Var} U_n$. These pathologies can be exactly traced back to the diagonal elements generated by plugging in the empirical measure as made clear in the expansions of $s^2_{L,n}(h_n)$ and $s^2_{W,n}(h_n)$ in terms of the optimal U-statistics $\upsilon_{1,n}(h_n)$ and $\upsilon_{3,n}(h_n)$ obtained in Section A.5 of the Supplementary Material. Proposition 4.4 shows that decoupling the bandwidth sequences allows us to bypass these limitations by restoring consistency and adaptivity to both regimes. This solution, however, does not match the optimal variance estimator of Section 4.1. First, the rates of convergence are upper bounded by the best achievable rates for $s^2_{L,n}(h_n)$ and $s^2_{W,n}(h_n)$, namely $n^{2s/(2s+1)}$ instead of the much faster $n^{1/2} \wedge n^{4s/(4s+1)}$ for $V_{U,n}(h_n)$. Moreover, the estimator $S^2_{U,n}(h_n, h'_n)$ still suffers from consequential finite-sample bias as implied by the exact finite-sample bias characterization of $s^2_{L,n}$ and $s^2_{W,n}$ obtained in Lemma A.8 of the Supplementary Material. Finally, the optimal bandwidth to restore consistency and achieve the upper bound rates is different from the bandwidth used for optimal estimation with U_n.

Proposition 4.3. *Let $f_0 \in L^\infty$. For any bandwidth sequence $h_n \to 0$, it holds that*

$$\mathbb{E}\,[S^2_{U,n}(h_n, h_n)] - \operatorname{Var} U_n(h_n) = (1 + o(1)) \left[\frac{4}{n^2 h_n} \sigma^2_W + \frac{2}{n^3 h_n^2} K^2(0) \right] .$$

Moreover, if $n^2 h_n \to \infty$, then it holds that

$$S^2_{U,n}(h_n, h_n) = \frac{4}{n} \sigma^2_L + \frac{6}{n^2 h_n} \sigma^2_W + \frac{2}{n^3 h_n^2} K^2(0) + o_p\left(\frac{1}{n} \vee \frac{1}{n^2 h_n} \vee \frac{1}{n^3 h_n^2} \right) .$$

Remark 4.5. The asymptotic equivalence for the bias of $S^2_{U,n}(h_n, h_n)$ in Proposition 4.3 makes explicit the leading factors that drive its behavior in sufficiently large samples. The equivalence is sufficient to illustrate the inflating effects of the diagonal terms that were purged in the construction of $V_{U,n}$. Exact expansions of the finite-sample biases of $s^2_{L,n}$, $s^2_{W,n}$, and $S^2_{U,n}(h_n, h_n)$ can be obtained. These expansions capture more precisely the finite-sample behavior of the estimators, but their convoluted nature partly hides the dominating factors inflating the plug-in variance estimators in practice. These expansions are collected in Lemma A.8 of the Supplementary Material.

Lemma 4.3. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$. If $h_n \asymp n^{-1/(2s+1)}$, then*

$$s^2_{L,n}(h_n) = \sigma^2_L + O_P\left(n^{-2s/(2s+1)}\right).$$

Lemma 4.4. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$. If $h_n \asymp n^{-1/(2s+1)}$, then*

$$s^2_{W,n}(h_n) = \sigma^2_W + O_P\left(n^{-2s/(2s+1)}\right).$$

Remark 4.6. The bandwidth sequence $h_n \asymp n^{-1/(2s+1)}$ is optimal for both $s^2_{L,n}$ and $s^2_{W,n}$. It differs from the bandwidth sequence used for the optimal estimation of $\int_{\mathbb{R}} f_0^2(x)\,dx$ with U_n. If we use the bandwidth sequence $h_n \asymp n^{-2/(4s+1)}$ with $s^2_{L,n}$ and $s^2_{W,n}$, then both estimators fail to be consistent when $s \in (0, 1/4]$. While $s^2_{L,n}$ can still be theoretically used in its regime of application, $s^2_{W,n}$ cannot be used anymore. Moreover, even if $s^2_{L,n}$ remains consistent when $s \in (1/4, 1/2)$, its rates of convergence degrade to $n^{(4s-1)/(4s+1)}$ compared to the parametric rate achieved by $v_{L,n}$. These rates are pathologically slow the closer the regularity is to the regime threshold $s = 1/4$.

Proposition 4.4. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$. If $h_n \asymp n^{-2/(4s+1)}$ and $h'_n \asymp n^{-1/(2s+1)}$, then*

$$\frac{S^2_{U,n}(h_n, h'_n)}{\operatorname{Var} U_n(h_n)} = 1 + O_P\left(n^{-2s/(2s+1)}\right).$$

Remark 4.7. Bandwidth decoupling allows us to restore the consistency of the plug-in variance estimator in the nonparametric regime $s \in (0, 1/4)$. This correction comes, however, at the cost of hard-to-control finite-sample behavior that may plague inference in practice. This can be seen by computing the exact finite-sample bias of $S^2_{U,n}(h_n, h'_n)$ using Lemma A.8 in the Supplementary Material or more directly by noting that the asymptotic bias of $S^2_{U,n}(h_n, h'_n)$ includes a term, typically negative, of the order $O(h'^{2s}_n/n)$. By some (un)fortunate choice of (the constant factors in) h'_n, this term can inadvertently overcome the positive diagonal bias buffer created by $4\sigma^2_W/n^2 h'_n + 2K^2(0)/n^3 h_n h'_n$. A similar issue plagues the estimator $S^2_{U,n}(h'_n, h'_n)$ for the estimation of $\operatorname{Var} U_n(h_n)$ with $h_n = o(h'_n)$. The estimator can be used to improve the rates of convergence in the parametric regime $s \in (1/4, 1/2)$ to the upper bound $n^{2s/(2s+1)}$, but it remains inconsistent in the nonparametric regime $s \in (0, 1/4)$. Its finite-sample bias also displays a series of negative terms that may have detrimental consequences for inference in practice.

4.3 Nonparametric bootstrap variance estimator

Another standard approach for variance estimation is by way of the bootstrap and its variants. It is natural given the results on the plug-in variance estimators to expect a number of failures for the nonparametric bootstrap. To state this formally, we introduce some notations. Let $\mathcal{X}_n = \{X_1, X_2, \dots, X_n\}$ be the original i.i.d. sample and $\mathcal{X}^*_n = \{X^*_1, X^*_2, \dots, X^*_n\}$ another sample obtained by uniformly sampling n times from $\mathcal{X}_n$ with replacement. Denote by $P^*, \mathbb{E}^*, \operatorname{Var}^*, \operatorname{cov}^*$ the probability, expectation, variance, and covariance taken with respect to the sampling distribution conditional on $\mathcal{X}_n$. Consider the estimator

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

which is, conditional on $\mathcal{X}_n$, the same second-order U-statistic with n-dependent kernel k_n as U_n but computed over the sample $\mathcal{X}_n^*$. Conditional on $\mathcal{X}_n$, the sample $\mathcal{X}_n^*$ is i.i.d. with marginal the uniform discrete distribution on the non-random values $\{X_1, X_2, \ldots, X_n\}$. By the fact that $\mathcal{X}_n^*$ conditional on $\mathcal{X}_n$ is i.i.d., we directly obtain that

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

Since the marginal sampling distribution conditional on $\mathcal{X}_n$ is discrete uniform, we derive by direct computation that

$$\begin{aligned} \operatorname{Var}^*(k_n(X_1^*)) &= s_{L,n}^2(h_n) , \\ h_n \operatorname{Var}^*(k_n(X_1^*, X_2^*)) &= s_{W,n}^2(h_n) . \end{aligned}$$

From these results, the following equivalence holds

$$\operatorname{Var}^* U_n^*(h_n) = S_{U,n}^2(h_n, h_n) .$$

For the nonparametric bootstrap variance estimator, bandwidth decoupling is impossible: the variance estimator necessarily inherits the bandwidth sequence used for the computation of U_n^* in the bootstrap world. While it is possible to compute U_n^* with a different bandwidth h_n', the bootstrap variance estimator is then equal to $\operatorname{Var}^* U_n^*(h_n') = S_{U,n}^2(h_n', h_n')$.

By virtue of the equivalence with the uncorrected plug-in variance estimator, the nonparametric bootstrap variance estimator $\operatorname{Var}^* U_n^*(h_n)$ directly inherits all the pathologies of $S_{U,n}^2(h_n, h_n)$ gathered in Section 4.2 and in Section A.5 of the Supplementary Material. For the sake of exposition, we collect two results. Corollary 4.5, which follows immediately from Proposition 4.3, implies that $\operatorname{Var}^* U_n^*(h_n)$ with $h_n \asymp n^{-2/(4s+1)}$ is consistent when $s \in (1/4, 1/2)$ but inconsistent when $s \in (0, 1/4]$. The inconsistency in the nonparametric regime is not exclusive to the use of $h_n \asymp n^{-2/(4s+1)}$. In fact, the estimator completely fails to replicate the structure of $\operatorname{Var} U_n$ for any bandwidth that makes the quadratic term dominate asymptotically. Corollary 4.6, which is proved in Section A.6 of the Supplementary Material, makes explicit the rates of convergence and divergence of $\operatorname{Var}^* U_n^*(h_n)$ with $h_n \asymp n^{-2/(4s+1)}$ in the respective regimes. The estimator converges at the pathologically slow rates $n^{(4s-1)/(4s+1)}$ when $s \in (1/4, 1/2)$ and diverges at the exploding rates $n^{(1-4s)/(4s+1)}$ when $s \in (0, 1/4)$. The case $s = 1/4$ is tackled in Remark 4.8.

Corollary 4.5. *Let $f_0 \in L^\infty$. For any bandwidth sequence $h_n \to 0$, it holds that*

$$\mathbb{E}\left[\operatorname{Var}^* U_n^*(h_n)\right] - \operatorname{Var} U_n(h_n) = (1 + o(1)) \left[\frac{4}{n^2 h_n} \sigma_W^2 + \frac{2}{n^3 h_n^2} K^2(0) \right] .$$

Moreover, if $n^2 h_n \to \infty$, then it holds that

$$\operatorname{Var}^* U_n^*(h_n) = \frac{4}{n}\sigma_L^2 + \frac{6}{n^2 h_n}\sigma_W^2 + \frac{2}{n^3 h_n^2} K^2(0) + o_P\Big(\frac{1}{n} \vee \frac{1}{n^2 h_n} \vee \frac{1}{n^3 h_n^2}\Big)\,.$$

Corollary 4.6. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$. If $h_n \asymp n^{-2/(4s+1)}$, then*

$$\frac{\operatorname{Var}^* U_n^*(h_n)}{\operatorname{Var} U_n(h_n)} = \begin{cases} 1 + O_P\big(n^{-(4s-1)/(4s+1)}\big) & \text{if } s > 1/4\,, \\ O_P\big(n^{(1-4s)/(4s+1)}\big) & \text{if } s < 1/4\,. \end{cases}$$

Remark 4.8. The case $s = 1/4$ is omitted for readability. The rates are immaterial in this case and the probability limit can be exactly characterized at no cost from the first corollary. In particular, it holds that $n[\operatorname{Var}^* U_n^*(h_n) - \operatorname{Var} U_n(h_n)] = \frac{4}{c}\sigma_W^2 + \frac{2}{c^2}K^2(0) + o_P(1)$ where $nh_n \to c > 0$.

Remark 4.9. The rates of convergence of the bootstrap variance estimator can be improved by choosing any bandwidth $h'_n \asymp n^{-1/(2s+1)}$ leading to the estimator $\operatorname{Var}^* U_n^*(h'_n) = S^2_{U,n}(h'_n, h'_n)$. This improves the rates in the parametric regime $s \in (1/4, 1/2)$ from the pathologically slow $n^{(4s-1)/(4s+1)}$ to the suboptimal $n^{2s/(2s+1)}$. However, this choice does not solve the inconsistency issue in the nonparametric regime $s \in (0, 1/4)$. It also comes with hard-to-control finite-sample properties as made clear at the end of Section 4.2 for the estimator $S^2_{U,n}(h'_n, h'_n)$ with $h_n = o(h'_n)$.

Remark 4.10. The inconsistency of the nonparametric bootstrap variance in the nonparametric regime from Corollary 4.5 does not immediately imply that the standardized bootstrap statistic fails to converge in distribution to the same limit as the original statistic (in probability, conditional on the original data). Indeed, the standardized bootstrap statistic $n^2 h_n (U_n^* - U_n)^2$ is directly seen to be not uniformly integrable when $nh_n \to 0$. It can be shown, however, that the inconsistency of the nonparametric bootstrap variance in the nonparametric regime is sufficient to break the validity of any inferential method using the nonparametric bootstrap distribution in this regime. To see this, we note first that the bootstrap statistic is biased: we have $\mathbb{E}^*[U_n^*] = \frac{n-1}{n}U_n + \frac{K(0)}{nh_n}$. This was already reported in Cattaneo and Jansson (2022) and rightly associated with bootstrap strong inconsistency in the parametric regime as soon as $nh_n^2 \to 0$. This defect can be easily corrected by considering the debiased statistic $U_n^* - \frac{K(0)}{nh_n}$. This is enough to restore the strong consistency of the nonparametric bootstrap in the parametric regime even under $nh_n^2 \to 0$, but not sufficient to prevent its failure in the nonparametric regime. Indeed, the variance mismatch reported in Corollary 4.5 when $nh_n \to 0$ is sufficient in this case to prevent the debiased standardized statistic $n\sqrt{h_n}(U_n^* - \frac{K(0)}{nh_n} - U_n)$ from being tight under the bootstrap measure P^*. These results are formally proved in a separate manuscript where a seemingly novel bootstrap failure of independent interest is obtained for the studentized statistic $(\operatorname{Var}^* U_n^*)^{-1/2}(U_n^* - \frac{K(0)}{nh_n} - U_n)$.

4.4 Subsampling variance estimator

A natural solution to address the pathologies of the nonparametric bootstrap variance estimators is to consider subsampling without replacement. The idea is to obtain a sampling scheme that automatically purges some of the diagonal terms responsible for the deficiencies. As we show in this section, subsampling without replacement suppresses the constant and quadratic diagonal terms from resampling with the empirical measure but not the linear diagonal terms. These partial deletions are enough to obtain a consistent variance estimator in both the parametric and nonparametric regimes since the linear diagonal terms are asymptotically negligible under the standard assumption on the subsample size that $m/n \to 0$. Nonetheless, the deletions obtained by subsampling are not sufficient to restore unbiasedness in finite samples nor rate-optimality of the variance estimator. To show this, we introduce some notations. We define the set $\mathcal{X}_m^* = \{X_{i_1}^*, \ldots, X_{i_m}^*\}$ obtained by randomly sampling from $\mathcal{X}_n$ without replacement. We denote by $S = \{i_1, i_2, \ldots, i_m\}$ the corresponding set of indices randomly sampled from $\{1, \ldots, n\}$ without replacement. We define the subsampled statistic

$$U_S^*(h_n) = \frac{2}{m(m-1)} \sum_{\substack{i,j \in S \\ i<j}} k_n(X_i^*, X_j^*)$$

where the kernel k_n is unchanged and defined for the bandwidth sequence h_n. Conditional on $\mathcal{X}_n$, the estimator U_S^* is still a second-order U-statistic from the sample $\mathcal{X}_m^*$, but the sample $\mathcal{X}_m^*$ is not i.i.d. anymore. By virtue of the simple sampling scheme, the moments of U_S^* can still be computed exactly by combinatorial arguments – the computations have been carried out in the finite population literature – see, e.g., Section 2.5 in Lee (1990) or Bloznelis and Götze (2001). In particular, we directly have by a combinatorial calculation that $\mathbb{E}^*[U_S^*] = U_n$ and, using (2.6) in Bloznelis and Götze (2001), that

$$\operatorname{Var}^* U_S^*(h_n) = \frac{4}{m}\frac{n-m}{n-1}\psi_{L,n}^2(h_n) + \frac{2}{m(m-1)}\frac{(n-m)(n-m-1)}{(n-2)(n-3)}\psi_{W,n}^2(h_n)$$

where

$$\psi_{L,n}^2(h_n) = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{n-2}\sum_{j\neq i} k_n(X_i, X_j) - \frac{n-1}{n-2}U_n(h_n)\right]^2,$$

$$\psi_{W,n}^2(h_n) = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\Bigg[k_n(X_i, X_j) - \frac{1}{n-2}\sum_{k\neq i} k_n(X_i, X_k) - \frac{1}{n-2}\sum_{k\neq j} k_n(X_k, X_j) + \frac{n}{n-2}U_n(h_n)\Bigg]^2.$$

The terms $\psi_{L,n}^2(h_n)$ and $\psi_{W,n}^2(h_n)$ are the variances of the properly orthogonalized first-order and second-order terms in the finite-population Hoeffding decomposition of U_S^*. These terms, which only depend on the original sample $\mathcal{X}_n$, can be further decomposed in terms of $\upsilon_{3,n}$ and $s_{L,n}^2$ as proved in Lemma A.9 of the Supplementary Material. The results of Section 4.1 and Section 4.2 can then be directly leveraged to derive the most important properties of $\operatorname{Var}^* U_S^*(h_n)$ in view of uncertainty quantification. Proposition 4.7 characterizes the finite-sample bias of $\operatorname{Var}^* U_S^*(h_n)$ as well as its probability limit. It implies that if $h_n \asymp n^{-2/(4s+1)}$ and the subsample size is chosen so that $mh_n \to \infty$, then $\mathrm{Var}^* U_S^*(h_n)$ is consistent in the parametric regime $s \in (1/4, 1/2)$ when rescaled by m/n and consistent in the nonparametric regime $s \in (0, 1/4)$ when rescaled by m^2/n^2. In the nonparametric case, the condition that $m/n \to 0$ is sufficient to ensure $mh_n \to 0$ without any further constraint on the subsample size. By virtue of the different scaling in each regime, the subsampling variance estimator fails to be adaptive in the same way $V_{U,n}$ is – see also Remark 4.11 for the case $s = 1/4$. Proposition 4.8 makes explicit the rates of convergence of $\mathrm{Var}^* U_S^*(h_n)$ for the bandwidth sequence $h_n \asymp n^{-2/(4s+1)}$ and subsample size $m = n/\log(n)$. In the parametric regime $s \in (1/4, 1/2)$, the subsampling variance estimator converges at the pathologically slow rates $n^{4s/(4s+1)}$ with an additional $\log(n)$ penalty emerging from the choice of m. In the nonparametric regime $s \in (0, 1/4)$, the situation is markedly different: the rates $n^{(1-4s)/(4s+1)} \wedge n^{4s/(4s+1)}$ are optimal in the very-low-regularity regime $s \in (0, 1/8]$; moreover, the subsample size has no direct effect on the rates. The phase transition at $s = 1/8$ is surprising and seemingly novel. The proofs of both propositions are collected in Section A.7 of the Supplementary Material.

Proposition 4.7. *Let $f_0 \in L^\infty$. For any bandwidth $h_n \to 0$, any $m \geq 3$ and any $n \geq 3$, it holds that*

$$\begin{aligned}\mathbb{E}\left[\mathrm{Var}^*(U_S^*(h_n))\right] - \mathrm{Var}(U_n(h_n)) =& \left(\frac{4(m-2)}{m(m-1)} - \frac{8(n-2)}{n(n-1)}\right)\mathrm{Var}(k_n(X_1)) \\ &+ \left(\frac{2}{m(m-1)} - \frac{4}{n(n-1)}\right)\mathrm{Var}(k_n(X_1, X_2)).\end{aligned}$$

Moreover, if $m \to \infty$, $m/n \to 0$, and $n^2 h_n \to \infty$, then

$$\mathrm{Var}^*(U_S^*(h_n)) = \frac{4}{m}\sigma_L^2 + \frac{2}{m^2 h_n}\sigma_W^2 + o_P\left(\frac{1}{m} \vee \frac{1}{m^2 h_n}\right).$$

Proposition 4.8. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$. Suppose that $h_n \asymp n^{-2/(4s+1)}$ and $m = n/\log(n)$.*

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

$$\frac{1}{\log(n)} \frac{\mathrm{Var}^* U_S^*(h_n)}{\mathrm{Var}\, U_n(h_n)} = 1 + O_P\left(\log(n) n^{-(4s-1)/(4s+1)}\right).$$

2. *If $s < 1/4$, then*

$$\frac{1}{(\log(n))^2} \frac{\mathrm{Var}^* U_S^*(h_n)}{\mathrm{Var}\, U_n(h_n)} = 1 + O_P\left(n^{(4s-1)/(4s+1)} \vee n^{-4s/(4s+1)}\right).$$

Remark 4.11. The necessity to rescale the subsampling variance estimator $\mathrm{Var}^* U_S^*(h_n)$ with different factors in each regime for consistency prevents a direct application in the case $s = 1/4$. In fact, there is no factor that makes the subsampling variance estimator consistent when $s = 1/4$.

Remark 4.12. The subsample size $m = n/\log(n)$ satisfies $mh_n \to \infty$ when $h_n \asymp n^{-2/(4s+1)}$ and $s \in (1/4, 1/2)$. Other subsample sizes satisfying the conditions can be considered, leading to similar results provided the scaling m/n and m^2/n^2 and the rate-penalty in the parametric regime are modified accordingly. The condition $mh_n \to \infty$ is not trivial. For instance, a standard polynomial choice $m = n^\gamma$ with $\gamma \in (0, 1)$ leads to $mh_n \to \infty$ when $h_n \asymp n^{-2/(4s+1)}$ and $s \in (0, 1/2)$ if and only if $\gamma \in (2/(4s + 1), 1)$. This prevents, for instance, the standard choice $m = \sqrt{n}$.

Remark 4.13. The rates of convergence in the parametric regime $s \in (1/4, 1/2)$ can be improved from $n^{(4s-1)/(4s+1)}$ to $n^{2s/(2s+1)}$ by choosing $h'_n \asymp n^{-1/(2s+1)}$ for the computation of $\mathrm{Var}^* U_S^*(h'_n)$. However, this bandwidth sequence leads to an inconsistent estimator in the nonparametric regime as well as to hard-to-control finite-sample properties (as manifest from the addition of negative terms in the finite-sample bias of $\mathrm{Var}^* U_S^*(h'_n)$ compared to that of $\mathrm{Var}^* U_S^*(h_n)$).

Remark 4.14. The fact that the subsampling scheme purges the quadratic and constant diagonal terms of the statistics obtained from U_n should be enough to obtain the strong consistency of the subsampled statistic in both regimes (and so without debiasing since $\mathbb{E}^*[U_S^*] = U_n$). The result is beyond the scope of this paper and relegated to another manuscript.

5 Monte Carlo simulations in $B^s_{2,\infty}(\mathbb{R})$

5.1 Parametric and nonparametric Wald-type confidence intervals

When the level of smoothness $s \in (0, 1/2)$ of f_0 is known, the normal limits of Section 3 and the variance estimators of Section 4 can be combined to form Wald-type confidence intervals for $\int_{\mathbb{R}} f_0^2(x)\,dx$ based on U_n. The use of an optimal bandwidth $h_n \asymp n^{-2/(4s+1)}$ for U_n opens the way for constructing Wald-type confidence intervals that shrink at the optimal rates $\sqrt{n} \wedge n^{4s/(4s+1)}$. By virtue of the phase transition at $s = 1/4$, we shall consider two forms of Wald-type confidence intervals to ensure nominal coverage properties asymptotically. In the parametric regime $s \in (1/4, 1/2)$, the bias of U_n is guaranteed to be $o(n^{-1/2})$ and we shall directly consider parametric intervals of the form

$$\mathrm{CI}_n^{\mathrm{SP}}(\hat{V}_n) = \left[U_n - z_{1-\alpha/2}\hat{V}_n^{1/2}, U_n + z_{1-\alpha/2}\hat{V}_n^{1/2}\right],$$

where $\hat{V}_n$ is a placeholder for one of the variance estimators of Section 4 and $z_{1-\alpha/2}$ is the $(1-\alpha/2)$-quantile of the standard normal distribution. If $\hat{V}_n$ is negative, we shall consider $\max(0, \hat{V}_n)$ instead. In the nonparametric regime $s \in (0, 1/4)$, the bias of the estimator does not necessarily converge to zero at the optimal rates $n^{4s/(4s+1)}$ but is at worst bounded at these rates. As made clear in Remark 3.5, this limitation is not a defect of the optimal estimator U_n but an inherent feature of the nonparametric regime. The construction of Wald-type confidence intervals with asymptotic coverage guarantees in this case requires us to account for the bias. Several solutions exist, but they all require leveraging some a priori bounds on the asymptotic bias of U_n given that its estimation is impractical. For generality and simplicity, we shall directly expand the confidence intervals by subtracting an upper bound and a lower bound accounting for the asymptotic bias of U_n, that is,

$$\mathrm{CI}_n^{\mathrm{NP}}(\hat{V}_n, b_L, b^U) = \left[U_n - z_{1-\alpha/2}\hat{V}_n^{1/2} - n^{-4s/(4s+1)}b_U, U_n + z_{1-\alpha/2}\hat{V}_n^{1/2} - n^{-4s/(4s+1)}b^L\right],$$

where we assume that $b_L \leq n^{4s/(4s+1)}\left(\mathbb{E}\left[U_n\right] - \int_{\mathbb{R}} f_0^2(x)\,dx\right) \leq b^U$ for n large enough. If $\hat{V}_n$ is negative, we shall again consider $\max(0, \hat{V}_n)$. Provided the estimator $\hat{V}_n$ is consistent when $s \in (0, 1/4]$, the intervals $\mathrm{CI}_n^{\mathrm{NP}}(\hat{V}_n, b_L, b^U)$ are guaranteed to asymptotically cover $\int_{\mathbb{R}} f_0^2(x)\,dx$ at least at the nominal level. In simulations, we can control the true density f_0 so as to make the asymptotic bias of U_n practically negligible. In this case, we shall only consider the intervals $\mathrm{CI}_n^{\mathrm{SP}}(\hat{V}_n) = \mathrm{CI}_n^{\mathrm{NP}}(\hat{V}_n, 0, 0)$ without loss of generality. Heuristics for the choice of b_L and b^U as well as the possibility of improving on $\mathrm{CI}_n^{\mathrm{NP}}(\hat{V}_n, b_L, b^U)$ are important topics but fall beyond the scope of this paper. The goal of the simulations is to illustrate the effects of the variance estimators of Section 4 on Wald-type confidence intervals that are commonly considered in practice.

From an asymptotic viewpoint, the confidence intervals obtained from U_n with the optimal bandwidth $h_n \asymp n^{-2/(4s+1)}$ are indistinguishable provided the variance estimators are consistent in their respective regimes. The intervals are asymptotically valid by construction, shrink at the optimal rates $\sqrt{n} \wedge n^{4s/(4s+1)}$ by virtue of the optimality of U_n, and exhibit the same rates of convergence of their coverage error by virtue of the bias of U_n. In finite samples, however, the properties of the variance estimators obtained in Section 4 suggest the existence of significant differences with important consequences for uncertainty quantification that worsen the less regular the density f_0 is. To characterize these finite-sample consequences, we shall estimate by simulations the average length and average coverage of the confidence intervals obtained for each of the variance estimators of Section 4 for different levels of smoothness $s \in (0, 1/2)$ and different sample sizes $n \in \mathbb{N}$. We shall also use the simulations to illustrate the inconsistency of the plug-in and nonparametric bootstrap variance estimators in the nonparametric regime.

5.2 Monte Carlo simulations based on Weierstrass perturbations

For the simulations, we compute U_n using a standard Gaussian kernel and the bandwidth sequence $h_n = n^{-2/(4s+1)}$. We consider two designs each based on perturbing a regular probability density with a Weierstrass function. This allows us to control exactly the Besov regularity of the density by varying the amplitude of the cosine waves with respect to their frequencies. More precisely, we define for any $x \in \mathbb{R}$ and any $s \in (0, 1/2)$,

$$\bar{f}(x; s, f_\infty) := M f_\infty(x)\left(1 + m\sum_{k=0}^{\infty} 2^{-ks}\cos\left(2\pi\, 2^k x\right)\right)$$

where M is a normalizing constant to ensure the density integrates to 1, f_∞ is a regular bounded probability density function on $\mathbb{R}$, m is a scaling constant to ensure the density is strictly positive, and $W_s := \sum_{k=0}^{\infty} 2^{-ks}\cos(2\pi\, 2^k x)$ is a standard Weierstrass function. To guarantee positivity, we directly take $m = 0.5(\sum_{k=0}^{\infty} 2^{-ks})^{-1}$ for any $s \in (0, 1/2)$. This leads to the following density defined, for any $x \in \mathbb{R}$ and any $s \in (0, 1/2)$, by

$$f(x; s, f_\infty) := M f_\infty(x)\Big(1 + 0.5(1 - 2^{-s}) \sum_{k=0}^{\infty} 2^{-ks} \cos\left(2\pi\, 2^k x\right)\Big).$$

The same argument as in the proof of Proposition 3.3 in Section A.3 of the Supplementary Material shows that $f(\cdot; s, f_\infty)$ is a proper probability density such that $f(\cdot; s, f_\infty) \in L^\infty \cap B^s_{2,\infty}$ for any $s \in (0, 1/2)$. For the first design, we take f_∞ as the standard normal density. For the second design, we take f_∞ as a mixture of Laplace distributions. That is, we define for any $x \in \mathbb{R}$,

Design 1:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2};$$

Design 2:

$$f_{\text{Lap}}(x) = 0.2e^{-|x-2|} + 0.3e^{-|x+2|}.$$

For each design and each value of $s \in (0, 1/2)$, the normalizing constant M can be computed beyond standard machine precision by evaluating the first terms of the series. (In Design 1, the value of M is (almost) indistinguishable from 1.) For each design f_∞ and each $s \in (0, 1/2)$, we can similarly evaluate the true value of the parameter $\int_{\mathbb{R}} f^2(x; s, f_\infty)\, dx$ by expanding the square and evaluating the first terms of the corresponding series. The values are recorded (up to 0.0001 precision) in the following tables for $s \in \{0.05, 0.20, 0.30, 0.45\}$.

Table 1: Parameter value in the normal design

	$s = 0.05$	$s = 0.2$	$s = 0.3$	$s = 0.45$
$\int_{\mathbb{R}} f^2(x; s, \phi)\, dx$	0.2827	0.2845	0.2858	0.2876

Table 2: Parameter value in the mixture Laplace design

	$s = 0.05$	$s = 0.2$	$s = 0.3$	$s = 0.45$
$\int_{\mathbb{R}} f^2(x; s, f_{\text{Lap}})\, dx$	0.1417	0.1437	0.1450	0.1468

For each design, the bias of U_n is seen to be lower bounded by h_n^{2s} and some constant following the proof of Proposition 3.3 in Section A.3 of the Supplementary Material. Nonetheless, direct computations show that, when scaled by h_n^{-2s}, the bias is negligible compared to the magnitude of $\int_{\mathbb{R}} f^2(x; s, f_\infty)\, dx$. Hereafter, we shall consider without loss of generality the parametric intervals $\text{CI}_n^{\text{SP}}(\hat{V}_n) = \text{CI}_n^{\text{NP}}(\hat{V}_n, 0, 0)$ in both the parametric regime $s \in (1/4, 1/2)$ and in the nonparametric regime $s \in (0, 1/4)$. Designs where the asymptotic bias in the nonparametric regime is of the magnitude of the parameter can be conceived; these additional complexities are of secondary interest for the problem and are thus ignored in this paper.

To implement a sampler based on each design, we have to truncate the Weierstrass series to some fixed term $k_{\max}$. Using $k_{\max} = 50$ is enough to exceed standard machine precision in both designs. This leads to smooth approximations $f(x; s, f_\infty, k_{\max})$ that mimic the behavior of the original function beyond standard machine precision. We then build a sampler using these approximations based on rejection sampling. As a sanity check, we can recover from the sampler the true value of the parameter by evaluating the empirical average of $f(X; s, f_\infty, k_{\max})$ over a large enough sample obtained from the sampler (up to 0.001 precision for a sample of size 4, 000, 000). For illustration, we plot a graph of each density for $s \in \{0.05, 0.45\}$ as well as the respective empirical histogram obtained from rejection sampling. The graphs are collected in Section B of the Supplementary Material. For the simulations, we use the sampler to draw i.i.d. samples of size $n \in \{100, 500, 1000, 2000\}$ for each design and each value of $s \in \{0.05, 0.20, 0.30, 0.45\}$. Using Monte Carlo simulations over $N = 2,000$ runs, we compute the average length and average coverage of the intervals $\mathrm{CI}_n^{\mathrm{SP}}(\hat{V}_n)$ with $\alpha = 0.05$ where $\hat{V}_n$ is chosen within the list of (positive) variance estimators: optimal $\max(0, V_{U,n}(h_n))$, decoupled plug-in $S^2_{U,n}(h_n, h'_n)$, nonparametric bootstrap $\mathrm{Var}^* \, U_n^*(h_n)$, and subsampling $\mathrm{Var}^* U_S^*(h_n)$. For the hyper-parameters, we simply take $h_n = n^{-2/(4s+1)}$, $h'_n = n^{-1/(2s+1)}$, and $m = \lfloor n/\log(n) \rfloor$ without any further a posteriori refinements. The average coverage can be naturally improved for some fixed n for each interval by multiplying any of these hyper-parameters by some well-chosen constant. Improvements in this direction for some fixed n fall beyond the scope of this paper. The results of the simulations are collected in Table 3 and Table 4 in the Appendix.

Results from the Monte Carlo simulations

Table 3: Standard Normal ($N = 2,000$, $\alpha = 0.05$)

		Optimal Variance ($\max(0, V_{U,n})$)		Decoupled Plug-in ($S^2_{U,n}$)		Nonpar. Bootstrap ($\text{Var}^* U^*_n$)		Subsampling ($\text{Var}^* U^*_S$)	
Smoothness (s)	Sample Size (n)	Cov.	Length	Cov.	Length	Cov.	Length	Cov.	Length
$\mathbf{s = 0.45}$	$n = 100$	0.916	0.118	0.972	0.139	0.992	0.172	0.997	0.191
	$n = 500$	0.944	0.051	0.962	0.055	0.988	0.067	0.997	0.081
	$n = 1,000$	0.957	0.035	0.968	0.037	0.988	0.045	0.998	0.056
	$n = 2,000$	0.953	0.024	0.964	0.026	0.983	0.030	0.997	0.038
$\mathbf{s = 0.30}$	$n = 100$	0.914	0.152	0.985	0.189	1.0	0.273	1.0	0.287
	$n = 500$	0.932	0.067	0.968	0.076	0.998	0.115	0.999	0.137
	$n = 1,000$	0.946	0.046	0.970	0.052	1.0	0.079	1.0	0.099
	$n = 2,000$	0.949	0.032	0.964	0.035	0.998	0.054	1.0	0.071
$\mathbf{s = 0.20}$	$n = 100$	0.914	0.214	0.984	0.282	1.0	0.511	0.922	0.204
	$n = 500$	0.941	0.106	0.980	0.127	1.0	0.264	0.933	0.100
	$n = 1,000$	0.940	0.078	0.977	0.090	1.0	0.198	0.928	0.073
	$n = 2,000$	0.950	0.057	0.977	0.065	1.0	0.149	0.934	0.054
$\mathbf{s = 0.05}$	$n = 100$	0.810	0.660	0.992	1.112	1.0	4.925	0.810	0.662
	$n = 500$	0.876	0.523	0.992	0.814	1.0	6.303	0.875	0.520
	$n = 1,000$	0.892	0.476	0.993	0.713	1.0	7.046	0.890	0.472
	$n = 2,000$	0.915	0.430	0.992	0.625	1.0	7.887	0.916	0.426

Table 4: Laplace Mixture ($N = 2,000$, $\alpha = 0.05$)

		Optimal Variance ($\max(0, V_{U,n})$)		Decoupled Plug-in ($S^2_{U,n}$)		Nonpar. Bootstrap ($\text{Var}^* U^*_n$)		Subsampling ($\text{Var}^* U^*_S$)	
Smoothness (s)	Sample Size (n)	Cov.	Length	Cov.	Length	Cov.	Length	Cov.	Length
$\mathbf{s = 0.45}$	$n = 100$	0.916	0.085	0.978	0.104	0.993	0.130	0.995	0.138
	$n = 500$	0.935	0.037	0.965	0.041	0.991	0.050	0.997	0.058
	$n = 1,000$	0.946	0.025	0.964	0.027	0.984	0.033	0.998	0.040
	$n = 2,000$	0.959	0.018	0.967	0.019	0.985	0.022	0.999	0.027
$\mathbf{s = 0.30}$	$n = 100$	0.909	0.107	0.993	0.145	1.0	0.219	1.0	0.203
	$n = 500$	0.947	0.047	0.983	0.057	1.0	0.091	1.0	0.097
	$n = 1,000$	0.951	0.033	0.976	0.038	1.0	0.062	1.0	0.070
	$n = 2,000$	0.958	0.023	0.978	0.026	1.0	0.042	1.0	0.050
$\mathbf{s = 0.20}$	$n = 100$	0.894	0.149	0.996	0.226	1.0	0.445	0.897	0.144
	$n = 500$	0.944	0.075	0.989	0.099	1.0	0.234	0.932	0.071
	$n = 1,000$	0.943	0.055	0.985	0.070	1.0	0.176	0.930	0.052
	$n = 2,000$	0.947	0.041	0.985	0.049	1.0	0.133	0.932	0.038
$\mathbf{s = 0.05}$	$n = 100$	0.660	0.407	0.998	0.983	1.0	4.843	0.661	0.407
	$n = 500$	0.797	0.347	0.998	0.712	1.0	6.268	0.797	0.345
	$n = 1,000$	0.825	0.317	0.999	0.620	1.0	7.020	0.825	0.315
	$n = 2,000$	0.852	0.287	0.999	0.541	1.0	7.868	0.852	0.285

References

Andrews, D. W. (1994): "Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, pp. 43–72.

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.

Birgé, L., and P. Massart (1995): "Estimation of Integral Functionals of a Density," *The Annals of Statistics*, pp. 11–29.

Bloznelis, M., and F. Götze (2001): "Orthogonal Decomposition of Finite Population Statistics and Its Applications to Distributional Asymptotics," *The Annals of Statistics*, 29(3), 899–917.

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., 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.

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.

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.

de Jong, P. (1987): "A Central Limit Theorem for Generalized Quadratic Forms," *Probability Theory and Related Fields*, 75(2), 261–277.

Döbler, C. (2024): "The Berry–Esseen Bound in de Jong's CLT," *Statistics & Probability Letters*, 215, 110244.

Donoho, D. L., and M. Nussbaum (1990): "Minimax Quadratic Estimation of a Quadratic Functional," *Journal of Complexity*, 6(3), 290–323.

Efromovich, S., and M. Low (1996): "On Optimal Adaptive Estimation of a Quadratic Functional," *The Annals of Statistics*, 24(3), 1106–1125.

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 (2008): "A Simple Adaptive Estimator of the Integrated Square of a Density," *Bernoulli*, pp. 47–61.

——— (2021): *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press.

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.

HASMINSKII, R. Z., AND I. A. IBRAGIMOV (1979): "On the Nonparametric Estimation of Functionals," in *Proceedings of the Second Prague Symposium on Asymptotic Statistics*, vol. 473, pp. 474–482. North-Holland Amsterdam.

HOEFFDING, W. (1948): "A Class of Statistics with Asymptotically Normal Distribution," *The Annals of Mathematical Statistics*, pp. 293–325.

KERKYACHARIAN, G., AND D. PICARD (1996): "Estimating Nonquadratic Functionals of a Density Using Haar Wavelets," *The Annals of Statistics*, 24(2), 485–507.

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.

LEONENKO, N. N., L. PRONZATO, AND V. SAVANI (2008): "A Class of Renyi Information Estimators for Multidimensional Densities," *Annals of Statistics*, 36(5), 2153–2182.

LEPSKI, O. V., AND V. G. SPOKOINY (1997): "Optimal Pointwise Adaptive Methods in Nonparametric Estimation," *The Annals of Statistics*, 25(6), 2512–2546.

LEPSKII, O. (1991): "On a Problem of Adaptive Estimation in Gaussian White Noise," *Theory of Probability & Its Applications*, 35(3), 454–466.

LEVIT, B. Y. (1978): "Asymptotically Efficient Estimation of Nonlinear Functionals," *Problemy Peredachi Informatsii*, 14(3), 65–72.

LIU, R. C., AND L. D. BROWN (1993): "Nonexistence of Informative Unbiased Estimators in Singular Problems," *The Annals of Statistics*, pp. 1–13.

NEWEY, W. K. (1994): "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, pp. 1349–1382.

PFANZAGL, J. (1982): *Contributions to a General Asymptotic Statistical Theory*. Springer.

RITOV, Y., AND P. J. BICKEL (1990): "Achieving Information Bounds in Non and Semiparametric Models," *The Annals of statistics*, 18(2), 925–938.

ROBINS, J., L. LI, E. TCHETGEN, A. VAN DER VAART, ET AL. (2008): "Higher Order Influence Functions and Minimax Estimation of Nonlinear Functionals," in *Probability and statistics: essays in honor of David A. Freedman*, vol. 2, pp. 335–422. Institute of Mathematical Statistics.

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.

Supplementary Material for "Integrated Square of a Density: Semiparametric Inference Beyond the $o_p(n^{-1/4})$ Rule"

Paul Delatte

The supplementary material is composed of two sections: Section A contains most of the proofs of the results in the main text; Section B contains supporting material for the simulations. The notations used in the main text and in the supplementary material are (re)introduced below.

Notation. We denote by λ the Lebesgue measure on $\mathbb{R}$ and we write $\int_{\mathbb{R}} g\, d\lambda = \int_{\mathbb{R}} g(x)\, d\lambda(x) = \int_{\mathbb{R}} g(x)\, dx$ for the Lebesgue integral of a measurable function $g\colon \mathbb{R} \to \mathbb{R}$. For $p \in [1, \infty]$, we denote by $L^p = L^p(\mathbb{R}; \lambda)$ the Lebesgue spaces of measurable functions $g\colon \mathbb{R} \to \mathbb{R}$ such that $\|g\|_p < \infty$ where $\|g\|_p^p = \int_{\mathbb{R}} |g(x)|^p\, dx$ if $p \in [1, \infty)$ and $\|g\|_\infty = \inf\{C \geq 0 : |g(x)| \leq C \text{ for a.e. } x\}$. For $g \in L^1$, we define the Fourier transform by $F(g)(u) = \int_{\mathbb{R}} e^{-iux} g(x)\, dx$ and we extend it by continuity to L^2. We denote by $\rightsquigarrow$ weak convergence of random variables as $n \to \infty$. We use the Landau notations o, O as $n \to \infty$ and their stochastic variants o_p, O_p as $n \to \infty$. If $a_n = O(b_n)$ and $b_n = O(a_n)$, then we write $a_n \asymp b_n$. The density and the cumulative distribution function of a random variable $Z \sim N(0,1)$ are denoted by $\phi(\cdot)$ and $\Phi(\cdot)$, respectively. We make use of the following notational shortcut $K_h(\cdot) = K(\cdot/h_n)/h_n$ in the proofs.

A Proofs

A.1 Proofs of the moment bounds

Proof of Lemma 3.1. 1. By definition and the change of variables $u = (x-y)/h_n$, we have

$$\mathbb{E}\left[|k_n(X_i)|^r\right] = \int_{\mathbb{R}} \left| \int_{\mathbb{R}} K_h(x-y) f_0(y)\, dy \right|^r f_0(x)\, dx$$
$$= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} K(u) f_0(x - uh_n)\, du \right|^r f_0(x)\, dx.$$

Then using the fact that $f_0 \in L^\infty$, we have $|f_0(x - uh_n)| \leq \|f_0\|_\infty$ for all $x, u \in \mathbb{R}$. Since $K \in L^1$ and $\|f_0\|_1 = 1$, we directly get

$$\mathbb{E}\left[|k_n(X_i)|^r\right] \leq \|f_0\|_\infty^r \|K\|_1^r.$$

2. By definition, Fubini's theorem, and the change of variables $u = (x-y)/h_n$, we have

$$\mathbb{E}\left[|k_n(X_i, X_j)|^r\right] = \int_{\mathbb{R}} \int_{\mathbb{R}} \left|K_h(x-y)\right|^r f_0(x) f_0(y)\, dx\, dy$$
$$= \frac{1}{h_n^{r-1}} \int_{\mathbb{R}} \int_{\mathbb{R}} \left|K(u)\right|^r f_0(x - uh_n) f_0(x)\, dx\, du.$$

By using again the bound $|f_0(x - uh_n)| \leq \|f_0\|_\infty$ for all $x, u \in \mathbb{R}$ as well as the fact that $K \in L^r$ and $\|f_0\|_1 = 1$, we get

$$\mathbb{E}\left[|k_n(X_i, X_j)|^r\right] \leq \frac{1}{h_n^{r-1}} \|f_0\|_\infty \|K\|_r^r.$$

□

Proof of Lemma 3.2. The proof is obtained by induction over l by taking $p = 1$ after noticing that the bounds factorize over p by virtue of independence. The initialization follows directly from the previous result. For the inductive step, we can simply verify the formula with $R = r + q$, $l = 3$, and $p = 1$ for $\mathbb{E}\left[|k_n(X_i, X_j)|^r |k_n(X_i, X_k)|^q\right]$ where $r, q \geq 1$ are any integers. For values of $l > 3$, the proof is identical using the fact that $K \in L^p$ for any $1 \leq p \leq \infty$. By definition, Fubini's theorem, and the changes of variables $u = (x - y)/h_n$ and $v = (x - z)/h_n$, we get

$$\begin{aligned}
\mathbb{E}\left[|k_n(X_i, X_j)|^r |k_n(X_i, X_k)|^q\right] &= \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} \left|K_h(x - y)\right|^r \left|K_h(x - z)\right|^q f_0(x) f_0(y) f_0(z) \, dx \, dy \, dz \\
&= \frac{1}{h_n^{r+q-2}} \int_{\mathbb{R}} \int_{\mathbb{R}} \int_{\mathbb{R}} \left|K(u)\right|^r \left|K(v)\right|^q f_0(x - uh_n) f_0(x - vh_n) f_0(x) \, dx \, du \, dv.
\end{aligned}$$

By using again the bound $|f_0(x - uh_n)| \leq \|f_0\|_\infty$ for all $x, u \in \mathbb{R}$ as well as the fact that $K \in L^{r+q}$ and $\|f_0\|_1 = 1$, we get

$$\mathbb{E}\left[|k_n(X_i, X_j)|^r |k_n(X_i, X_k)|^q\right] \leq \frac{1}{h_n^{r+q-2}} \|f_0\|_\infty^2 \|K\|_r^r \|K\|_q^q.$$

This is enough to conclude the proof. □

Proof of Lemma 3.3. To obtain the result, it suffices to notice that the terms $|k_n(X_i)|$ only work as weights in the previous integrals. To show this, we can look at the case $\mathbb{E}\left[|k_n(X_i, X_j)|^r |k_n(X_i)|\right]$. From the proof of Lemma 3.1, we already know that for any $x \in \mathbb{R}$,

$$\left| \int_{\mathbb{R}} K_h(x - z) f_0(z) \, dz \right| \leq \|f_0\|_\infty \|K\|_1.$$

It follows then that

$$\begin{aligned}
\mathbb{E}\left[|k_n(X_i, X_j)|^r |k_n(X_i)|\right] &= \int_{\mathbb{R}} \int_{\mathbb{R}} \left|K_h(x - y)\right|^r \left| \int_{\mathbb{R}} K_h(x - z) f_0(z) \, dz \right| f_0(x) f_0(y) \, dx \, dy \\
&\leq \|f_0\|_\infty \|K\|_1 \mathbb{E}\left[|k_n(X_i, X_j)|^r\right],
\end{aligned}$$

and this is enough to conclude the proof from Lemma 3.2. □

Proof of Lemma 3.4. The proof follows immediately from Lemma 3.3 and the multinomial theorem. □

A.2 Proofs for the normal approximations

We break the proof of Proposition 3.1 into three parts: we first obtain closed-form expressions for the limiting variances; we then obtain the joint asymptotic normality of L_n and W_n when properly rescaled; we finally apply this result to obtain the asymptotic normality of $(\operatorname{Var} U_n)^{-1/2}(U_n - \mathbb{E}[U_n])$ under the general condition that $n^2 h_n \to \infty$.

Lemma A.1. *If $f_0 \in L^\infty$, then:*

1.

$$\lim_{n\to\infty} \operatorname{Var} k_n(X_1) = \int_{\mathbb{R}} f_0^3(x)\, dx - \left(\int_{\mathbb{R}} f_0^2(x)\, dx\right)^2 =: \sigma_L^2\,;$$

2.

$$\lim_{n\to\infty} h_n \operatorname{Var} k_n(X_1, X_2) = \int_{\mathbb{R}} f_0^2(x)\, dx \int_{\mathbb{R}} K^2(u)\, du =: \sigma_W^2.$$

Proof. 1. By definition and the change of variables $u = x - y$, we have that

$$\mathbb{E}[k_n(X_1)] = \int_{\mathbb{R}}\int_{\mathbb{R}} K_h(u) f_0(x-u) f_0(x)\, dx\, du$$

and

$$\mathbb{E}[k_n(X_1)^2] = \int_{\mathbb{R}}\left(\int_{\mathbb{R}} K_h(u) f_0(x-u)\, du\right)^2 f_0(x)\, dx.$$

Since $K \in L^1$ and $f_0 \in L^1 \cap L^2$, 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^1 and L^2 to f_0 as $h_n \to 0$. Then

$$\int_{\mathbb{R}}\left|\int_{\mathbb{R}} K_h(u) f_0(x-u)\, du - f_0(x)\right| f_0(x)\, dx \le \|f_0\|_\infty \int_{\mathbb{R}}\left|\int_{\mathbb{R}} K_h(u) f_0(x-u)\, du - f_0(x)\right| dx$$

and we directly obtain that

$$\lim_{n\to\infty} \mathbb{E}[k_n(X_1)]^2 = \left(\int_{\mathbb{R}} f_0^2(x)\, dx\right)^2.$$

Similarly, we have

$$\int_{\mathbb{R}}\left|\int_{\mathbb{R}} K_h(u) f_0(x-u)\, du - f_0(x)\right|^2 f_0(x)\, dx \le \|f_0\|_\infty \int_{\mathbb{R}}\left|\int_{\mathbb{R}} K_h(u) f_0(x-u)\, du - f_0(x)\right|^2 dx$$

and we directly obtain that

$$\lim_{n\to\infty} \mathbb{E}[k_n(X_1)^2] = \int_{\mathbb{R}} f_0^3(x)\, dx.$$

2. From (1), we directly have that $\lim_{n\to\infty} h_n \mathbb{E}[k_n(X_1, X_2)]^2 = 0$. Now, by definition and the change of variables $u = x - y$, we have that

$$h_n \mathbb{E}[k_n(X_1, X_2)^2] = \int_{\mathbb{R}}\int_{\mathbb{R}} h_n K_h^2(u) f_0(x-u) f_0(x)\, dx\, du.$$

Since $K^2 \in L^1$ and $f_0 \in L^1$ and $h_n K_h^2(\cdot) = K^2(\cdot/h_n)/h_n$, we have by the mollification theorem for K^2 that $\int_{\mathbb{R}} h_n K_h^2(u) f_0(x-u)\, du$ converges in L^1 to $\int K^2(u)\, du\, f_0(x)$. Then

$$\int_{\mathbb{R}} \left| \int_{\mathbb{R}} h_n K_h^2(u) f_0(x-u)\, du - \int_{\mathbb{R}} K^2(u)\, du\, f_0(x) \right| f_0(x)\, dx \\ \leq \|f_0\|_\infty \int_{\mathbb{R}} \left| \int_{\mathbb{R}} K_h^2(u) f_0(x-u)\, du - \int_{\mathbb{R}} K^2(u)\, du\, f_0(x) \right| dx$$

and we directly obtain that

$$\lim_{n\to\infty} h_n \mathbb{E}\left[k_n(X_1, X_2)^2\right] = \int_{\mathbb{R}} K^2(u)\, du \int_{\mathbb{R}} f_0^2(x)\, dx.$$

□

Lemma A.2. *If $f_0 \in L^\infty$ and $n^2 h_n \to \infty$, then*

$$\begin{pmatrix} \sqrt{n} L_n \\ \frac{1}{\sqrt{2}} \sqrt{n(n-1)h_n} W_n \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_L^2 & 0 \\ 0 & \sigma_W^2 \end{pmatrix} \right).$$

Proof of Lemma A.2. To prove this result, we verify the conditions for the central limit theorem in Eubank and Wang (1999) which extends the central limit theorem of de Jong (1987) to also handle the case where the linear term and the quadratic term may be equivalent. To this end, let us define

$$L(n) = \sqrt{n} L_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(X_i)\,,$$

$$W(n) = \frac{1}{\sqrt{2}} \sqrt{n(n-1)h_n} W_n = \frac{\sqrt{2h_n}}{\sqrt{n(n-1)}} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w(X_i, X_j)\,,$$

where l and w are defined in Lemma 3.4. Since $\operatorname{Var}(L(n) + W(n)) = O(1)$, the normalizing variances in all the Lyapunov-type conditions can be taken to be 1. Conditions (1.3) to (1.6) in Eubank and Wang (1999) then rewrite as

$$\frac{2}{n(n-1)} h_n \max_{1\leq i\leq n} \sum_{j=1}^{n} \operatorname{Var}(w(X_i, X_j)) \to 0\,, \tag{EW1.3}$$

$$\mathbb{E}\left[W(n)^4\right] / (\operatorname{Var} W(n))^2 \to 3\,, \tag{EW1.4}$$

$$\frac{1}{n^2} \sum_{i=1}^{n} \mathbb{E}\left[l(X_i)^4\right] \to 0\,, \tag{EW1.5}$$

$$\frac{2}{n(n-1)} n^{-1} h_n \mathbb{E}\left[\left(\sum_{i=2}^{n} \sum_{j=1}^{i-1} \mathbb{E}\left[w(X_i, X_j) l(X_i) | X_1, \ldots, X_{i-1}\right] \right)^2 \right] \to 0\,. \tag{EW1.6}$$

Since $(X_i)_{i=1,\dots,n}$ are i.i.d., the previous conditions respectively simplify to

$$n^{-1}h_n \mathrm{Var}(w(X_1,X_2)) \to 0, \tag{EW1.3bis}$$

$$\mathbb{E}\left[W(n)^4\right]/(\mathrm{Var}\, W(n))^2 \to 3, \tag{EW1.4}$$

$$n^{-1}\mathbb{E}\left[l(X_1)^4\right] \to 0, \tag{EW1.5bis}$$

$$n^{-1}h_n \mathrm{Var}\big(\mathbb{E}\left[w(X_2,X_1)l(X_2)|X_1\right]\big) \to 0. \tag{EW1.6bis}$$

To handle (EW1.4), we make use of the expansion of $\mathbb{E}\left[W(n)^4\right]$ 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 $G_{\mathrm{I}}, G_{\mathrm{II}}, G_{\mathrm{III}}, G_{\mathrm{IV}}$ tend to zero and the term G_{V} is asymptotically equivalent to $\mathrm{Var}(W(n))^2/2$. Since $(X_i)_{i=1,\dots,n}$ are i.i.d., this reduces to the following conditions

$$n^{-2}h_n^2\mathbb{E}\left[w(X_1,X_2)^4\right] \to 0 \tag{dJ.G_{I}}$$

$$n^{-1}h_n^2\mathbb{E}\left[w(X_1,X_2)^2 w(X_1,X_3)^2\right] \to 0 \tag{dJ.G_{II}}$$

$$n^{-1}h_n^2\mathbb{E}\left[w(X_1,X_2)^2 w(X_1,X_3)w(X_3,X_2)\right] \to 0 \tag{dJ.G_{III}}$$

$$h_n^2\mathbb{E}\left[w(X_1,X_2)w(X_1,X_3)w(X_4,X_2)w(X_4,X_3)\right] \to 0 \tag{dJ.G_{IV}}$$

$$h_n\mathbb{E}\left[w(X_1,X_2)^2\right]/\mathrm{Var}(W(n)) \to 1 \tag{dJ.G_{V}}$$

where the last equivalence follows from $3\binom{n}{4}\binom{n}{2}^{-2} = 1/2 + o(1)$.

We can now use Lemma 3.4 to verify that all the limits hold. For (EW1.3bis), we have $n^{-1}h_n\mathrm{Var}(w(X_1,X_2)) = O(n^{-1})$. For (dJ.$G_{\mathrm{I}}$), we have $n^{-2}h_n^2\mathbb{E}\left[w(X_1,X_2)^4\right] = O(h_n^{-1}n^{-2})$. For (dJ.$G_{\mathrm{II}}$), we have $n^{-1}h_n^2\mathbb{E}\left[w(X_1,X_2)^2w(X_1,X_3)^2\right] = O(n^{-1})$. For (dJ.$G_{\mathrm{III}}$), we have $n^{-1}h_n^2\mathbb{E}\left[w(X_1,X_2)^2w(X_1,X_3)w(X_3,X_2)\right] = O(n^{-1})$. For (dJ.$G_{\mathrm{IV}}$), we have $h_n^2\mathbb{E}\left[w(X_1,X_2)w(X_1,X_3)w(X_4,X_2)w(X_4,X_3)\right] = O(h_n)$. For (dJ.$G_{\mathrm{V}}$), the result follows immediately from the definition of $W(n)$ and Lemma A.1. For (EW1.5bis), we have $n^{-1}\mathbb{E}\left[l(X_1)^4\right] = O(n^{-1})$. For (EW1.6bis), we have by monotonicity, conditional Hölder's inequality, independence, and the tower property that $\mathrm{Var}(\mathbb{E}\left[w(X_2,X_1)l(X_2)|X_1\right]) \le \mathbb{E}\left[w(X_2,X_1)^2\right](\mathbb{E}\left[l(X_2)^2\right])$ and so $n^{-1}h_n\mathrm{Var}(\mathbb{E}\left[w(X_2,X_1)l(X_2)|X_1\right]) = O(n^{-1})$. All the conditions for the central limit theorem in Eubank and Wang (1999) are hence verified. This concludes the proof. □

Proof of Proposition 3.1. From Equation (2.2) and Lemma A.1, we have

$$\mathrm{Var}\, U_n = \frac{4}{n}\Big(\sigma_L^2 + o(1)\Big) + \frac{2}{n(n-1)}\frac{1}{h_n}\Big(\sigma_W^2 + o(1)\Big). \tag{A.1}$$

By distinguishing three cases if necessary, the result then follows immediately from Equation (2.1), Slutsky's theorem, Lemma A.2, and the normality of the marginals of bivariate normals. □

Proof of Proposition 3.2. To prove the result, we show that Berry–Esseen bounds hold uniformly for the linear term L_n and the quadratic term W_n, respectively. The result for U_n then follows by resorting to the standard inequality: for any $\varepsilon > 0$,

$$\sup_x \Big| P(X+Y \le x) - \Phi(x) \Big| \le \sup_x \Big| P(X \le x) - \Phi(x) \Big| + \frac{\varepsilon}{\sqrt{2\pi}} + P(|Y| > \varepsilon).$$

For the linear term, we can directly apply the Berry–Esseen theorem to the i.i.d. random variables $(k_n(X_i))_{i=1,\ldots,n}$. Indeed, we have $\mathbb{E}[k_n(X_1)] = 0$ and $\mathbb{E}[|k_n(X_1)|^3] \le \|f_0\|_\infty^3 \|K\|_1^3$ from Lemma 3.1. If we define $L^N(n) = \sqrt{n} L_n / \sigma_L$, then it holds that

$$\sup_{x \in \mathbb{R}} \Big| P(L^N(n) \le x) - \Phi(x) \Big| \le \frac{C \|f_0\|_\infty^3 \|K\|_1^3}{\sigma_L^3 \sqrt{n}}$$

and so

$$\sup_{f \in \mathcal{F}(M,\delta)} \sup_{x \in \mathbb{R}} \Big| P(L^N(n) \le x) - \Phi(x) \Big| \le \frac{C M^3 \|K\|_1^3}{\delta^3 \sqrt{n}}.$$

For the quadratic term, we use the Berry–Esseen bounds derived in Döbler (2024) and apply it to the properly normalized sequence $W^N(n) = (\sqrt{2}\sigma_W)^{-1} \sqrt{n(n-1)h_n} W_n$. From the expansion of $\mathbb{E}[W_n^4]$ in the proof of Lemma A.2, we directly obtain the bound

$$\Big| \mathbb{E}[W^N(n)^4] - 3 \Big| \le C^2 \Big(\frac{c_1(K)\|f_0\|_\infty}{n^2 h_n} + \frac{c_2(K)\|f_0\|_\infty^2}{n} + c_3(K)\|f_0\|_\infty^3 h_n \Big)$$

where $c_1(K), c_2(K), c_3(K)$ are constants that depend on the L^p-norms of K. Then from Corollary 2.2 in Döbler (2024), we get

$$\sup_{x \in \mathbb{R}} \Big| P(W^N(n) \le x) - \Phi(x) \Big| \le C \Big(\frac{c_1(K)\|f_0\|_\infty}{n^2 h_n} + \frac{c_2(K)\|f_0\|_\infty^2}{n} + c_3(K)\|f_0\|_\infty^3 h_n \Big)^{1/2} + \frac{38}{\sqrt{n}}$$

and so

$$\sup_{f \in \mathcal{F}(M,\delta)} \sup_{x \in \mathbb{R}} \Big| P(W^N(n) \le x) - \Phi(x) \Big| \le C \Big(\frac{c_1(K) M}{n^2 h_n} + \frac{c_2(K) M^2}{n} + c_3(K) M^3 h_n \Big)^{1/2} + \frac{38}{\sqrt{n}}.$$

□

A.3 Proof of the bias bound

Proof of Proposition 3.3. 1. By definition, the change of variables $u = (x-y)/h_n$, and the fact that $\int_{\mathbb{R}} K(u)\,du = 1$, we get

$$\begin{aligned} \mathbb{E}[U_n] - \int_{\mathbb{R}} f^2(x)\,dx &= \int_{\mathbb{R}} \int_{\mathbb{R}} K(u) f(x - u h_n) f(x)\,dx\,du - \int_{\mathbb{R}} f^2(x)\,dx \\ &= \int_{\mathbb{R}} K(u) \Big(\int_{\mathbb{R}} f(x - u h_n) f(x)\,dx - \int_{\mathbb{R}} f^2(x)\,dx \Big) du. \end{aligned}$$

We can notice that the inner term is $\langle f(\cdot - u h_n), f(\cdot) \rangle_2 - \|f(\cdot)\|_2^2$. Then by expanding the square and noting that $\|f(\cdot - u h_n)\|_2^2 = \|f(\cdot)\|_2^2$ by translation invariance, we get $\langle f(\cdot - u h_n), f(\cdot) \rangle_2 - \|f(\cdot)\|_2^2 =$

$-\frac{1}{2}\|f(\cdot) - f(\cdot - uh_n)\|_2^2$. It follows that

$$\mathbb{E}\left[U_n\right] - \int_{\mathbb{R}} f^2(x)\, dx = -\frac{1}{2}\int_{\mathbb{R}} K(u)\|f(\cdot) - f(\cdot - uh_n)\|_2^2\, du\,.$$

By the characterization of $B^s_{2,\infty}$ through the L^2-modulus of continuity, we directly get that $\|f(\cdot) - f(\cdot - uh_n)\|_2 \leq B|uh_n|^s$. It follows then that

$$\left|\mathbb{E}\left[U_n\right] - \int_{\mathbb{R}} f^2(x)\, dx\right| \leq \frac{B^2 h_n^{2s}}{2}\int_{\mathbb{R}} |u|^{2s}\left|K(u)\right| du.$$

This concludes the proof with $C(B, K) = \frac{1}{2}B^2 \int |u|^{2s}\left|K(u)\right| du < \infty$.

2. For some $m > 0$, define the function f for any $x \in \mathbb{R}$ by

$$f(x) = M\phi(x)\left(1 + m\sum_{k=1}^{\infty} 2^{-ks}\cos(2\pi\, 2^k x)\right),$$

where $M > 0$ is a normalizing constant and ϕ is the density of the standard normal distribution. The function f integrates to 1 and is bounded from standard properties of Weierstrass functions. In view of the boundedness of Weierstrass functions, the value m can be chosen so as to ensure the positivity of f. Since ϕ is infinitely smooth with fast decaying tails and the Weierstrass perturbation belongs to $B^s_{2,\infty}$, we directly have that $f \in B^s_{2,\infty}$. By the reverse triangle inequality and bounds on the translation errors for the Gaussian density and the Weierstrass function, we get that there exist $C(m, M)$ and $\delta > 0$ such that for all $|z| < \delta$,

$$\|f(\cdot - z) - f(\cdot)\|_2^2 \geq C(m, M)|z|^{2s}\,.$$

By the same expansion as used in the first part of the proof, we directly get

$$\left|\mathbb{E}\left[U_n\right] - \int_{\mathbb{R}} f^2(x)\, dx\right| \geq \frac{C(m, M)}{2} h_n^{2s} \int_{|u|<\delta/h_n} K(u)|u|^{2s}\, du.$$

From the non-negativity of the kernel, we have for n large enough that $\int_{|u|<\delta/h_n} K(u)|u|^{2s}\, du$ is bounded from below by some constant $C(K) > 0$. Taking $C_f = C(K)C(m, M)/2$ is enough to conclude the proof. □

A.4 Proofs for the optimal variance estimators

Lemma A.3. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$. Then there is a constant $C(f_0, K) > 0$ depending only on f_0 and K such that*

$$\left|\mathbb{E}\left[v_{1,n}\right] - \int_{\mathbb{R}} f_0^3(x)\, dx\right| \leq C(f_0, K)h_n^{2s}\,.$$

Moreover, it holds that

$$\upsilon_{1,n} = \int_{\mathbb{R}} f_0^3(x)\,dx + O_p\Big(h_n^{2s} \vee \frac{1}{\sqrt{n}} \vee \frac{1}{n^{3/2}h_n}\Big).$$

Proof of Lemma A.3. (*Part A*) We have the following decomposition

$$\begin{aligned}\mathbb{E}\left[\upsilon_{1,n}\right] - \int_{\mathbb{R}} f_0^3(x)\,dx &= \int_{\mathbb{R}} \big(K_h * f_0(x)\big)^2 f_0(x)\,dx - \int_{\mathbb{R}} f_0^3(x)\,dx \\ &= 2\int_{\mathbb{R}} f_0^2(x)\big(K_h * f_0(x) - f_0(x)\big)\,dx + \int_{\mathbb{R}} f_0(x)\big(K_h * f_0(x) - f_0(x)\big)^2\,dx\,,\end{aligned}$$

where $K_h * f_0(x) = \int_{\mathbb{R}} K_h(x-y)f_0(y)\,dy$. We analyze both terms successively.

For the second term, we have that

$$\begin{aligned}\Big|\int_{\mathbb{R}} f_0(x)\big(K_h * f_0(x) - f_0(x)\big)^2\,dx\Big| &\le \|f_0\|_\infty \big\|K_h * f_0(\cdot) - f_0(\cdot)\big\|_2^2 \\ &\le \|f_0\|_\infty \Big(\int_{\mathbb{R}} |K(u)|\big\|f_0(\cdot - uh_n) - f_0(\cdot)\big\|_2\,du\Big)^2 \\ &\le \|f_0\|_\infty B^2 h_n^{2s}\Big(\int |K(u)||u|^s\,du\Big)^2 \\ &= C_1(f_0, K)h_n^{2s}\end{aligned}$$

where the second inequality follows from Minkowski's integral inequality and the change of variables $u = (x-y)/h_n$, while the third inequality follows from the fact that $f_0 \in B^s_{2,\infty}$ by the same argument as in the proof of Proposition 3.3 (where $B \ge 0$ is any constant bounding the L^2 modulus of continuity of f_0).

For the first term, we have that

$$\begin{aligned}2\int_{\mathbb{R}} f_0^2(x)\big(K_h * f_0(x) - f_0(x)\big)\,dx &= \int_{\mathbb{R}} K(u)\int_{\mathbb{R}} f_0^2(x)\Big(f_0(x-uh_n) + f_0(x+uh_n) - 2f_0(x)\Big)dx\,du \\ &= -\int_{\mathbb{R}} K(u)\int_{\mathbb{R}}\Big(f_0^2(x-uh_n) - f_0^2(x)\Big)\Big(f_0(x-uh_n) - f_0(x)\Big)dx\,du\end{aligned}$$

where the first equality follows from the symmetry of the kernel and the changes of variables $u = (x-y)/h_n$ and $v = -u$, while the second equality follows from the change of variables $w = x + uh_n$ and a simple factorization equality. For any $u \in \mathbb{R}$, we have

$$\begin{aligned}\Big|\int_{\mathbb{R}}\Big(f_0^2(x-uh_n) - f_0^2(x)\Big)\Big(f_0(x-uh_n) - f_0(x)\Big)dx\Big| &\le \|f_0^2(\cdot - uh_n) - f_0^2(\cdot)\|_2\|f_0(\cdot - uh_n) - f_0(\cdot)\|_2 \\ &\le 2\|f_0\|_\infty\|f_0(\cdot - uh_n) - f_0(\cdot)\|_2^2\,,\end{aligned}$$

where the first inequality follows from Hölder's inequality, while the second inequality follows from a simple factorization equality and the fact that $f_0 \in L^\infty$. Indeed, it is seen that $\|f_0^2(\cdot - uh_n) - f_0^2(\cdot)\|_2 \le 2\|f_0\|_\infty\|f_0(\cdot - uh_n) - f_0(\cdot)\|_2$ from the simple factorization $f_0^2(x-uh_n) - f_0^2(x) =$

$(f_0(x-uh_n)-f_0(x))(f_0(x-uh_n)+f_0(x))$. Plugging the inequality in, we obtain

$$\begin{aligned}\left|2\int_{\mathbb{R}} f_0^2(x)\big(K_h * f_0(x)-f_0(x)\big)\,dx\right| &\le 2\|f_0\|_\infty \int_{\mathbb{R}} |K(u)|\|f_0(\cdot-uh_n)-f_0(\cdot)\|_2^2\,du\\ &\le 2\|f_0\|_\infty B^2 h_n^{2s}\int_{\mathbb{R}}|K(u)||u|^{2s}\,du\\ &= C_2(f_0,K)h_n^{2s}\,,\end{aligned}$$

where the second inequality follows again from the fact that $f_0 \in B^s_{2,\infty}$.

By taking $C(f_0,K)=C_1(f_0,K)+C_2(f_0,K)$, this concludes the proof.

(*Part B*) Given the previous bound on the bias of $\upsilon_{1,n}$, it suffices now to bound the variance of $\upsilon_{1,n}$. Since it is a third-order U-statistic based on an i.i.d. sample, it admits a simple Hoeffding decomposition which leads to a simple expression for its variance. We then simply need to bound $\mathbb{E}[k_n^S(X_1)^2]$, $\mathbb{E}[k_n^S(X_1,X_2)^2]$, and $\mathbb{E}[k_n^S(X_1,X_2,X_3)^2]$ where $k_n^S(X_1)=\mathbb{E}[k_n(X_1,X_2)k_n(X_1,X_3)|X_1]$, $k_n^S(X_1,X_2)=\mathbb{E}[k_n(X_1,X_2)k_n(X_1,X_3)|X_1,X_2]$ and $k_n^S(X_1,X_2,X_3)=k_n(X_1,X_2)k_n(X_1,X_3)$. By the properties of the conditional expectation, we can bound these terms directly from Lemma 3.3. Indeed, it is seen that

$$\begin{aligned}&\mathbb{E}[k_n^S(X_1)^2]=\mathbb{E}[k_n(X_1)^4]=O(1)\,,\\ &\mathbb{E}[k_n^S(X_1,X_2)^2]=\mathbb{E}[k_n(X_1,X_2)^2k_n(X_1)^2]=O(h_n^{-1})\,,\\ &\mathbb{E}[k_n^S(X_1,X_2,X_3)^2]=\mathbb{E}[k_n(X_1,X_2)^2k_n(X_1,X_3)^2]=O(h_n^{-2}).\end{aligned}$$

It follows directly from these bounds that $\operatorname{Var}\upsilon_{1,n}=O(n^{-1}\vee n^{-3}h_n^{-2})$ by noting that the variance of the quadratic term of order $n^{-2}h_n^{-1}$ never dominates (when distinguishing the cases $nh_n\to 0$ and $nh_n\to\infty$ if necessary). This is enough to obtain the result in combination with (*Part A*). □

Lemma A.4. *Let $s\in(0,1/2)$ and $f_0\in L^\infty\cap B^s_{2,\infty}$. Then there is a constant $C(f_0,K)>0$ depending only on f_0 and K such that*

$$\left|\mathbb{E}[\upsilon_{2,n}]-\left(\int_{\mathbb{R}} f_0^2(x)\,dx\right)^2\right|\le C(f_0,K)h_n^{2s}\,.$$

Moreover, if $n^2h_n\to\infty$, then it holds that

$$\upsilon_{2,n}=\left(\int_{\mathbb{R}} f_0^2(x)\,dx\right)^2+O_P\left(h_n^{2s}\vee\frac{1}{\sqrt{n}}\vee\frac{1}{n\sqrt{h_n}}\right).$$

Proof of Lemma A.4. (*Part A*) Since $(X_i)_{i=1,\dots,n}$ is an i.i.d. sample, it is directly seen that

$$\mathbb{E}[\upsilon_{2,n}]=\mathbb{E}[k_n(X_1,X_2)]\mathbb{E}[k_n(X_3,X_4)]=\mathbb{E}[k_n(X_1,X_2)]^2.$$

By a simple factorization inequality, we get

$$\begin{aligned}\left|\mathbb{E}\left[v_{2,n}\right]-\left(\int_{\mathbb{R}} f_0^2(x)\,dx\right)^2\right| &= \left|\mathbb{E}\left[k_n(X_1,X_2)\right]-\int_{\mathbb{R}} f_0^2(x)\,dx\right|\left|\mathbb{E}\left[k_n(X_1,X_2)\right]+\int_{\mathbb{R}} f_0^2(x)\,dx\right| \\ &\le \left(\|f_0\|_\infty\|K\|_1+\|f_0\|_2^2\right)\left|\mathbb{E}\left[k_n(X_1,X_2)\right]-\int_{\mathbb{R}} f_0^2(x)\,dx\right| \\ &\le \left(\|f_0\|_\infty\|K\|_1+\|f_0\|_2^2\right)C(f_0,K)h_n^{2s},\end{aligned}$$

where the first inequality follows from Lemma 3.1 and the second inequality from Proposition 3.3. This is enough to conclude the proof.

(*Part B*) Given the previous bound on the bias of $v_{2,n}$, it suffices again to bound the variance of $v_{2,n}$. Since it is a fourth-order U-statistic based on an i.i.d. sample, it admits a simple Hoeffding decomposition which leads to a simple expression for its variance. We then simply need to bound $\mathbb{E}\left[k_n^T(X_1)^2\right]$, $\mathbb{E}\left[k_n^T(X_1,X_2)^2\right]$, $\mathbb{E}\left[k_n^T(X_1,X_2,X_3)^2\right]$, and $\mathbb{E}\left[k_n^T(X_1,X_2,X_3,X_4)^2\right]$ where $k_n^T(X_1)=\mathbb{E}\left[k_n(X_1,X_2)k_n(X_3,X_4)|X_1\right]$, $k_n^T(X_1,X_2)=\mathbb{E}\left[k_n(X_1,X_2)k_n(X_3,X_4)|X_1,X_2\right]$, $k_n^T(X_1,X_2,X_3)=\mathbb{E}\left[k_n(X_1,X_2)k_n(X_3,X_4)|X_1,X_2,X_3\right]$, and $k_n^T(X_1,X_2,X_3,X_4)=k_n(X_1,X_2)k_n(X_3,X_4)$. This can again be done directly from the bounds in Lemma 3.3. It is seen, in particular, that

$$\begin{aligned}&\mathbb{E}\left[k_n^T(X_1)^2\right]=\mathbb{E}\left[k_n(X_1,X_2)\right]^2\mathbb{E}\left[k_n(X_1)^2\right]=O(1)\,,\\ &\mathbb{E}\left[k_n^T(X_1,X_2)^2\right]=\mathbb{E}\left[k_n(X_1,X_2)\right]^2\mathbb{E}\left[k_n(X_1,X_2)^2\right]=O(h_n^{-1})\,,\\ &\mathbb{E}\left[k_n^T(X_1,X_2,X_3)^2\right]=\mathbb{E}\left[k_n(X_1,X_2)^2\right]\mathbb{E}\left[k_n(X_1)^2\right]=O(h_n^{-1})\,,\\ &\mathbb{E}\left[k_n^T(X_1,X_2,X_3,X_4)^2\right]=\mathbb{E}\left[k_n(X_1,X_2)^2k_n(X_3,X_4)^2\right]=O(h_n^{-2})\,.\end{aligned}$$

It follows directly from these bounds that $\operatorname{Var} v_{2,n}=O(n^{-1}\vee n^{-2}h_n^{-1})$ by noting that the variance of the cubic term is of order $n^{-3}h_n^{-1}$ and that of the quartic term of order $n^{-4}h_n^{-2}$ and using the assumption that $n^2h_n\to\infty$. This is enough to obtain the result. □

Proof of Lemma 4.1. Since $s\in(0,1/2)$ and $h_n\asymp n^{-2/(4s+1)}$, it holds that $n^2h_n\to\infty$. From the proofs of Lemma A.3 and Lemma A.4, it follows immediately that

$$v_{L,n}=\int_{\mathbb{R}} f_0^3(x)\,dx-\left(\int_{\mathbb{R}} f_0^2(x)\,dx\right)^2+O_p\left(h_n^{2s}\vee\frac{1}{\sqrt{n}}\vee\frac{1}{n^{3/2}h_n}\right).$$

By plugging $h_n\asymp n^{-2/(4s+1)}$, the result follows directly. □

Proof of Lemma 4.2. The estimator $h_nv_{3,n}$ is a second-order U-statistic with kernel $h_nk_n^2(x,y)=K^2((x-y)/h_n)/h_n$. The proof for the bias bound of U_n in Proposition 3.3 immediately transposes with kernel K^2 instead of K. The same variance bounds as for U_n obtain from the Hoeffding decomposition of $h_nv_{3,n}$. It thus follows that

$$h_nv_{3,n}=\int_{\mathbb{R}} f_0^2(x)\,dx\int_{\mathbb{R}} K^2(u)\,du+O_p\left(h_n^{2s}\vee\frac{1}{\sqrt{n}}\vee\frac{1}{n\sqrt{h_n}}\right).$$

Since $s\in(0,1/2)$ and $h_n\asymp n^{-2/(4s+1)}$, it holds that $n^2h_n\to\infty$. From Lemma A.4, we then know that

$$h_n \upsilon_{2,n} = O(h_n) + O_p\Big(h_n^{2s+1} \vee \frac{h_n}{\sqrt{n}} \vee \frac{\sqrt{h_n}}{n}\Big).$$

It follows that

$$\upsilon_{W,n} = \int_{\mathbb{R}} f_0^2(x)\, dx \int_{\mathbb{R}} K^2(u)\, du + O_p\Big(h_n^{2s} \vee \frac{1}{\sqrt{n}} \vee \frac{1}{n\sqrt{h_n}}\Big).$$

By plugging $h_n \asymp n^{-2/(4s+1)}$, the result follows directly. □

Proof of Proposition 4.1. By independence and the law of iterated expectations, we already know that $\mathbb{E}[\upsilon_{1,n}] = \mathbb{E}[k_n^2(X_1)]$. It follows that $\mathbb{E}[\upsilon_{1,n}] = \operatorname{Var} k_n(X_1) + \mathbb{E}[k_n(X_1)]^2 = \operatorname{Var} k_n(X_1) + \mathbb{E}[k_n(X_1, X_2)]^2$. Similarly, by independence, we already obtained that $\mathbb{E}[\upsilon_{2,n}] = \mathbb{E}[k_n(X_1, X_2)]^2$. It follows directly that $\mathbb{E}[\upsilon_{L,n}] = \operatorname{Var} k_n(X_1)$. By independence, we have that $\mathbb{E}[\upsilon_{3,n}] = \mathbb{E}[k_n^2(X_1, X_2)]$ and so $\mathbb{E}[\upsilon_{3,n}] = \operatorname{Var}(k_n(X_1, X_2)) + \mathbb{E}[k_n(X_1, X_2)]^2$. It follows that $\mathbb{E}[\upsilon_{W,n}]/h_n = \operatorname{Var}(k_n(X_1, X_2))$. From the Hoeffding decomposition for the variance of U_n, we directly obtain $\mathbb{E}[V_{U,n}] = \operatorname{Var} U_n$ and this concludes the proof. □

Lemma A.5. *Let $s \in (0, 1/2)$ and $f_0 \in L^\infty \cap B^s_{2,\infty}$. For any bandwidth sequence $h_n \to 0$, it holds that*

$$\operatorname{Var} U_n = \frac{4(n-2)}{n(n-1)}\sigma_L^2 + \frac{2}{n(n-1)}\frac{1}{h_n}\sigma_W^2 + O\Big(\frac{h_n^{2s}}{n} \vee \frac{h_n^{2s}}{n^2 h_n}\Big).$$

Proof of Lemma A.5. The variance of U_n admits the decomposition

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

We know from the proof of Proposition 4.1 that $\operatorname{Var} k_n(X_1) = \mathbb{E}[\upsilon_{L,n}]$ and $\operatorname{Var}(k_n(X_1, X_2)) = \mathbb{E}[\upsilon_{W,n}]/h_n$. From the proofs of Lemma 4.1 and Lemma 4.2, we know that $\mathbb{E}[\upsilon_{L,n}] = \sigma_L^2 + O(h_n^{2s})$ and $\mathbb{E}[\upsilon_{W,n}] = \sigma_W^2 + O(h_n^{2s})$. The result then follows immediately by substituting these approximations in the variance decomposition. □

Proof of Proposition 4.2. We have from the proof of Lemma 4.1 and Lemma 4.2 that

$$V_{U,n} - \operatorname{Var} U_n = \frac{1}{n} O_p\Big(h_n^{2s} \vee \frac{1}{\sqrt{n}} \vee \frac{1}{n^{3/2} h_n}\Big) + \frac{1}{n^2 h_n} O_p\Big(h_n^{2s} \vee \frac{1}{\sqrt{n}} \vee \frac{1}{n\sqrt{h_n}}\Big) + O\Big(\frac{h_n^{2s}}{n} \vee \frac{h_n^{2s}}{n^2 h_n}\Big).$$

From Lemma A.5, we also have $\operatorname{Var} U_n \asymp n^{-1} \vee n^{-2} h_n^{-1}$. Then take $h_n \asymp n^{-2/(4s+1)}$. If $s \geq 1/4$, then $(\operatorname{Var} U_n)^{-1}(V_{U,n} - \operatorname{Var} U_n) = O_p(n^{-1/2}) + n^{(1-4s)/(4s+1)} O_p(n^{-1/2}) + O(n^{-4s/(4s+1)}) = O_p(n^{-1/2})$. If $s < 1/4$, then $(\operatorname{Var} U_n)^{-1}(V_{U,n} - \operatorname{Var} U_n) = n^{(4s-1)/(4s+1)} O_p(n^{(1-12s)/(8s+2)}) + O_p(n^{-4s/(4s+1)}) + O(n^{-4s/(4s+1)}) = O_p(n^{-4s/(4s+1)})$. By combining the two cases, this is enough to conclude. □

A.5 Proofs for the plug-in variance estimators

Lemma A.6. *Let $f_0 \in L^\infty$. It holds that*

$$s_{L,n}^2 = \frac{(n-1)}{n^2}\upsilon_{3,n} + \frac{(n-1)(n-2)}{n^2}\upsilon_{1,n} - \Big(\frac{n-1}{n}\Big)^2 U_n^2 .$$

Moreover, if $n^2 h_n \to \infty$, then

$$s_{L,n}^2 = \sigma_L^2 + \frac{1}{nh_n}\sigma_W^2 + o_P\Big(1 \vee \frac{1}{nh_n}\Big).$$

Proof of Lemma A.6. The first thing to note is that $s_{L,n}^2$ rewrites as a leave-one-out estimator since the diagonal terms $\frac{1}{h_n}K(0)$ cancel out. That is,

$$\begin{aligned} s_{L,n}^2 &= \frac{1}{n}\sum_{i=1}^n \Big(\frac{1}{n}\sum_{j\neq i} k_n(X_i, X_j)\Big)^2 - \Big(\frac{1}{n^2}\sum_{i=1}^n\sum_{j\neq i} k_n(X_i, X_j)\Big)^2 \\ &= \frac{1}{n}\sum_{i=1}^n \Big(\frac{1}{n}\sum_{j\neq i} k_n(X_i, X_j)\Big)^2 - \Big(\frac{n-1}{n}\Big)^2 U_n^2. \end{aligned}$$

It is then seen that the first term can be exactly decomposed into appropriately rescaled versions of $\upsilon_{3,n}$ and $\upsilon_{1,n}$. It follows that

$$s_{L,n}^2 = \frac{(n-1)}{n^2}\upsilon_{3,n} + \frac{(n-1)(n-2)}{n^2}\upsilon_{1,n} - \Big(\frac{n-1}{n}\Big)^2 U_n^2.$$

Suppose now that $n^2 h_n \to \infty$. From the results of Section 3, we have that $U_n = \int_{\mathbb{R}} f_0^2(x)\,dx + o_P(1)$ and so, by continuity, we obtain that $U_n^2 = (\int_{\mathbb{R}} f_0^2(x)\,dx)^2 + o_P(1)$. From the results of Section A.4, we have that $\upsilon_{1,n} = \int_{\mathbb{R}} f_0^3(x)\,dx + o_P(1)$ and that $h_n\upsilon_{3,n} = \sigma_W^2 + o_P(1)$. By combining the approximations in the expansion for $s_{L,n}^2$, we directly obtain that

$$s_{L,n}^2 = \sigma_L^2 + \frac{1}{nh_n}\sigma_W^2 + o_P\Big(1 \vee \frac{1}{nh_n}\Big).$$

□

Proof of Lemma 4.3. From the results of Section 3, we have that $U_n = \int_{\mathbb{R}} f_0^2(x)\,dx + O_P(h_n^{2s} \vee n^{-1/2} \vee n^{-1}h_n^{-1/2})$. It follows by expanding the square that $U_n^2 = (\int_{\mathbb{R}} f_0^2(x)\,dx)^2 + O_P(h_n^{2s} \vee n^{-1/2} \vee n^{-1}h_n^{-1/2})$. From the results of Section A.4, we know that $\upsilon_{1,n} = \int_{\mathbb{R}} f_0^3(x)\,dx + O_P(h_n^{2s} \vee n^{-1/2} \vee n^{-3/2}h_n^{-1})$ and that $h_n\upsilon_{3,n} = \sigma_W^2 + O_P(h_n^{2s} \vee n^{-1/2} \vee n^{-1}h_n^{-1/2})$. By combining all the approximations in the expansion, we get that

$$s_{L,n}^2 = \sigma_L^2 + O_P\Big(h_n^{2s} \vee \frac{1}{nh_n}\Big).$$

By taking $h_n \asymp n^{-1/(2s+1)}$, we obtain the result. □

Lemma A.7. *Let $f_0 \in L^\infty$. It holds that*

$$s_{W,n}^2 = h_n \frac{(n-1)}{n} \upsilon_{3,n} + \frac{(n-1)}{n^2 h_n} K^2(0) - h_n \Big(\frac{n-1}{n}\Big)^2 U_n^2 - \frac{2(n-1)}{n^2} K(0) U_n .$$

Moreover, if $n^2 h_n \to \infty$, then

$$s_{W,n}^2 = \sigma_W^2 + \frac{1}{nh_n} K^2(0) + o_p\Big(1 \vee \frac{1}{nh_n}\Big) .$$

Proof of Lemma A.7. By isolating the diagonal term and expanding the square, it is directly seen that

$$s_{W,n}^2 = h_n \frac{(n-1)}{n} \upsilon_{3,n} + \frac{(n-1)}{n^2 h_n} K^2(0) - h_n \Big(\frac{n-1}{n}\Big)^2 U_n^2 - \frac{2(n-1)}{n^2} K(0) U_n .$$

Suppose now that $n^2 h_n \to \infty$. From the results of Section 3, we know that $U_n = O_p(1)$ and $U_n^2 = O_p(1)$ so that $U_n/n = o_p(1)$ and $h_n U_n^2 = o_p(1)$. We then know from the proof of Lemma 4.2 that $h_n \upsilon_{3,n} = \sigma_W^2 + o_p(1)$. Since $\frac{K^2(0)}{nh_n}$ is deterministic, we directly obtain by combining the terms that

$$s_{W,n}^2 = \sigma_W^2 + \frac{1}{nh_n} K^2(0) + o_p\Big(1 \vee \frac{1}{nh_n}\Big) .$$

□

Proof of Lemma 4.4. Since $s \in (0, 1/2)$ and $h_n \asymp n^{-1/(2s+1)}$, it holds that $n^2 h_n \to \infty$. We know that $U_n = \int_{\mathbb{R}} f_0^2(x)\, dx + O_p(h_n^{2s} \vee n^{-1/2} \vee n^{-1} h_n^{-1/2})$ and $U_n^2 = (\int_{\mathbb{R}} f_0^2(x)\, dx)^2 + O_p(h_n^{2s} \vee n^{-1/2} \vee n^{-1} h_n^{-1/2})$ as well as $h_n \upsilon_{3,n} = \sigma_W^2 + O_p(h_n^{2s} \vee n^{-1/2} \vee n^{-1} h_n^{-1/2})$. It follows immediately that

$$s_{W,n}^2 = \sigma_W^2 + O_p\Big(h_n^{2s} \vee \frac{1}{nh_n}\Big) .$$

By taking $h_n \asymp n^{-1/(2s+1)}$, the result follows directly. □

Lemma A.8. *For any bandwidth sequence $h_n \to 0$ and any $n \geq 4$, it holds that:*

1.

$$\mathbb{E}\,[s_{L,n}^2] - \operatorname{Var} k_n(X_1) = -\frac{7n^2 - 14n + 8}{n^3} \operatorname{Var} k_n(X_1) + \frac{(n-1)(n-2)}{n^3} \operatorname{Var} k_n(X_1, X_2)$$

2.

$$\begin{aligned}\mathbb{E}\,[s_{W,n}^2] - h_n \operatorname{Var} k_n(X_1, X_2) = &- h_n \frac{4(n-1)(n-2)}{n^3} \operatorname{Var} k_n(X_1) - h_n \frac{n^2 + 2n - 2}{n^3} \operatorname{Var} k_n(X_1, X_2) \\ &+ \frac{(n-1)}{n^2 h_n} K^2(0) - \frac{2(n-1)}{n^2} K(0) \mathbb{E}\,[k_n(X_1, X_2)] \\ &+ h_n \frac{n-1}{n^2} \mathbb{E}\,[k_n(X_1, X_2)]^2\end{aligned}$$

3.

$$
\begin{aligned}
\mathbb{E}\,[S^2_{U,n}(h_n,h_n)] - \operatorname{Var} U_n(h_n) = &- \frac{4(n-2)(7n^2-12n+6)}{n^4(n-1)}\operatorname{Var} k_n(X_1) \\
&+ \frac{2(2n^3-11n^2+14n-6)}{n^4(n-1)}\operatorname{Var} k_n(X_1,X_2) \\
&+ \frac{2}{n^3h_n^2}K^2(0) - \frac{4}{n^3h_n}K(0)\mathbb{E}\,[k_n(X_1,X_2)] \\
&+ \frac{2}{n^3}\mathbb{E}\,[k_n(X_1,X_2)]^2
\end{aligned}
$$

Proof of Lemma A.8. 1. From the proof of Proposition 4.1, we have $\mathbb{E}\,[v_{1,n}] = \operatorname{Var} k_n(X_1) + \mathbb{E}\,[k_n(X_1,X_2)]^2$ and $\mathbb{E}\,[v_{3,n}] = \operatorname{Var} k_n(X_1,X_2) + \mathbb{E}\,[k_n(X_1,X_2)]^2$. We know that $\mathbb{E}\,[U_n^2] = \operatorname{Var} U_n + \mathbb{E}\,[k_n(X_1,X_2)]^2$. From the expansion of $s^2_{L,n}$, we get

$$
\mathbb{E}\,[s^2_{L,n}] = \frac{(n-1)}{n^2}\operatorname{Var} k_n(X_1,X_2) + \frac{(n-1)(n-2)}{n^2}\operatorname{Var} k_n(X_1) - \Big(\frac{n-1}{n}\Big)^2\operatorname{Var} U_n\,,
$$

since the terms $\mathbb{E}\,[k_n(X_1,X_2)]^2$ cancel out. By the Hoeffding variance equality for $\operatorname{Var} U_n$ and direct computations, the result follows.

2. The proof proceeds similarly from the expansion of $s^2_{W,n}$ by noting that $(n-1)K^2(0)/(n^2h_n)$ is deterministic and $\mathbb{E}\,[U_n] = \mathbb{E}\,[k_n(X_1,X_2)]$. The result then follows from the Hoeffding variance equality for $\operatorname{Var} U_n$ and direct computations. In this case, the terms $\mathbb{E}\,[k_n(X_1,X_2)]^2$ do not cancel out but reduce to $h_n(n-1)\mathbb{E}\,[k_n(X_1,X_2)]^2/n^2$.

3. The proof follows directly from combining (1) and (2). □

Proof of Proposition 4.3. (*Part A*) We recall that $\mathbb{E}\,[k_n(X_1,X_2)] = O(1)$, $\mathbb{E}\,[k_n(X_1,X_2)]^2 = O(1)$, $\operatorname{Var} k_n(X_1) = O(1)$, $h_n \operatorname{Var} k_n(X_1,X_2) = \sigma_W^2 + o(1)$. It follows, in particular, that $2(2n^3-11n^2+14n-6)\operatorname{Var} k_n(X_1,X_2)/(n^4(n-1)) + 2K^2(0)/n^3h_n^2 = D_n + o(D_n)$ where $D_n = 4\sigma_W^2/n^2h_n + 2K^2(0)/n^3h_n^2$. All other terms in the expansion of the finite-sample bias of $S^2_{U,n}(h_n,h_n)$ from Lemma A.8 are seen to be negligible with respect to D_n. The result then follows immediately.

(*Part B*) The result follows immediately from combining Lemma A.6 and Lemma A.7. □

Proof of Proposition 4.4. Suppose that $h'_n \asymp n^{-1/(2s+1)}$. From Lemma A.5, Lemma 4.3, and Lemma 4.4, we directly have for any h_n

$$
S^2_{U,n}(h_n,h'_n) - \operatorname{Var} U_n = \Big[\frac{4(n-2)}{n(n-1)} + \frac{2}{n(n-1)h_n}\Big]O_P\Big(n^{-2s/(2s+1)}\Big) + O\Big(\frac{h_n^{2s}}{n} \vee \frac{h_n^{2s}}{n^2h_n}\Big)\,.
$$

We know that $\operatorname{Var} U_n \asymp n^{-1} \vee n^{-2}h_n^{-1}$. Suppose now that $h_n \asymp n^{-2/(4s+1)}$. For any $s \in (0,1/2)$, we directly obtain from the previous expansion that $(\operatorname{Var} U_n)^{-1}(S^2_{U,n}(h_n,h'_n) - \operatorname{Var} U_n) = O_P(n^{-2s/(2s+1)}) + O(n^{-4s/(4s+1)})$. Since $O(n^{-4s/(4s+1)}) = o(n^{-2s/(2s+1)})$, the result follows immediately. □

A.6 Proof for the nonparametric bootstrap variance estimators

Proof of Corollary 4.6. From Lemma A.5, we have that

$$\operatorname{Var} U_n(h_n) = \frac{4(n-2)}{n(n-1)}\sigma_L^2 + \frac{2}{n(n-1)}\frac{1}{h_n}\sigma_W^2 + o\Big(\frac{1}{n} \vee \frac{1}{n^2 h_n}\Big).$$

It follows from Proposition 4.3 that

$$\begin{aligned}\frac{\operatorname{Var}^* U_n^*(h_n)}{\operatorname{Var} U_n(h_n)} &= 1 + \frac{1}{nh_n}\left(\frac{4\sigma_W^2 + \frac{2K^2(0)}{nh_n}}{4\sigma_L^2 + \frac{2\sigma_W^2}{nh_n}}\right)(1+o_p(1)) \\ &= 1 + O_p\Big(\frac{1}{nh_n}\Big).\end{aligned}$$

Suppose that $h_n \asymp n^{-2/(4s+1)}$. If $s > 1/4$, then $1 + O_p(n^{-1}h_n^{-1}) = 1 + O_p(n^{-(4s-1)/(4s+1)})$. If $s < 1/4$, then $1 + O_p(n^{-1}h_n^{-1}) = O_p(n^{(1-4s)/(4s+1)})$. This is enough to conclude. □

A.7 Proofs for the subsampling variance estimators

Lemma A.9. *Let $f_0 \in L^\infty$. For any bandwidth sequence $h_n \to 0$ and any $n \geq 4$, it holds that:*

1.
$$\psi_{L,n}^2(h_n) = \Big(\frac{n}{n-2}\Big)^2 s_{L,n}^2(h_n)\,;$$

2.
$$\psi_{W,n}^2(h_n) = \upsilon_{3,n}(h_n) - \frac{2n^2}{(n-1)(n-2)} s_{L,n}^2(h_n) - U_n^2(h_n)\,.$$

Proof of Lemma A.9. 1. By expanding the square in the definition of $\psi_{L,n}^2$ and noticing that the cross terms are equal to $2(n-1)^2 U_n^2$, we get

$$\begin{aligned}\psi_{L,n}^2 &= \Big(\frac{1}{n-2}\Big)^2 \Big[\frac{1}{n}\sum_{i=1}^n \Big(\sum_{j\neq i} k_n(X_i,X_j)\Big)^2 - (n-1)^2 U_n^2\Big] \\ &= \Big(\frac{n}{n-2}\Big)^2 \Big[\frac{1}{n}\sum_{i=1}^n \Big(\frac{1}{n}\sum_{j\neq i} k_n(X_i,X_j)\Big)^2 - \Big(\frac{n-1}{n}\Big)^2 U_n^2\Big].\end{aligned}$$

This is enough to conclude from the proof of Lemma A.6.

2. By expanding the square in the definition of $\psi_{W,n}^2$ and simplifying the coefficients for each term, we get

$$\psi_{W,n}^2 = \upsilon_{3,n} - \frac{2(n-1)}{n(n-2)}\sum_{i=1}^n \Big(\frac{1}{n-1}\sum_{j\neq i} k_n(X_i,X_j)\Big)^2 + \frac{n}{n-2}U_n^2\,.$$

By noticing then that

$$\frac{1}{n}\sum_{i=1}^n \Big(\frac{1}{n-1}\sum_{j\neq i} k_n(X_i,X_j)\Big)^2 = \Big(\frac{n}{n-1}\Big)^2 s_{L,n}^2 + U_n^2\,,$$

we directly get

$$\psi^2_{W,n} = \upsilon_{3,n} - \frac{2n^2}{(n-1)(n-2)} s^2_{L,n} - U^2_n.$$

□

Proof of Proposition 4.7. (*Part A*) We start by computing $\mathbb{E}\,[\psi^2_{L,n}]$ and $\mathbb{E}\,[\psi^2_{W,n}]$ using the expansions of Lemma A.9 and the bias results for $\upsilon_{3,n}$ and $s^2_{L,n}$ of Section A.4 and Section A.5. From Lemma A.9, we have

$$\mathbb{E}\,[\psi^2_{L,n}] = \frac{n-1}{n(n-2)}\Big((n-4)\operatorname{Var} k_n(X_1) + \operatorname{Var} k_n(X_1, X_2)\Big).$$

Similarly, by expanding and grouping the terms, we get

$$\mathbb{E}\,[\psi^2_{W,n}] = \frac{n-3}{n-1}\Big(-2\operatorname{Var} k_n(X_1) + \operatorname{Var} k_n(X_1, X_2)\Big).$$

By combining and grouping the terms together, we obtain that

$$\begin{aligned}\mathbb{E}\,[\operatorname{Var}^* U^*_S] &= \frac{4(n-m)}{m(n-1)}\mathbb{E}\,[\psi^2_{L,n}] + \frac{2(n-m)(n-m-1)}{m(m-1)(n-2)(n-3)}\mathbb{E}\,[\psi^2_{W,n}] \\ &= \frac{4(n-m)(mn-2m-2n+2)}{mn(m-1)(n-1)}\operatorname{Var} k_n(X_1) + \frac{2(n-m)(n+m-1)}{mn(m-1)(n-1)}\operatorname{Var} k_n(X_1, X_2)\end{aligned}$$

Subtracting from $\mathbb{E}\,[\operatorname{Var}^* U^*_S]$ the Hoeffding decomposition of the variance $\operatorname{Var} U_n$ and grouping the terms yields the result.

(*Part B*) From Lemma A.9 and Lemma A.6, we directly have

$$\psi^2_{L,n} = \sigma^2_L + \frac{1}{nh_n}\sigma^2_W + o_p\Big(1 \vee \frac{1}{nh_n}\Big).$$

From Lemma A.9, Lemma A.6, and the proof of Lemma 4.2, we obtain that

$$\begin{aligned}\psi^2_{W,n} &= \frac{1}{h_n}\sigma^2_W + o_p\Big(\frac{1}{h_n}\Big) - 2\sigma^2_L - \frac{2}{nh_n}\sigma^2_W + o_p\Big(1 \vee \frac{1}{nh_n}\Big) + O(1) \\ &= \frac{1}{h_n}\sigma^2_W + o_p\Big(\frac{1}{h_n}\Big).\end{aligned}$$

Suppose now that $m/n \to 0$. Then from the definition of $\operatorname{Var}^* U^*_S$ and by noticing that the terms $O_p((mnh_n)^{-1})$ and $O(m^{-2})$ are $o_p(m^{-1} \vee m^{-2}h_n^{-1})$, we can directly conclude that

$$\operatorname{Var}^* U^*_S = \frac{4}{m}\sigma^2_L + \frac{2}{m^2h_n}\sigma^2_W + o_p\Big(\frac{1}{m} \vee \frac{1}{m^2h_n}\Big).$$

□

Proof of Proposition 4.8. From Lemma A.9, the proofs of Proposition 4.2 and Proposition 4.3, we have that

$$\psi_{L,n}^2 = \sigma_L^2 + O_P\Big(h_n^{2s} \vee \frac{1}{nh_n}\Big);$$
$$h_n\psi_{W,n}^2 = \sigma_W^2 + O_P\Big(h_n^{2s} \vee \frac{1}{\sqrt{n}} \vee \frac{1}{n\sqrt{h_n}}\Big).$$

Suppose that $m = n/\log(n)$. If we denote the coefficients of $\psi_{L,n}^2$ and $\psi_{W,n}^2$ in $\mathrm{Var}^* U_S^*$ by A_n and B_n, respectively, then $A_n \asymp 4\log(n)/n$ and $B_n \asymp 2(\log(n))^2/n^2$.

1. Suppose that $s > 1/4$ and $h_n \asymp n^{-2/(4s+1)}$. From the previous approximations, we obtain that
$$\frac{1}{\log(n)}\mathrm{Var}^* U_S^* = \frac{4}{n}\Big(\sigma_L^2 + O_P\Big(h_n^{2s} \vee \frac{1}{nh_n}\Big)\Big) + \frac{2\log(n)}{n^2 h_n}\sigma_W^2(1 + o_P(1)).$$

Since $\mathrm{Var}\, U_n \asymp 4\sigma_L^2/n$, we obtain that

$$\frac{1}{\log(n)}\frac{\mathrm{Var}^* U_S^*}{\mathrm{Var}\, U_n} = 1 + O_P\Big(\frac{1}{nh_n}\Big) + O_P\Big(\frac{\log(n)}{nh_n}\Big) = 1 + O_P\Big(\frac{\log(n)}{nh_n}\Big).$$

By noting that $nh_n \asymp n^{(4s-1)/(4s+1)}$, the result obtains.

2. Suppose that $s < 1/4$ and $h_n \asymp n^{-2/(4s+1)}$. From the earlier approximations, we obtain that

$$\begin{aligned}\frac{1}{(\log(n))^2}\mathrm{Var}^* U_S^* - \mathrm{Var}\, U_n &= \frac{4}{n\log(n)}\Big(\sigma_L^2 - \log(n)\sigma_L^2 + O_P\Big(h_n^{2s} \vee \frac{1}{nh_n}\Big)\Big) \\ &\quad + \frac{1}{n^2 h_n} O_P\Big(h_n^{2s} \vee \frac{1}{\sqrt{n}} \vee \frac{1}{n\sqrt{h_n}}\Big) + O\Big(\frac{h_n^{2s}}{n} \vee \frac{h_n^{2s}}{n^2 h_n}\Big) \\ &= O\Big(\frac{1}{n}\Big) + \frac{1}{n^2 h_n} O_P(h_n^{2s})\end{aligned}$$

Since $\mathrm{Var}\, U_n \asymp 2\sigma_W^2/n^2 h_n$, it follows that

$$\frac{1}{(\log(n))^2}\frac{\mathrm{Var}^* U_S^*}{\mathrm{Var}\, U_n} - 1 = O_P\Big(nh_n \vee h_n^{2s}\Big).$$

By noting that $nh_n \asymp n^{(4s-1)/(4s+1)}$, we obtain the result. □

B Supporting material for the simulations

B.1 Plots for the standard normal design

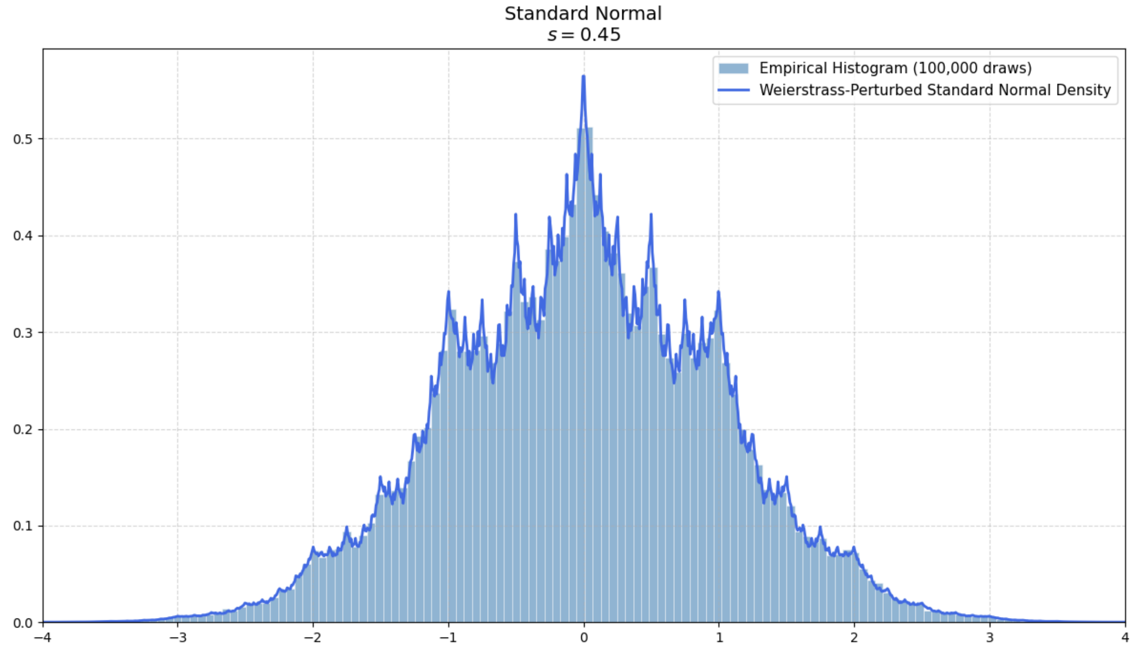

Figure 1: Density and Empirical Histogram, Standard Normal Design with $s = 0.45$

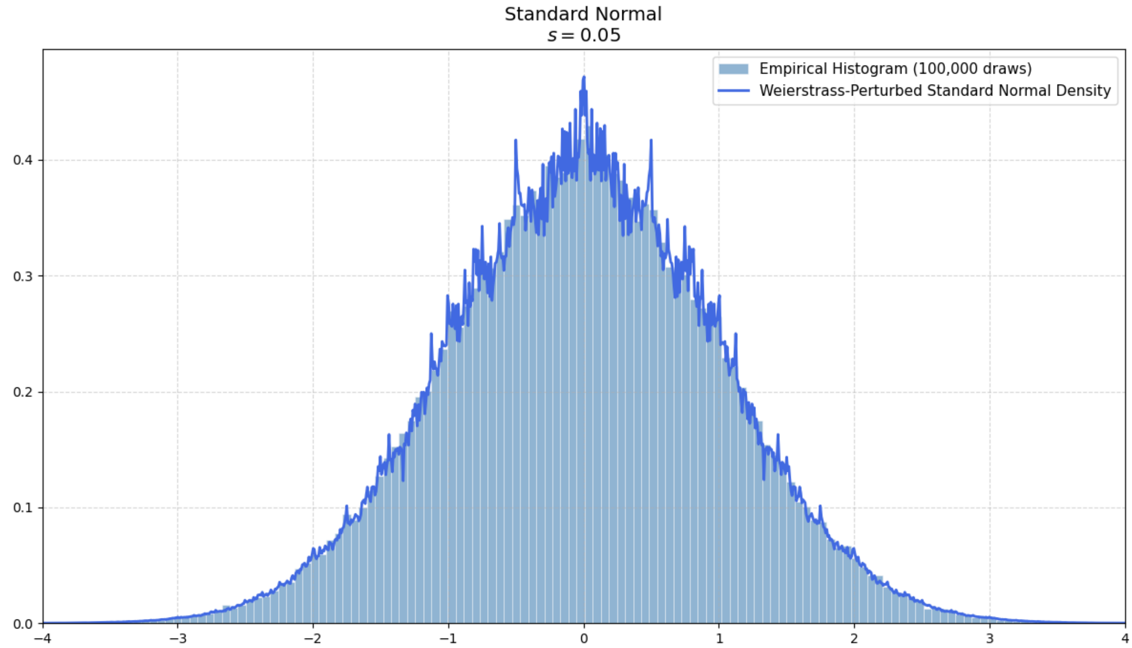

Figure 2: Density and Empirical Histogram, Standard Normal Design with $s = 0.05$

B.2 Plots for the Laplace mixture design

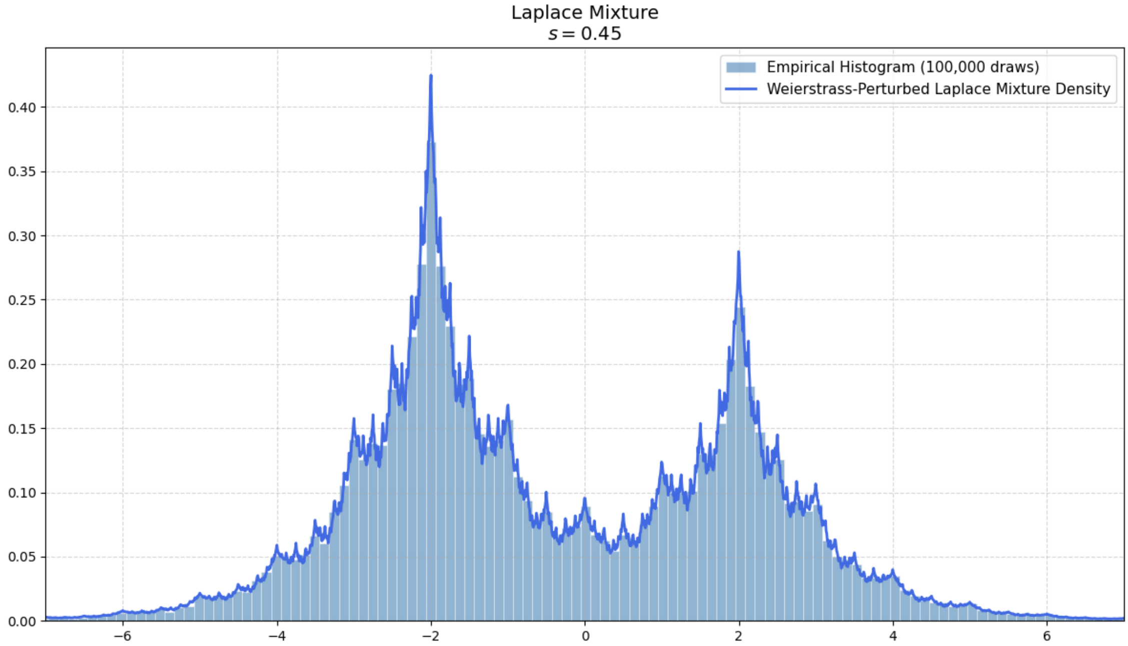

Figure 3: Density and Empirical Histogram, Laplace Mixture Design with $s = 0.45$

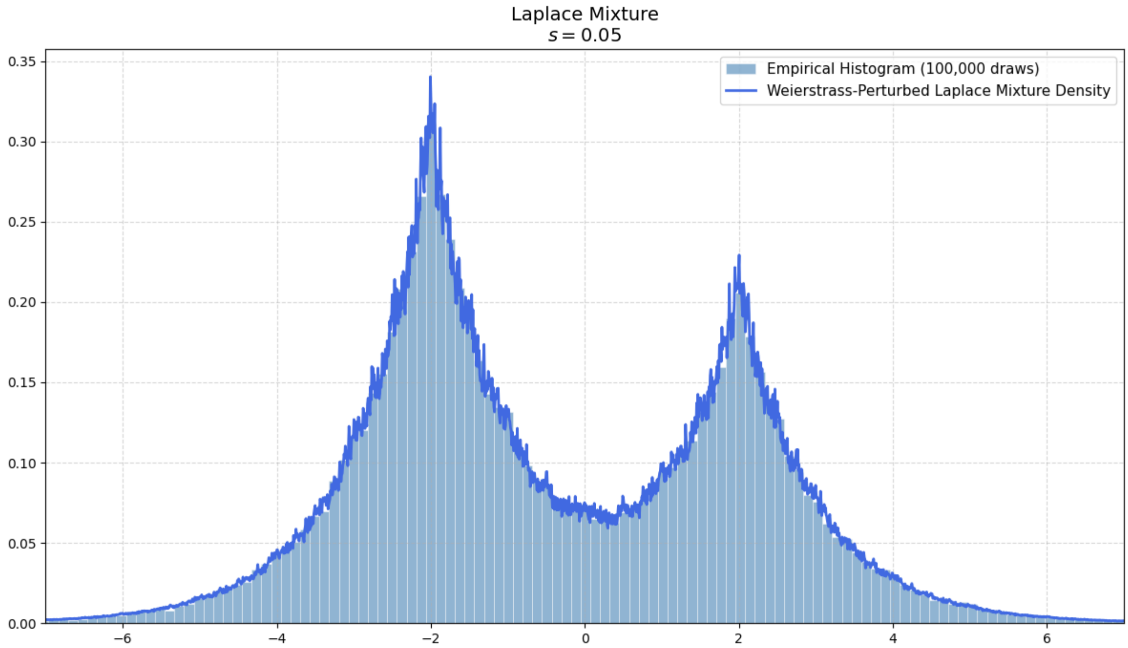

Figure 4: Density and Empirical Histogram, Laplace Mixture Design with $s = 0.05$