Higher-order Refinements of Small Bandwidth Asymptotics for Density-Weighted Average Derivative Estimators[*]

Matias D. Cattaneo[†]   Max H. Farrell[‡]   Michael Jansson[§]   Ricardo Masini[¶]

December 31, 2022

Abstract

The density weighted average derivative (DWAD) of a regression function is a canonical parameter of interest in economics. Classical first-order large sample distribution theory for kernel-based DWAD estimators relies on tuning parameter restrictions and model assumptions leading to an asymptotic linear representation of the point estimator. Such conditions can be restrictive, and the resulting distributional approximation may not be representative of the underlying sampling distribution of the statistic of interest, in particular not being robust to bandwidth choices. Small bandwidth asymptotics offers an alternative, more general distributional approximation for kernel-based DWAD estimators that allows for, but does not require, asymptotic linearity. The resulting inference procedures based on small bandwidth asymptotics were found to exhibit superior finite sample performance in simulations, but no formal theory justifying that empirical success is available in the literature. Employing Edgeworth expansions, this paper shows that small bandwidth asymptotics lead to inference procedures with demonstrable superior higher-order distributional properties relative to procedures based on asymptotic linear approximations.

Keywords: density weighted average derivatives, Edgeworth expansions, small bandwidth asymptotics.

1 Introduction

Identification, estimation and inference in the context of two-step semiparametric models has a long tradition in econometrics (Powell, 1994). Canonical two-step semiparametric parameters are finite dimensional functionals of some other unknown infinite dimensional parameters in the model (e.g., a density or regression function), a leading example being the density weighted average derivative (DWAD) of a regression function (Stoker, 1986). This paper seeks to honor the many contributions of Jim Powell to semiparametric theory in econometrics by juxtaposing the higher-order distributional properties of Powell *et al.* (1989)'s two-step kernel-based DWAD estimator under two alternative large sample approximation regimes: one based on the classical asymptotic linear representation, and the other based on a more general quadratic distributional approximation.[1]

In a landmark contribution, Powell *et al.* (1989) proposed a kernel-based DWAD estimator and obtained first-order, asymptotically linear distribution theory employing ideas from the U-statistics literature, along with plug-in standard error estimators, to develop valid inference procedures in large samples. This work sparked a wealth of subsequent developments in the econometrics literature: Robinson (1995) obtained Berry-Esseen bounds, Powell and Stoker (1996) considered mean square error expansions, Nishiyama and Robinson (2000, 2001, 2005) developed Edgeworth expansions, and Newey *et al.* (2004) investigated bias properties, just to mention a few contributions. The two-step semiparametric estimator in this literature employs a preliminary kernel-based estimator of a density function, which requires choosing two main tuning parameters (a bandwidth and a kernel function), and their "optimal" choices depend on the goal of interest (e.g., point estimation vs. inference) as well as the features of the underlying data generating process (e.g., smoothness of the unknown density and dimensionality of the covariates).

Classical first-order distribution theory for kernel-based DWAD estimators has focused on cases where tuning parameter restrictions and model assumptions lead to an asymptotic linear representation of the two-step semiparametric point estimator (Newey and McFadden, 1994; Ichimura and Todd, 2007, for overviews), that is, the two-step estimator is approximated by a sample average based on the so-called influence function. This approach can lead to semiparametric efficient inference procedures in large samples, but the implied distributional approximation may not be "robust" to tuning parameter choices and/or model features. More specifically, the limiting distribution obtained based on the asymptotic linear representation is invariant to the way that the preliminary nonparametric estimators are constructed, and requires potentially high smoothness levels of the underlying unknown functions and thus the use of higher-order kernels. At its core, asymptotic linear approximations assume away the contribution of additional terms forming the statistic of interest, despite the fact that these terms do contribute to the sampling variability of the two-step semiparametric estimator and, more importantly, do reflect the effect of tuning parameter

---

[1]Jim Powell's contributions to semiparametric theory are numerous. Honoré and Powell (1994), Powell and Stoker (1996), Blundell and Powell (2004), Aradillas-Lopez *et al.* (2007), Ahn *et al.* (2018), and Graham *et al.* (2023) are some of the most closely connected to the our work. These papers employ U-statistics methods for two-step kernel-based estimators similar to those considered herein. See Powell (2017) for more discussion and references.

choices in finite samples.

Cattaneo *et al.* (2014a) proposed an alternative distributional approximation for kernel-based DWAD estimators that allows for, but does not require, asymptotic linearity. The key idea is to capture the joint contribution to the sampling distribution of both linear and quadratic terms forming the kernel-based DWAD estimator. To operationalize this idea, Cattaneo *et al.* (2014a) introduced an asymptotic experiment where the bandwidth sequence is allowed to vanish at a speed that would render the classical asymptotic linear representation invalid because the quadratic term becomes first order even in large samples, which they termed "small bandwidth" asymptotics. This framework was carefully developed to obtain a distributional approximation that explicitly depends on both linear and quadratic terms, thereby forcing a more careful analysis of how the quadratic term contributes to the sampling distribution of the statistic.

Small bandwidth asymptotics inference methods for kernel-based DWAD estimators were found to perform well in simulations (Cattaneo *et al.*, 2010, 2014a,b), but no formal justification for its finite sample success is available in the literature. Methodologically, this alternative distributional approximation leads to a new way of conducting inference (e.g., constructing confidence interval estimators) because the original standard error formula proposed by Powell *et al.* (1989) must be modified to make the asymptotic approximation valid across the full range of allowable bandwidths (including the region where asymptotic linearity fails). Theoretically, however, the empirical success of small bandwidth asymptotics could in principle come from two distinct sources: (i) it could deliver a better distributional approximation to the sampling distribution of the point estimator; or (ii) it could deliver a better distributional approximation to the sampling distribution of the Studentized t-statistic because the standard error formula was modified.

Employing Edgeworth expansions (Bhattacharya and Rao, 1976; Hall, 1992), this paper shows that the small bandwidth asymptotics approximation framework leads to inference procedures with demonstrable superior higher-order distributional properties relative to procedures based on asymptotic linear approximations. We study both standardized and Studentized t-statistics, under both asymptotic linearity and small bandwidth asymptotic regimes, and show that both standardized and Studentized t-statistics emerging from the small bandwidth regime offer higher-order corrections as measured by the second cummulant underlying their Edgeworth expansions. An immediate implication of our results is that the small bandwidth asymptotic framework delivers both a better distributional approximation (Theorem 1, standardized t-statistic) and leads to a better standard error construction (Theorem 2, Studentized t-statistic). Therefore, our results have both theoretical and practical implications for empirical work in economics, in addition to providing a theory-based explanation for prior simulation-based findings exhibiting better numerical performance of inference procedures constructed using small bandwidth asymptotics relative to those constructed using classical distributional approximations.

The closest antecedent to our work is Nishiyama and Robinson (2000, 2001), who also studied Edgeworth expansions for kernel-based DWAD estimators. Their expansions, however, were motivated by the asymptotic linear approximation to the point estimator, and hence can not be used

to compare and contrast to the distributional approximation emerging from the alternative small bandwidth asymptotic regime. Therefore, from a technical perspective, this paper also offers novel Edgeworth expansions that allow for different standardization and Studentization schemes, thereby allowing us to plug-and-play when comparing the two competing asymptotic frameworks. More specifically, Theorem 1 below concerns a generic standardized t-statistic and is proven based on Theorem A in the appendix, which may be of independent technical interest due to is generality. Theorem 2 below concerns a more specialized class of Studentized t-statistic because establishing valid Edgeworth expansions is considerably harder when dealing with Studentization.

The idea of employing alternative (more general) asymptotic approximation frameworks that do not enforce asymptotic linearity for two-step semiparametric estimators has also featured in other context such as partially linear series-based, many covariates and many instrument estimation as well as certain network estimation settings (Cattaneo *et al.*, 2018a,b; Matsushita and Otsu, 2021), as well as other non-linear two-step semiparametric settings (Cattaneo *et al.*, 2013; Cattaneo and Jansson, 2018; Cattaneo *et al.*, 2019). While our theoretical developments and results focus specifically on the case of kernel-based DWAD estimation, their main conceptual conclusions can be extrapolated to those settings as well. The main takeaway is that employing alternative asymptotic frameworks can deliver improved inference with smaller higher-order distributional approximation errors, thereby offering more robust inference procedures in finite samples.

The paper continues as follows. Section 2 introduces the setup and main assumptions. Section 3 reviews the classical first-order distributional approximation based on asymptotic linearity and the more general small bandwidth distributional approximation, along with their corresponding choices of standard error formulas. Section 4 presents the main results of our paper. Section 5 concludes. The appendix is organized in three parts: Appendix A provides a self-contained generic Edgeworth expansion for second-order U-statistics, which may be of independent technical interest, Appendix B gives the proof of Theorem 1 (standardized t-statistic), and Appendix C gives the proof of Theorem 2 (Studentized t-statistic).

2 Setup and Assumptions

Suppose $Z_i = (Y_i, X_i')'$, $i = 1, \ldots, n$, is a random sample from the distribution of the random vector $Z = (Y, X')'$, where Y is an outcome variable of interest and X takes value on $\mathbb{R}^d$ with Lebesgue density f. We consider the density weighted average derivative of the regression function $g(X) = \mathbb{E}[Y|X]$ given by

$$\theta := \mathbb{E}[f(X)\dot{g}(X)],$$

where for any function a we define $\dot{a}(x) := \frac{\partial}{\partial x} a(x)$. To save notation, we also define $e(X) := f(X)g(X)$ and $v(X) := \mathbb{E}[Y^2|X]$. We impose the following conditions on the underlying data generating process. Let $\|\cdot\|$ be the Euclidean norm.

Assumption 1.

(a) $\mathbb{E}[|Y|^p] < \infty$, *for some* $p \geq 3$.

(b) $\Sigma := \mathbb{E}[\psi(Z)\psi(Z)']$ *is positive definite, where* $\psi(Z) := 2[\dot{e}(X) - Y\dot{f}(X) - \theta]$.

(c) f *is* $(S+1)$ *times differentiable, and* f *and its* $(S+1)$ *derivatives are bounded, for* $2S > d+2$;

(d) g *is* $(S+1)$ *times differentiable and its first three derivatives are bounded;*

(e) e *and its first* $(S+1)$ *derivatives are bounded;*

(f) v *is twice diferentiable, and its first two derivatives are bounded, and* $v\dot{f}$ *and* $\mathbb{E}[|Y|^3|X]f(X)$ *are bounded;*

(g) f, gf, $\dot{g}f$ *and* vf *vanish on the boundaries of their convex supports;*

(h) Cramér Condition: $\displaystyle \sup_{\nu \in \mathbb{R}^d : \|v\|=1} \limsup_{|t| \to \infty} |\mathbb{E}\exp(\iota t \ell_1/\bar{\sigma}_\nu)| < 1$ *where* $\bar{\sigma}_\nu := \nu'\Sigma\nu$.

Under Assumption 1 and using integration by parts, the DWAD vector can be expressed as

$$\theta = -2\mathbb{E}[Y\dot{f}(X)],$$

which motivates the celebrated plug-in analog estimator of Powell *et al.* (1989) given by

$$\widehat{\theta} = -2\frac{1}{n}\sum_{i=1}^n Y_i \widehat{\dot{f}}_i(X_i), \qquad \widehat{f}_i(x) = \frac{1}{n-1}\sum_{j=1, j \neq i}^n \frac{1}{h^d}K\left(\frac{X_j - x}{h}\right),$$

where $\widehat{f}_i(\cdot)$ is a "leave-one-out" kernel density estimator for kernel function $K : \mathbb{R}^d \to \mathbb{R}$ and positive vanishing (bandwidth) sequence h. For the kernel function, we impose the following conditions.

Assumption 2.

(a) K *is even, differentiable, and* $\dot{K}$ *is bounded;*

(b) $\int_{\mathbb{R}^d} \dot{K}(u)\dot{K}(u)'\mathrm{d}u$ *is positive definite;*

(c) For some $P \geq 2$,

$$\int_{\mathbb{R}^d} |K(u)|(1 + \|u\|^P)\mathrm{d}u + \int_{\mathbb{R}^d} \|\dot{K}(u)\|(1 + \|u\|^2)\mathrm{d}u < \infty$$

and

$$\int_{\mathbb{R}^d} u^a K(u)\mathrm{d}u = \begin{cases} 1, & \text{if } [a] = 0, \\ 0, & \text{if } 0 < [a] < P \\ \mu_a < \infty, & \text{if } [a] = P, \end{cases}$$

where $a \in \mathbb{Z}_+^d$ *is a multi-index.*[2]

---

[2] We employ standard multi-index notation. For $a := (a_1, \ldots, a_d)$ we have (i) $[a] := a_1 + \cdots + a_d$, (ii) $a! := a_1! \ldots a_d!$, (iii) $x^a := x_1^{a_1} \ldots x_d^{a_d}$ for $x \in \mathbb{R}^d$ and (iv) $q^{(a)}(x) = \frac{\partial^{[a]}q}{\partial^{a_1}x_1 \ldots \partial^{a_d}x_d}$ for smooth enough $q : \mathbb{R}^d \to \mathbb{R}$.

The estimator $\widehat{\theta}$ can be expressed as a second-order U-statistic with n-varying kernel:

$$\widehat{\theta} = \binom{n}{2}^{-1} \sum_{i<j}^{n} U_{ij}, \qquad U_{ij} = -\frac{1}{h^{d+1}} \dot{K}\left(\frac{X_i - X_j}{h}\right)(Y_i - Y_j), \tag{2.1}$$

where $\sum_{i<j}^{n}$ is shorthand notation for $\sum_{i=1}^{n-1} \sum_{j=i+1}^{n}$.

3 First-order Theory

Before presenting our main results concerning the higher-order distributional properties of different statistics based on $\widehat{\theta}$, we overview conventional and alternative asymptotic distributional approximations, and the variance estimation methods proposed in the literature emerging from those distinct approximation frameworks. Limits are taken as $h \to 0$ and $n \to \infty$ unless otherwise noted.

3.1 Distributional Approximation

In a landmark contribution, Powell *et al.* (1989) studied the first-order large sample distributional properties of $\widehat{\theta}$. They showed that, under appropriate restrictions on h and K, the estimator $\widehat{\theta}$ is asymptotically linear with (efficient) influence function $\psi(z)$, and thus with semiparametric (efficient) asymptotic variance Σ. More precisely, Powell *et al.* (1989) showed that if Assumptions 1 and 2 hold, and if $nh^{2\min(P,S)} \to 0$ and $nh^{d+2} \to \infty$, then

$$\sqrt{n}(\widehat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi(Z_i) + o_{\mathbb{P}}(1) \rightsquigarrow \mathcal{N}(0, \Sigma). \tag{3.1}$$

This result follows from the U-statistic representation in (2.1) and its Hoeffding decomposition, which gives $\widehat{\theta} = \mathbb{E}[U_{ij}] + \bar{L} + \bar{Q}$, where

$$\bar{L} = \frac{1}{n} \sum_{i=1}^{n} L_i, \qquad L_i = 2(\mathbb{E}[U_{ij}|Z_i] - \mathbb{E}[U_{ij}]),$$

and

$$\bar{Q} = \binom{n}{2}^{-1} \sum_{i<j}^{n} Q_{ij}, \qquad Q_{ij} = U_{ij} - \mathbb{E}[U_{ij}|Z_i] - \mathbb{E}[U_{ij}|Z_j] + \mathbb{E}[U_{ij}],$$

both mean zero random vectors. Because $\mathbb{E}[U_{ij}] = \theta + O(h^{\min(P,S)})$ and $\bar{Q} = O_{\mathbb{P}}(n^{-1}h^{-(d+2)/2})$, it follows that

$$\sqrt{n}(\widehat{\theta} - \theta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\mathbb{E}[U_{ij}|Z_i] - \mathbb{E}[U_{ij}]\right) + O_{\mathbb{P}}\left(\sqrt{n}h^{\min(P,S)} + \frac{1}{\sqrt{nh^{d+2}}}\right),$$

from which the asymptotic linear representation based on the (efficient) influence function in (3.1) is established upon noting that $\mathbb{E}[\|\bar{L} - \sum_{i=1}^{n} \psi(Z_i)/n\|^2] = O(n^{-1}h)$.

Conceptually, the Hoeffding decomposition and subsequent analysis of each of its terms shows that the estimator admits a bilinear form representation in general, which then is reduced to a sample average approximation by assuming a bandwidth sequence and kernel shape that makes both the misspecification error (smoothing bias) and the variability introduced by $\bar{Q}$ ("quadratic term" term) negligible in large samples. As a result, provided that such tuning parameter choices are feasible, the estimator will be asymptotically linear.

Asymptotic linearity of a semiparametric estimator has several distinct features that may be considered attractive from a theoretical point of view (Newey, 1994). In particular, it is a necessary condition for semiparametric efficiency and it leads to a limiting distribution that is invariant to the choice of the first-step nonparametric estimator entering the two-step semiparametric procedure. However, insisting on asymptotic linearity may also have its drawbacks because it requires several potentially strong assumptions and leads to a large sample theory that may not accurately represent the finite sample behavior of the statistic. In the case of $\widehat{\theta}$, asymptotic linearity requires $P > 2$ unless $d = 1$, thereby forcing restrictive smoothness conditions ($S \geq P$) and the use of higher-order kernels or similar debiasing techniques (see, e.g., Chernozhukov *et al.*, 2022, and references therein). In addition, classical asymptotic linear theory (whenever valid) leads to a limiting experiment which is invariant to the particular choices of smoothing (K) and bandwidth (h) tuning parameters involved in the construction of the estimator, and therefore it is unable to "adapt" to changes in those choices. As a result, asymptotically linear large sample distribution theory is silent with respect to the impact that tuning parameter choices may have on the finite sample behavior of the two-step semiparametric statistic.

To address the aforementioned limitations with classical asymptotic distribution theory, Cattaneo *et al.* (2014a) proposed a more general distributional approximation for kernel-based DWAD estimators that accommodates but does not enforces asymptotic linearity. The core idea is to characterize the joint asymptotic distributional features of both the linear ($\bar{L}$) and quadratic ($\bar{Q}$) terms jointly, and in the process develop an alternative first-order asymptotic theory that accommodates weaker assumptions than those imposed in the classical asymptotically linear distribution theory. Formally, if Assumptions 1 and 2 hold, and if $\min(nh^{d+2}, 1)nh^{2\min(P,S)} \to 0$ and $n^2h^d \to \infty$, then

$$(\mathbb{V}[\widehat{\theta}])^{-1/2}(\widehat{\theta} - \theta) \rightsquigarrow \mathcal{N}(0, I), \tag{3.2}$$

where

$$\mathbb{V}[\widehat{\theta}] = \mathbb{V}[\bar{L}] + \mathbb{V}[\bar{Q}], \qquad \mathbb{V}[\bar{L}] = \frac{1}{n}\big[\Sigma + o(1)\big], \qquad \mathbb{V}[\bar{Q}] = \binom{n}{2}^{-1} h^{-d-2}\big[\Delta + o(1)\big],$$

and $\Delta = 2\mathbb{E}[v(X)f(X)] \int_{\mathbb{R}^d} \dot{K}(u)\dot{K}(u)' \mathrm{d}u$.

This more general distributional approximation was developed explicitly in an attempt to better characterize the finite sample behavior of $\widehat{\theta}$. The result in (3.2) shows that the conditions on the bandwidth sequence may be considerably weakened without invalidating the limiting Gaussian distribution, albeit the asymptotic variance formula may change. Importantly, if nh^{d+2} is bounded

then $\widehat{\theta}$ is no longer asymptotically linear and its limiting distribution will cease to be invariant with respect to the underlying preliminary nonparametric estimator. In particular, if $nh^{d+2} \to c > 0$ then $\widehat{\theta}$ is root-n consistency but not asymptotically linear. In addition, because the bandwidth is allowed to be "smaller" than usual, the bias of the estimator is controlled in a different way, removing the need for higher-order kernels. Interestingly, (3.2) allows for the point estimator to not even be consistent for θ, for sufficiently small bandwidth sequences.

Beyond the aforementioned technical considerations, the result in (3.2) can conceptually be interpreted as a more refined first-order distributional approximation for the standarized statistics $(\mathbb{V}[\widehat{\theta}])^{-1/2}(\widehat{\theta} - \theta)$, which by relying on a quadratic approximation (i.e., capturing the stochastic contributions of both $\bar{L}$ and $\bar{Q}$) it is expected to offer a "better" distributional approximation. The idea of standarizing a U-statistic by the joint variance of the linear and quadratic terms underlying its Hoeffding decomposition can be traced back to the original paper of Hoeffding (1948, p. 307). Furthermore, the asymptotic distribution theory proposed by Cattaneo *et al.* (2014a) can be viewed as highlighting the well known trade-off between robustness and efficiency in two-step semiparametric settings: $\widehat{\theta}$ is semiparametric efficient if and only if $nh^{d+2} \to \infty$, while it seems possible to construct more robust inference procedures under considerably weaker conditions that would not be semiparametric efficient. Simulation evidence reported in Cattaneo *et al.* (2010, 2014a,b) corroborated those conceptual interpretations numerically, but no formal justification is available in the literature. Theorem 1 below will offer the first theoretical result in the literature highlighting specific robustness features of the distributional approximation in (3.2) by showing that such approximation has a demonstrably smaller higher-order distributional approximation error.

3.2 Variance Estimation

Based on the asymptotically linear distributional approximation in (3.1), Powell *et al.* (1989) also proposed the following variance estimator

$$\widehat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\widehat{L}_i\widehat{L}_i', \qquad \widehat{L}_i = 2\Big[\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}U_{ij} - \widehat{\theta}\Big],$$

and proved its consistency (i.e., $\widehat{\Sigma} \to_{\mathbb{P}} \Sigma$) under the same bandwidth sequences ($nh^{2\min(P,S)} \to 0$ and $nh^{d+2} \to \infty$) required for asymptotic linearity. This result justifies employing the Studentized statistic

$$\widehat{\Sigma}^{-1/2}\sqrt{n}(\widehat{\theta} - \theta) \rightsquigarrow \mathcal{N}(0, I) \tag{3.3}$$

for inferences purposes, that is, to construct a confidence interval for θ and smooth trasformations thereof, or to carry out statistical hypothesis testing in the usual way.

However, motivated by their alternative asymptotic approximation, Cattaneo *et al.* (2014a) showed that

$$\frac{1}{n}\widehat{\Sigma} = \frac{1}{n}[\Sigma + o_{\mathbb{P}}(1)] + 2\binom{n}{2}^{-1}h^{-d-2}[\Delta + o_{\mathbb{P}}(1)],$$

which implies that the consistency result $\widehat{\Sigma} \to_{\mathbb{P}} \Sigma$ is valid if and only if $nh^{d+2} \to \infty$; otherwise, $\widehat{\Sigma}$ is in general asymptotically upwards biased relative to $\mathbb{V}[\widehat{\theta}]$ in (3.2). Because $\widehat{\Sigma}$ is asymptotically equivalent to the jackknife variance estimator of $\widehat{\theta}$, Cattaneo *et al.* (2014b) also noted that the asymptotic bias of $\widehat{\Sigma}$ is a result of a more generic phenomena underlying jackknife variance estimators studied in Efron and Stein (1981). See also Matsushita and Otsu (2021) for related discussion.

To conduct asymptotically valid inference under the more general small bandwidth asymptotic regime, Cattaneo *et al.* (2014a) proposed several "debiased" variance estimators, including the following

$$\widehat{V} = \frac{1}{n}\widehat{\Sigma} - \binom{n}{2}^{-1} h^{-d-2}\widehat{\Delta}, \qquad \widehat{\Delta} = h^{d+2}\binom{n}{2}^{-1} \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} U_{ij}U_{ij}',$$

and show that $\widehat{\Delta} \to_{\mathbb{P}} \Delta$ under the same bandwidth sequences ($nh^{2\min(P,S)} \to 0$ and $n^2h^d \to \infty$) required for (3.2) to hold. The estimator $\widehat{\Delta}$ is asymptotically equivalent to the debiasing procedure proposed in Efron and Stein (1981). This result justifies employing the Studentized statistic

$$\widehat{V}^{-1/2}(\widehat{\theta} - \theta) \rightsquigarrow \mathcal{N}(0, I) \tag{3.4}$$

for more "robust" inferences purposes relative to those constructed using (3.3).

Heuristically, robustness manifests in two distinct ways. First, the underlying Gaussian distributional approximation holds under weaker bandwidth restrictions and does not require asymptotic linearity, thereby making the limiting distribution explicitly depend on tuning parameter choices. Second, the new standard error formula $\widehat{V}$ is derived from the more general small bandwidth approximation and make explicit the contribution of terms regarded as higher-order by classical large sample distributional approximations.

While not reproduced here to conserve space, the in-depth Monte Carlo evidence reported in Cattaneo *et al.* (2010, 2014a,b) also showed that employing inference procedures based on (3.4) lead to large improvements in terms of "robustness" to bandwidth choice and other tuning inputs, when compared to classical asymptotically linear inference procedures based on (3.3). Theorem 2 below will study those two feasible statistics and show formally that the distributional approximation (3.4) has demonstrably smaller higher-order errors than the distributional approximation (3.3).

4 Higher-order Distribution Theory

We present Edgeworth expansions for scalar standarized and studentized statistics based on $\widehat{\theta}_\nu - \theta_\nu$ with $\widehat{\theta}_\nu := \nu'\widehat{\theta}$ and $\theta_\nu := \nu'\theta$, where $\nu \in \mathbb{R}^d$ is a fixed non-random vector. Considering scalar statistics substantially simplify the developments and proofs without affecting the main conceptual and theoretical takeaways. The sequence ϑ will first be non-random, thereby allowing us to investigate the role of classical distributional approximations based on asymptotic linearity vis-à-vis the more general distributional approximations based on small bandwidth asymptotics for standarized

statistics. Then, the sequence ϑ will be taken to be random based on the two alternative variance estimators introduced in the previous section, thereby allowing us to investigate the role of variance estimation on the performance of distributional approximations for Studentized statistics.

4.1 Distributional Approximation

Our first theorem offers a valid Edgeworth expansion for the sampling distribution function

$$F_\vartheta(t) := \mathbb{P}\left[\frac{\widehat{\theta}_\nu - \theta_\nu}{\vartheta} \le t\right], \qquad t \in \mathbb{R},$$

with precise characterization of the first three cummulants determining the leading errors in distributional approximation of the Studentized statistic. Define the following key quantities:

$$\beta := 2(-1)^P \sum_{[k]=P} \frac{\mu_k}{k!} \mathbb{E}\Big[g(X)\frac{\partial^k}{\partial X^k}\nu'\dot{f}(X)\Big], \qquad \sigma^2 := \mathbb{V}[\widehat{\theta}_\nu],$$

$$\kappa_1 := \mathbb{E}[\nu'\psi(Z)^3], \qquad \kappa_2 := 4\mathbb{E}[\delta(Z)\dot{\eta}(Z)] - 8\mathbb{E}[\delta(Z)^2]\theta_\nu + 4\theta_\nu^3,$$

where $\delta(Z) := \nu'\psi(Z)/2 + \theta_\nu$ and $\eta(Z_2) = \lim_{n\to\infty}\mathbb{E}[\delta(Z_1)\nu'U_{12}|Z_2]$.

Theorem 1 (Standardized). *Suppose Assumptions 1 and 2 hold. If $\sqrt{n}h^P \to 0$ and $nh^{d+2} \to \infty$, then for any positive non-random sequence ϑ such that $\vartheta/\sigma \to 1$,*

$$\sup_{t\in\mathbb{R}} \big|F_\vartheta(t) - G_\vartheta(t)\big| = O(R_n) + o(n^{-1/2})$$

with

$$G_\vartheta(t) := \Phi(t) - \phi(t)\Big\{\frac{\beta}{\vartheta}h^P + \Big(\frac{\sigma^2}{\vartheta^2} - 1\Big) + \frac{\kappa_1 + \kappa_2}{6n^2\vartheta^3}(t^2 - 1)\Big\},$$

and $R_n := nh^{2P} + \Big(\frac{(\log n)^3}{nh^{d+2}}\Big)^{3/2} + \frac{h^{d/3+1}}{nh^{d+2}} + \Big(\frac{h^{d/9+2/3}}{nh^{d+2}}\Big)^{3/2}$, where Φ and ϕ are the c.d.f. and p.d.f. of a standard Gaussian distribution. Furthermore, if $\frac{(\log n)^3}{nh^{d+2}} \to 0$, then $R_n = o\big(\sqrt{n}h^P + \frac{1}{nh^{d+2}}\big)$.

This theorem is proven by verifying the high-level conditions of a result in Appendix A establishing a valid Edgeworth Expansion for a generic class of U-statistics with n-varying kernels, which may be of independent theoretical interest. Specifically, Theorem A.1 and its corollary A.1 improve on Jing and Wang (2003) by allowing for n-varying kernels under more general condition suitable for the semiparametric problem of interest herein. Theorem 1 also improves on Nishiyama and Robinson (2000, Theorem 1) in two respects: (i) it allows for a generic standardization scheme ϑ instead of their specific choice $\sqrt{\nu'\Sigma\nu/n}$; and (ii) it presents a valid Edgeworth expansion with precise error rates with respect to the bandwidth. These improvements enable us to compare the two different distributional approximations of interest, (3.1) vs. (3.2).

The main conclusion in Theorem 1 follows the expected logic underlying Edgeworth Expansions: $\frac{\beta}{\vartheta}h^P$, $\frac{\sigma^2}{\vartheta^2} - 1$ and $\frac{\kappa_1+\kappa_2}{6n^2\vartheta^3}$ capture, respectively, the standardized bias, variance and higher moments of

9

the statistic. Inspection of these terms lead to interesting implications for large sample distribution theory, in particular leading to a sharp contrast between distribution theory based on asymptotic linear representations vis-à-vis alternative asymptotics, each with either fixed-bandwidth or leading asymptotic variance standardization. More specifically, we can consider four distinct standarization schemes: from first-order asymptotic linear theory (3.1) we have

$$\vartheta_{\text{AL}}^2 := \mathbb{V}[\nu'\bar{L}] = \frac{1}{n}\mathbb{V}[\nu'L_i] \qquad \text{and} \qquad \breve{\vartheta}_{\text{AL}}^2 := \frac{1}{n}\nu'\Sigma\nu,$$

while from small bandwidth distribution theory (3.2) we have

$$\vartheta_{\text{SB}}^2 := \mathbb{V}[\widehat{\theta}_\nu] = \sigma^2 \qquad \text{and} \qquad \breve{\vartheta}_{\text{SB}}^2 := \frac{1}{n}\nu'\Sigma\nu + \binom{n}{2}^{-1}h^{-d-2}\nu'\Delta\nu.$$

The standardizations ϑ_{AL} and ϑ_{SB} correspond to those constructed using the pre-asymptotic variance of the point estimator, each justified according to the asymptotic regime considered (asymptotic linear and small bandwidth, respectively). In contrast, the standardizations $\breve{\vartheta}_{\text{AL}}$ and $\breve{\vartheta}_{\text{SB}}$ correspond to employing the leading term only in the large sample approximation of the pre-asymptotic variance of the point estimator, again keeping only those terms that are justified by the asymptotic regime considered. That is, $\vartheta_{\text{AL}} = \breve{\vartheta}_{\text{AL}} + o(n^{-1})$ and $\vartheta_{\text{SB}} = \breve{\vartheta}_{\text{SB}} + o(n^{-1})$ under the assumptions of Theorem 1. For comparison, Nishiyama and Robinson (2000, Theorem 1) used $\breve{\vartheta}_{\text{AL}}$.

Employing Theorem 1 we can now compare the different approaches to standardization and their associated errors generated in the distributional approximation. Firstly, it is easy to see that employing $\breve{\vartheta}_{\text{AL}}$ and $\breve{\vartheta}_{\text{SB}}$ will generate larger distributional approximation errors relative to their pre-asymptotic counterparts, ϑ_{AL} and ϑ_{SB}, respectively. See the proof in the appendix for exact rates, which are not reproduced here to conserve space. The main conceptual message is that one should always employ variance formulas that capture the full variability of the statistic whenever possible, as opposed to employing those that capture only the leading variability in large samples. See Calonico et al. (2018, 2022) for closely related results in the context of nonparametric kernel-based density and local polynomial regression estimation and inference.

Secondly, and more importantly for our purposes, Theorem 1 shows that even if the full finite-sample variance of the point estimator is captured for standardization purposes, it is still crucial to incorporate the variability of both the linear and quadratic terms. More precisely, setting $\vartheta = \vartheta_{\text{AL}}$ then $\frac{\sigma^2}{\vartheta^2} - 1 = O(n^{-1}h^{-d-2})$, while setting $\vartheta = \vartheta_{\text{SB}}$ implies that $\frac{\sigma^2}{\vartheta^2} - 1 = 0$. As a consequence, our first main result shows that employing the pre-asymptotic variance of the statistic, which is naturally justified by the more general asymptotic distributional approximation (3.2), leads to the smallest error in the distributional approximation of the sampling distribution of the standardized statistic. This result thus provides theory-based evidence in favor of employing small bandwidth asymptotics for kernel-based DWAD methods whenever the goal is to minimize errors of inference procedures relying on large sample Gaussian approximations.

The methodological implications of our first theoretical result can be illustrated by analyzing the

coverage error of standardized confidence intervals. According to Theorem 1, for any $\alpha \in (0,1)$, a $100(1-\alpha)\%$ two-sided confidence interval based on asymptotic linearity satisfy

$$\mathbb{P}\left[\theta_\nu \in \left[\widehat{\theta}_\nu \pm \Phi_{1-\alpha/2}\vartheta_{\mathtt{AL}}\right]\right] = 1 - \alpha + \frac{\mathfrak{K}_{\mathtt{AL}}}{nh^{d+2}} + o\left(\sqrt{n}h^P + n^{-1}h^{-d-2} + n^{-1/2}\right),$$

where $\Phi_\alpha = \Phi^{-1}(\alpha)$, and $\mathfrak{K}_{\mathtt{AL}} = 2\Phi_{1-\alpha/2}\phi(1-\alpha/2)n^{-1}h^{-d-2}(\sigma^2/\vartheta_{\mathtt{AL}}^2 - 1) = O(1 + h^2)$, with the exact form of the leading terms described in the appendix. On the other hand, under the conditions in Theorem 1, a $100(1-\alpha)\%$ two-sided confidence intervals based on small bandwidth asymptotics satisfy

$$\mathbb{P}\left[\theta_\nu \in \left[\widehat{\theta}_\nu \pm \Phi_{1-\alpha/2}\vartheta_{\mathtt{SB}}\right]\right] = 1 - \alpha + o\left(\sqrt{n}h^P + n^{-1}h^{-d-2} + n^{-1/2}\right),$$

implying a smaller coverage error distortion in large samples.

The above coverage error comparison is conceptually useful, but it does not directly translate to practice because the confidence intervals are infeasible. To complement the results in this section, we consider next the implications of constructing variance estimators and hence study feasible (Studentized) inference procedures.

4.2 Variance Estimation

We study the role of Studentization and thus obtain valid Edgeworth expansion for the sampling distribution functions

$$F_{\mathtt{AL}}(t) := \mathbb{P}\left[\frac{\widehat{\theta}_\nu - \theta_\nu}{\widehat{\vartheta}_{\mathtt{AL}}} \leq t\right], \qquad \widehat{\vartheta}_{\mathtt{AL}} := \frac{1}{n}\nu'\widehat{\Sigma}\nu$$

and

$$F_{\mathtt{SB}}(t) := \mathbb{P}\left[\frac{\widehat{\theta}_\nu - \theta_\nu}{\widehat{\vartheta}_{\mathtt{SB}}} \leq t\right], \qquad \widehat{\vartheta}_{\mathtt{SB}} := \frac{1}{n}\nu'\widehat{\Sigma}\nu - \binom{n}{2}^{-1}h^{-d-2}\nu'\widehat{\Delta}\nu.$$

Crucially, the estimators $\widehat{\Sigma}$ and $\widehat{\Delta}$ target the total variability $n\mathbb{V}[\bar{L}] = \mathbb{V}[L_i]$ and $\binom{n}{2}h^{d+2}\mathbb{V}[\bar{Q}] = h^{d+2}\mathbb{V}[Q_{ij}]$, respectively, and not just their leading quantities Σ and Δ. Therefore, in light of the results reported in the previous section, we do not explicitly consider naïve plug-in estimators of $\breve{\vartheta}_{\mathtt{AL}}$ and $\breve{\vartheta}_{\mathtt{SBA}}$ such as $\frac{2}{n^2}\sum_{i=1}^n(\nu'[\widehat{\dot{e}}(X_i) - y\widehat{\dot{f}}(X_i) - \widehat{\theta}])^2$ for the former, where $\widehat{\dot{e}}(x)$ and $\widehat{\dot{f}}(x)$ are plug-in nonparametric estimators of $\dot{e}(x)$ and $\dot{f}(x)$, respectively. These alternative Studentization schemes will lead to larger higher-order distributional approximation errors when compared to $\widehat{\vartheta}_{\mathtt{AL}}$ and $\widehat{\vartheta}_{\mathtt{SB}}$.

Theorem 2 (Studentized). *Suppose Assumptions 1 and 2 hold with $p \geq 8$. If $\sqrt{n}h^P \to 0$ and $nh^{d+2}/(\log n)^9 \to \infty$, then*

$$\sup_{t\in\mathbb{R}} \left|F_{\mathtt{AL}}(t) - G_{\mathtt{AL}}(t)\right| = o(r_n)$$

with

$$G_{\mathtt{AL}}(t) := \Phi(t) - \phi(t)\left\{\frac{\sqrt{n}h^P\beta}{\nu'\Sigma\nu} - \frac{1}{nh^{d+2}}\frac{\nu'\Delta\nu}{\nu'\Sigma\nu}t - \frac{1}{\sqrt{n}6(\nu'\Sigma\nu)^3}\left[\kappa_1(2t^2+1) + \kappa_2(t^2+1)\right]\right\},$$

11

and

$$\sup_{t\in\mathbb{R}} \left| F_{\mathtt{SB}}(t) - G_{\mathtt{SB}}(t) \right| = o(r_n)$$

with

$$G_{\mathtt{SB}}(t) := \Phi(t) - \phi(t)\left\{ \frac{\sqrt{n}h^P\beta}{\nu'\Sigma\nu} - \frac{1}{\sqrt{n}6(\nu'\Sigma\nu)^3}\left[\kappa_1(2t^2+1) + \kappa_2(t^2+1) \right] \right\},$$

where $r_n := \sqrt{n}h^P + n^{-1}h^{-d-2} + n^{-1/2}$

This theorem shows that employing Studentization based on small bandwidth asymptotics offers demonstrable improvements in terms of distributional approximations for the resulting feasible t-test. The main practical implication of our second result can again be illustrated by analyzing the coverage error of Studentized confidence intervals. According to Theorem 2, and as it was the case for stdentized confidence intervals, a $100(1-\alpha)\%$ two-sided confidence intervals based on asymptotic linearity satisfy

$$\mathbb{P}\left[\theta_\nu \in \left[\widehat{\theta}_\nu \pm \Phi_{1-\alpha/2}\widehat{\vartheta}_{\mathtt{AL}} \right] \right] = 1 - \alpha + \frac{1}{nh^{d+2}}2\Phi_{1-\alpha/2}\phi(1-\alpha/2)\frac{\nu'\Delta\nu}{\nu'\Sigma\nu} + o(r_n),$$

while, under the conditions in Theorem 2, a $100(1-\alpha)\%$ two-sided confidence intervals based on small bandwidth asymptotics satisfy

$$\mathbb{P}\left[\theta_\nu \in \left[\widehat{\theta}_\nu \pm \Phi_{1-\alpha/2}\widehat{\vartheta}_{\mathtt{SB}} \right] \right] = 1 - \alpha + o(r_n),$$

implying a smaller coverage error distortion in large samples. This result provides a theoretical justification to the simulation evidence reported in Cattaneo *et al.* (2014a,b, 2010) where feasible confidence intervals based on small bandwidth asymptotics were shown to offer better finite sample performance in terms of coverage error than their counterparts based classical asymptotic linear approximations.

5 Conclusion

Employing Edgeworth expansions, we study the higher-order properties of two alternative first-order distributional approximations and their associated inference procedures (e.g., confidence intervals) for the kernel-based DWAD estimator of Powell *et al.* (1989). We showed that small bandwidth asymptotics not only give demonstrable better distributional approximations than asymptotic linear approximations, but also justify employing a variance estimator for Studentization purposes that also improves the distributional approximation. The main take away from our results is that in two-step semiparametric settings and related problems, alternative asymptotic approximations that capture higher-order terms ignored by classic asymptotic linear approximation can deliver better distributional approximations and, by implication, better inference procedures with improved performance in finite samples.

While beyond the scope of this paper, it would be of interest to develop analogous Edgeworth

expansions for non-linear two-step semiparamtric procedures developed using alternative asymptotic approximations and resampling methods (Cattaneo *et al.*, 2013; Cattaneo and Jansson, 2018; Cattaneo *et al.*, 2019). For the special case of kernel-based DWAD estimators (a linear two-step kernle-based semiparametric estimator), Nishiyama and Robinson (2005) present results that could be contrasted with those obtained under under small bandwidth asymptotics (Cattaneo *et al.*, 2014b). We relegate such developments for future research due to the substantial amount of additional technical work required.

A Edgeworth Expansion for Second-Order U-Statistic

Consider the sequence of maps $(u_n : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, n \in \mathbb{N})$ where $u := u_n$ is symmetric in terms of the permutation of its two arguments for every $n \in \mathbb{N}$. Given a random sample $Z_1, \ldots, Z_n$ for $n \geq 2$ of the random variable Z taking values on $\mathbb{R}^d$, the object of interest in the second order U-statistics with an n-varying kernel given by

$$\bar{U} := \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n}^{n} u(Z_i, Z_j). \tag{A.1}$$

We drop the subscript n to simplify notation. By the Hoeffding decomposition,

$$\frac{\bar{U} - \theta}{\vartheta} = B + L + Q,$$

where $B := (\mathbb{E}u(Z_1, Z_2) - \theta)/\vartheta$, $L := \frac{1}{\vartheta n} \sum_{i=1}^{n} \ell_i$ and $Q := \frac{1}{\vartheta} \binom{n}{2}^{-1} \sum_{1 \leq i < j \leq n}^{n} q_{ij}$, where $\ell_i := \ell(Z_i)$ and $q_{ij} := q(Z_i, Z_j)$ with $\ell(Z_1) := 2[\mathbb{E}u(Z_1, Z_2|Z_1) - \mathbb{E}u(Z_1, Z_2)]$ and $q(Z_1, Z_2) := u(Z_1, Z_2) - \ell(Z_1)/2 - \ell(Z_2)/2 - \mathbb{E}u(Z_1, Z_2)$. Given the decomposition above,

$$\sigma^2 := \mathbb{V}[\bar{U}] = \frac{1}{n}\sigma_\ell^2 + \binom{n}{2}^{-1} \sigma_q^2, \tag{A.2}$$

where $\sigma_\ell^2 := \mathbb{E}\ell_1^2$ and $\sigma_q^2 := \mathbb{E}q_{12}^2$.

We establish a valid third-order Edgeworth expansion for the the sampling distribution of the centered and standardized version of $\bar{U}$:

$$F(t) := \mathbb{P}\left[\frac{\bar{U} - \theta}{\vartheta} \leq t\right], \qquad t \in \mathbb{R}, \tag{A.3}$$

where $\theta \in \mathbb{R}$ and $\vartheta > 0$ are non-random.

Theorem A.1. *Let the following conditions hold:*

(a) $\mathbb{E}\left[(\ell_1/\sigma_\ell)^3\right] = O(1)$ *and* $\mathbb{E}[|q_{12}|]^{2+\delta} < \infty$, *and* $\sigma_\ell > 0$

(b) $\frac{\sigma_q}{\sqrt{n}\sigma_\ell} \to 0$ *and* $\frac{\sigma}{\vartheta} \to 1$.

(c) $\limsup_{n\to\infty} \limsup_{|t|\to\infty} |\mathbb{E}\exp(\iota t \ell_1/\sigma_\ell)| < 1$.

Then, $\sup_{t\in\mathbb{R}} |F(t) - G(t)| = O(\mathcal{E}) + o(n^{-1/2})$ where G is the distribution function with characteristic function

$$\chi_G(t) := e^{\iota t B - \frac{t^2}{2}} \left[1 + \sum_{j=2}^{9} (\iota t)^j \gamma_j \right],$$

with $\iota := \sqrt{-1}$,

$$\gamma_2 = \frac{1}{2}\left(\frac{\sigma^2}{\vartheta^2} - 1\right), \quad \gamma_3 = \frac{1}{6\vartheta^3 n^2}\left[\mathbb{E}\ell_1^3 + 6\mathbb{E}\ell_1\ell_2 q_{12}\right], \quad \gamma_4 = \frac{1}{4\vartheta^2}\left(\frac{\sigma^2}{\vartheta^2} - 1\right)\binom{n}{2}^{-1}\sigma_q^2$$

$$\gamma_5 = \frac{1}{12 n^2 \vartheta^5}\left[\binom{n}{2}^{-1}(\mathbb{E}\ell_1^3)\sigma_q^2 + 6\vartheta^2\left(\frac{\sigma_\ell^2}{\vartheta^2 n} - 1\right)\mathbb{E}\ell_1\ell_2 q_{12}\right]$$

$$\gamma_6 = \frac{1}{6\vartheta^6 n^4}\left[(\mathbb{E}\ell_1^3)\mathbb{E}\ell_1\ell_2 q_{12} + 12\binom{n}{2}^{-2}\binom{n}{4}[\mathbb{E}\ell_1\ell_2 q_{12}]^2\right], \quad \gamma_7 = 0,$$

$$\gamma_8 = \frac{1}{4\vartheta^6 n^4}\left(\frac{\sigma_\ell^2}{\vartheta^2 n} - 1\right)\binom{n}{2}^{-2}\binom{n}{4}[\mathbb{E}\ell_1\ell_2 q_{12}]^2, \quad \gamma_9 = \frac{1}{12\vartheta^9 n^6}\binom{n}{2}^{-2}\binom{n}{4}\mathbb{E}\ell_1^3[\mathbb{E}\ell_1\ell_2 q_{12}]^2,$$

and

$$\mathcal{E} := \left(\frac{\log n}{n^{3/2}\sigma_\ell}\right)^{2+\delta}\Pi_{2+\delta}(n) + \left(\frac{(\log n)^{\frac{4+\delta}{2+\delta}}\sigma_q^2}{n\sigma_\ell^2}\right)^{\frac{2+\delta}{2}} + \left(\frac{\log n}{n\sigma_\ell}\right)^{2+\delta}\Pi_{2+\delta}(\log n)$$

$$+ \frac{1}{\sigma_\ell^4 n}\mathbb{E}|\ell_1^2\ell_2 q_{12}| + \frac{1}{\sigma_\ell^5 n^{3/2}}\mathbb{E}|\ell_1^2\ell_2^2 q_{12}| + \frac{1}{\sigma_\ell^2 n^2}\mathbb{E}|\ell_1 q_{12}^2| + \frac{1}{\sigma_\ell^5 n^{3/2}}\mathbb{E}|\ell_1\ell_2\ell_3 q_{13}q_{23}|$$

$$+ \frac{1}{\sigma_\ell^7 n^{3/2}}(\mathbb{E}\ell_1\ell_2 q_{12})(\mathbb{E}|\ell_1^2\ell_2 q_{12}|) + \frac{1}{\sigma_\ell^8 n^2}(\mathbb{E}\ell_1\ell_2 q_{12})(\mathbb{E}|\ell_1^2\ell_2^2 q_{12}|),$$

with $\Pi_{2+\delta}(m) := \mathbb{E}|\sum_{i=1}^{[m]-1}\sum_{j=i+1}^{n} q_{ij}|^{2+\delta}$ for real $m > 1$ and $[\cdot]$ denoting the floor operator.

Corollary A.1. *Let the assumptions of Theorem A.1 hold. If $B \to 0$, then*

$$\sup_{t\in\mathbb{R}} |F(t) - G(t)| = O\left(B^2 + \mathcal{E}\right) + o(n^{-1/2}),$$

with

$$\chi_G(t) := e^{-\frac{t^2}{2}}\left[1 + B\iota t + \sum_{j=2}^{9}\left(\frac{\iota t}{\vartheta}\right)^j \gamma_j\right].$$

Remark A.1. Lemma A.2 below gives the following simpler bound

$$\Pi_{2+\delta}(m) \lesssim (nm\sigma_q^2)^{(2+\delta)/2} \vee mn^{1+\delta/2}\mathbb{E}\left[(\mathbb{E}(q_{12}^2|Z_1))^{1+\delta/2}\right] \vee nm\mathbb{E}|q_{12}|^{2+\delta},$$

where $\lesssim$ denotes bounded up to a fixed constant, and $a \vee b = \max\{a, b\}$.

⌐

Remark A.2. We can invert the characteristic function above to obtain a close form for F using the fact that for non-negative integer k, $\frac{1}{2\pi}\int_{\mathbb{R}}\exp\left(-\iota tx - t^2/2\right)(\iota t)^k \mathrm{d}t = H_k(x)\phi(x)$, where $H_k(x)$ is the k-th order Hermite polynomial (e.g., $H_0(k) = 1$, $H_1(x) = x$, $H_2(x) = x^2 - 1$, $H_3(x) = x^3 - 3x$). Therefore, the distribution function of $\chi_G(t)$ from Corollary A.1 is

$$G(x) = \Phi(x) - \phi(x)\left[\sum_{j=1}^{9}\gamma_j H_{j-1}(x)\right].$$

⌟

Remark A.3. To compare to Jing and Wang (2003), let $u(\cdot,\cdot)$ not dependent on n, $\theta = \mathbb{E}u(Z_1, Z_2)$, $\vartheta^2 = \sigma_\ell^2/n$, and $\mathbb{E}|q_{12}|^{2+\delta}$ bounded. Then, $\mathcal{E} = o(n^{-1/2})$ and $\chi_G(t) = \exp(-t^2/2)\left(1 - \frac{\iota\kappa_3 t^3}{6\sqrt{n}}\right) + o(n^{-1/2})$, giving

$$G(x) = \Phi(x) - \phi(x)\frac{1}{6\sqrt{n}}\left[\mathbb{E}\left(\frac{\ell_i}{\sigma_\ell}\right)^3 + \frac{6\mathbb{E}\ell_1\ell_2 q_{12}}{\sigma_\ell^3}\right](x^2 - 1).$$

⌟

A.1   Proof of Theorem A.1

Let χ_F denote the characteristic function F and g be the density of G. Using the well-known "smoothing inequality" (Bhattacharya and Rao, 1976; Hall, 1992), we write

$$\rho(F, G) \leq \frac{1}{\pi}\left[\int_{-\upsilon}^{\upsilon}\left|\frac{\chi_F(t) - \chi_G(t)}{t}\right|\mathrm{d}t + \frac{24\sup_{x\in\mathbb{R}}|g(x)|}{\upsilon}\right], \qquad \upsilon > 0$$

where ρ is the Kolmogorov distance. We set $\upsilon = \sqrt{n}\log n$ and split the range of integration into "low" frequencies and "high" frequencies. By the triangle inequality,

$$\rho(F, G) \lesssim I_1 + I_2 + I_3 + I_4 + \frac{1}{\sqrt{n}\log n}, \tag{A.4}$$

where

$$I_1 := \int_{|t|\leq\log n}\left|\frac{\chi_F(t) - \chi_G(t)}{t}\right|\mathrm{d}t, \qquad I_2 := \int_{\log n < |t| \leq c\sqrt{n}}\left|\frac{\chi_F(t)}{t}\right|\mathrm{d}t,$$

$$I_3 := \int_{c\sqrt{n} < |t| \leq \sqrt{n}\log n}\left|\frac{\chi_F(t)}{t}\right|\mathrm{d}t, \qquad I_4 := \int_{|t| > \log n}\left|\frac{\chi_G(t)}{t}\right|\mathrm{d}t;$$

Moreover, $c > 0$ is a fixed constant to be specified later.

We now bound each of these integrals in turn. We use extensively the fact hat

$$\left|\exp(\iota x) - \sum_{j=0}^{2}\frac{(\iota x)^j}{j!}\right| \leq |x|^{2+\delta} \quad , \quad \forall\delta \in [0,1]. \tag{A.5}$$

15

Also, define for $\psi(t) := \mathbb{E}\exp(\iota t\ell_1)$ for $t \in \mathbb{R}$ where σ_ℓ is positive by Assumption (a).

Bound for I_1

We start by decomposing $\chi_F(t) = \mathbb{E}\exp\left[\iota t(\frac{\bar{U}-\theta}{\vartheta})\right] = \exp(\iota tb)\chi_{L+Q}(t)$ where $\chi_{L+Q}(t) := \mathbb{E}\exp(\iota tL)\exp(\iota tQ)$. Use (A.5) to expand the second exponential in $\chi_{L+Q}(t)$ to write

$$\chi_{L+Q}(t) = \mathbb{E}\exp(\iota tL)[1 + \iota tQ - \tfrac{1}{2}(tQ)^2 + O((tQ)^{2+\delta})]. \tag{A.6}$$

Since $\ell_1, \ldots, \ell_n$ is a i.i.d sequence (for a given $n \geq 2$), the first term in (A.6) can be written as

$$\mathbb{E}\exp(\iota tL) = \mathbb{E}\exp\left(\frac{\iota t}{\vartheta n}\sum_{i=1}^{n}\ell_i\right) = \mathbb{E}\prod_{i=1}^{n}\exp\left(\frac{\iota t\ell_i}{\vartheta n}\right) = \prod_{i=1}^{n}\mathbb{E}\exp\left(\frac{\iota t\ell_i}{\vartheta n}\right) = \psi^n\left(\frac{t}{\vartheta n}\right).$$

For the second term in (A.6), we have

$$\mathbb{E}\exp(\iota tL)\iota tQ = \frac{\iota t}{\vartheta}\binom{n}{2}^{-1}\sum_{i<j}\mathbb{E}\prod_{k=1}^{n}\exp\left(\frac{\iota t}{\vartheta n}\ell_k\right)q_{ij}$$

$$= \frac{\iota t}{\vartheta}\binom{n}{2}^{-1}\sum_{i<j}\mathbb{E}\prod_{k\neq i,j}\exp\left(\frac{\iota t}{\vartheta n}\ell_k\right)\exp\left(\frac{\iota t}{\vartheta n}(\ell_i+\ell_j)\right)q_{ij}$$

$$= \frac{\iota t}{\vartheta}\binom{n}{2}^{-1}\sum_{i<j}\prod_{k\neq i,j}\mathbb{E}\exp\left(\frac{\iota t}{\vartheta n}\ell_k\right)\mathbb{E}\exp\left(\frac{\iota t}{\vartheta n}(\ell_i+\ell_j)\right)q_{ij}$$

$$= \frac{\iota t}{\vartheta}\psi^{n-2}\left(\tfrac{t}{\vartheta n}\right)\mathbb{E}\exp\left(\frac{\iota t}{\vartheta n}(\ell_1+\ell_2)\right)q_{12}.$$

Similarly, for the third term in (A.6), we use

$$\mathbb{E}\exp(\iota tL)(\iota tQ)^2 = \left[\frac{\iota t}{\vartheta}\binom{n}{2}^{-1}\right]^2 \times \left[\sum_{i<j}\mathbb{E}\prod_{k=1}^{n}\exp\left(\frac{\iota t}{\vartheta n}\ell_k\right)q_{ij}^2\right.$$

$$+ \sum_{i<j=k<l}\mathbb{E}\prod_{m\neq i,j,l}^{n}\exp\left(\frac{\iota t}{\vartheta n}\ell_m\right)q_{ij}q_{jl}$$

$$\left. + \sum_{i<j<k<l}\mathbb{E}\prod_{m\neq i,j,k,l}^{n}\exp\left(\frac{\iota t}{\vartheta n}\ell_m\right)q_{ij}q_{kl}\right]$$

$$= \left[\frac{\iota t}{\vartheta}\binom{n}{2}^{-1}\right]^2 \times \left[\psi^{n-2}\left(\tfrac{t}{\vartheta n}\right)\binom{n}{2}\mathbb{E}\exp\left(\frac{\iota t}{\vartheta n}(\ell_1+\ell_2)\right)q_{12}^2\right.$$

$$+ \psi^{n-3}\left(\tfrac{t}{\vartheta n}\right)\binom{n}{3}\mathbb{E}\exp\left(\frac{\iota t}{\vartheta n}(\ell_1+\ell_2+\ell_3)\right)q_{12}q_{23}$$

$$\left. + \psi^{n-4}\left(\tfrac{t}{\vartheta n}\right)\binom{n}{4}\left(\mathbb{E}\exp\left(\frac{\iota t}{\vartheta n}(\ell_1+\ell_2)\right)q_{12}\right)^2\right].$$

16

For the last in (A.6), we have

$$\left|\mathbb{E}\exp(\iota t L)(tQ)^{2+\delta}\right| \le \mathbb{E}|tQ|^{2+\delta} = \left[\frac{|t|}{\vartheta}\binom{n}{2}^{-1}\right]^{2+\delta}\mathbb{E}\left|\sum_{i<j}q_{ij}\right|^{2+\delta} = O\left(\left(\frac{|t|}{\vartheta n^2}\right)^{2+\delta}\Pi_{2+\delta}(n)\right).$$

Using the last four displays, we simplify (A.6) to

$$\begin{aligned}
\chi_{L+Q}(t) = {}& \psi^n\left(\tfrac{t}{\vartheta n}\right)\\
&+ \psi^{n-2}\left(\tfrac{t}{\vartheta n}\right)\left[\frac{\iota t}{\vartheta}\mathbb{E}\exp(\tfrac{\iota t}{\vartheta n}(\ell_1+\ell_2))q_{12} + \frac{(it)^2}{2\vartheta^2}\binom{n}{2}^{-1}\mathbb{E}\exp(\tfrac{\iota t}{\vartheta n}(\ell_1+\ell_2))q_{12}^2\right]\\
&+ \frac{1}{2}\left[\frac{\iota t}{\vartheta}\binom{n}{2}^{-1}\right]^2\psi^{n-3}\left(\tfrac{t}{\vartheta n}\right)\binom{n}{3}\mathbb{E}\exp(\tfrac{\iota t}{\vartheta n}(\ell_1+\ell_2+\ell_3))q_{13}q_{23}\\
&+ \frac{1}{2}\left[\frac{\iota t}{\vartheta}\binom{n}{2}^{-1}\right]^2\psi^{n-4}\left(\tfrac{t}{\vartheta n}\right)\binom{n}{4}\left[\mathbb{E}\exp(\tfrac{\iota t}{\vartheta n}(\ell_1+\ell_2))q_{12}\right]^2\\
&+ O\left[\left(\tfrac{|t|}{\vartheta n^2}\right)^{2+\delta}\Pi_{2+\delta}(n)\right]. \tag{A.7}
\end{aligned}$$

We now expand the exponentials inside the expectation and collect terms. For notation brevity, write $a := \frac{\iota t}{\vartheta n}$. For the first one, we have

$$\begin{aligned}
\mathbb{E}\exp(a(\ell_1+\ell_2))q_{12} = {}& \mathbb{E}\big(\exp(a\ell_1)-1\big)\big(\exp(a\ell_2)-1\big)q_{12}\\
= {}& \mathbb{E}\big[\big(\exp(a\ell_1)-1-a\ell_1\big)\big(\exp(a\ell_2)-1-a\ell_2\big)q_{12}\\
& + a\ell_1\big(\exp(a\ell_2)-1-a\ell_2\big)q_{12} + a\ell_2\big(\exp(a\ell_1)-1-a\ell_1\big)q_{12} + a^2\ell_1\ell_2 q_{12}\big]\\
= {}& a^2\mathbb{E}\ell_1\ell_2 q_{12} + O\left(|a|^3\mathbb{E}|\ell_1^2\ell_2 q_{12}| + |a|^4\mathbb{E}|\ell_1^2\ell_2^2 q_{12}|\right),
\end{aligned}$$

for the second term we have

$$\mathbb{E}\exp(a(\ell_1+\ell_2))q_{12}^2 = \sigma_q^2 + \mathbb{E}\big[\exp(a(\ell_1+\ell_2))-1\big]q_{12}^2 = \sigma_q^2 + O(|a|\mathbb{E}|\ell_1 q_{12}^2|),$$

and for the third term we have

$$\mathbb{E}\prod_{i=1}^{3}\exp(a\ell_i)q_{13}q_{23} = \mathbb{E}\prod_{i=1}^{3}\big(\exp(a\ell_i)-1\big)q_{13}q_{23} = O(|a|^3\mathbb{E}|\ell_1\ell_2\ell_3 q_{13}q_{23}|).$$

Plugging the above expansions back into (A.7) yields

$$\begin{aligned}
\chi_{L+Q}(t) = {}& \psi^n\left(\tfrac{t}{\vartheta n}\right) + \psi^{n-2}\left(\tfrac{t}{\vartheta n}\right)\left[\frac{(\iota t)^3}{\vartheta^3 n^2}\mathbb{E}\ell_1\ell_2 q_{12} + \frac{(it)^2}{2\vartheta^2}\binom{n}{2}^{-1}\sigma_q^2\right]\\
&+ \psi^{n-4}\left(\tfrac{t}{\vartheta n}\right)\frac{1}{2}\frac{(\iota t)^6}{\vartheta^6 n^4}\binom{n}{2}^{-2}\binom{n}{4}\left[\mathbb{E}\ell_1\ell_2 q_{12}\right]^2
\end{aligned}$$

17

$$+ O\left(\psi^{n-2}\left(\tfrac{t}{\vartheta n}\right)\left[\tfrac{t^4}{\vartheta^4 n^3}\mathbb{E}|\ell_1^2\ell_2 q_{12}| + \tfrac{|t|^5}{\vartheta^5 n^4}\mathbb{E}|\ell_1^2\ell_2^2 q_{12}| + \tfrac{|t|^3}{\vartheta^2 n^3}\mathbb{E}|\ell_1 q_{12}^2|\right]\right)$$

$$+ O\left(\psi^{n-3}\left(\tfrac{t}{\vartheta n}\right)\tfrac{|t|^5}{\vartheta^5 n^4}\mathbb{E}|\ell_1\ell_2\ell_3 q_{13}q_{23}|\right)$$

$$+ O\left(\psi^{n-4}\left(\tfrac{t}{\vartheta n}\right)\left[\tfrac{|t|^7}{\vartheta^7 n^5}(\mathbb{E}\ell_1\ell_2 q_{12})(\mathbb{E}|\ell_1^2\ell_2 q_{12}|) + \tfrac{|t|^8}{\vartheta^8 n^6}(\mathbb{E}\ell_1\ell_2 q_{12})(\mathbb{E}|\ell_1^2\ell_2^2 q_{12}|)\right]\right)$$

$$+ O\left[\left(\tfrac{|t|}{\vartheta n^2}\right)^{2+\delta}\Pi_{2+\delta}(n)\right]. \tag{A.8}$$

From the Edgeworth expansion theory for sum off i.i.d random variables (Bhattacharya and Rao, 1976; Hall, 1992), we have for $|t| \leq \delta^*\sqrt{n}$ for some small enough $\delta^* > 0$

$$\psi^n\left(\frac{t}{\sigma_\ell\sqrt{n}}\right) = \exp\left[-\tfrac{1}{2}t^2\right]\left[1 - \frac{\iota t^3}{6\sqrt{n}}\mathbb{E}\left(\frac{\ell_1}{\sigma_\ell}\right)^3\right] + o\left(\frac{(|t|^3 + t^6)}{\sqrt{n}}\exp(-t^2/4)\right).$$

Let $\alpha_k := \frac{\sigma_\ell\sqrt{n-k}}{\vartheta n}$ for $k \in \{0,2,3,4\}$. Since $\alpha_k \asymp 1$ by assumption, where $\asymp$ denotes proportional up to a fixed finite positive constant, we obtain

$$\psi^{n-k}\left(\frac{t}{\vartheta n}\right) = \psi^{n-k}\left(\frac{\alpha_k t}{\sigma_\ell\sqrt{n-k}}\right)$$

$$= \exp\left[-\tfrac{1}{2}\left(\alpha_k t\right)^2\right]\left[1 - \frac{\iota(\alpha_k t)^3}{6\sqrt{n-k}}\mathbb{E}\left(\frac{\ell_1}{\sigma_\ell}\right)^3\right] + o\left(\frac{(|t|^3 + t^6)}{\sqrt{n}}\exp(-(\alpha_k t)^2/4)\right).$$

A first-order Taylor expansion yields

$$\exp(-(\alpha_k t)^2/2) = \exp(-t^2/2)\left[1 - (\alpha_k^2 - 1)\tfrac{t^2}{2} + O(p(t)(\alpha_k^2 - 1)^2)\right],$$

and plugging it back in the previous expression, we have

$$\psi^{n-k}\left(\frac{t}{\vartheta n}\right) = \exp\left(-\tfrac{t^2}{2}\right)\left[1 - (\alpha_k^2 - 1)\frac{t^2}{2} - \frac{\iota(\alpha_k t)^3}{6\sqrt{n-k}}\mathbb{E}\left(\frac{\ell_1}{\sigma_\ell}\right)^3\right]$$

$$+ O\left((\alpha_k^2 - 1)^2 p(t)\exp\left(-t^2/2\right)\right) + o\left[\frac{(|t|^3 + t^6)}{\sqrt{n}}\exp(-(\alpha_k t)^2/4)\right].$$

Use the fact that $\alpha_k^2 = \alpha_0^2(1 - k/n) = \left(\frac{\sigma_\ell}{\vartheta\sqrt{n}}\right)^2 + O(n^{-1})$ to conclude that

$$\psi^{n-k}\left(\frac{t}{\vartheta n}\right) = \exp\left(-\tfrac{t^2}{2}\right)\left[1 - \left(\frac{\sigma_\ell^2}{\vartheta^2 n} - 1\right)\frac{t^2}{2} - \frac{\iota t^3}{6\vartheta^3 n^2}\mathbb{E}\ell_1^3\right]$$

$$+ O\left(\left(\frac{\sigma_\ell^2}{\vartheta^2 n} - 1\right)^2 p(t)\exp\left(-t^2/2\right)\right) + o\left(\frac{(|t|^3 + t^6)}{\sqrt{n}}\exp(-t^2/4)\right), \tag{A.9}$$

for $|t| \leq \delta^*\sqrt{n}$.

Combine (A.8) and (A.9) to conclude that, for $|t| \leq \delta^* \sqrt{n}$,

$$\chi_{L+Q}(t) = \exp\left(-\frac{t^2}{2}\right)\left[1 - \left(\frac{\sigma_\ell^2}{\vartheta^2 n} - 1\right)\frac{t^2}{2} - \frac{\iota t^3}{6\vartheta^3 n^2}\mathbb{E}\ell_1^3\right]$$

$$\times \left[1 + \frac{(it)^2}{2\vartheta^2}\binom{n}{2}^{-1}\sigma_q^2 + \frac{(\iota t)^3}{\vartheta^3 n^2}\mathbb{E}\ell_1\ell_2 q_{12} + \frac{1}{2}\frac{(\iota t)^6}{\vartheta^6 n^4}\binom{n}{2}^{-2}\binom{n}{4}[\mathbb{E}\ell_1\ell_2 q_{12}]^2\right]$$

$$+ O\left(\exp\left(-\frac{t^2}{2}\right)\left[1 + \left(\frac{\sigma_\ell^2}{\vartheta^2 n} - 1\right)t^2 + \left(\frac{\sigma_\ell^2}{\vartheta^2 n} - 1\right)^2 p(|t|) + \frac{|t|^3}{\sqrt{n}}\right]R(t)\right)$$

$$+ o\left(\exp\left(-\frac{t^2}{4}\right)\left[\frac{|t|^3 + t^6}{\sqrt{n}}\right]R(t)\right)$$

$$+ O\left(\left(\frac{|t|}{\vartheta n^2}\right)^{2+\delta}\Pi_{2+\delta}(n)\right), \tag{A.10}$$

where

$$R(t) := \frac{t^4}{\vartheta^4 n^3}\mathbb{E}|\ell_1^2\ell_2 q_{12}| + \frac{|t|^5}{\vartheta^5 n^4}\mathbb{E}|\ell_1^2\ell_2^2 q_{12}| + \frac{|t|^3}{\vartheta^2 n^3}\mathbb{E}|\ell_1 q_{12}^2| + \frac{|t|^5}{\vartheta^5 n^4}\mathbb{E}|\ell_1\ell_2\ell_3 q_{13}q_{23}|$$

$$+ \frac{|t|^7}{\vartheta^7 n^5}(\mathbb{E}\ell_1\ell_2 q_{12})(\mathbb{E}|\ell_1^2\ell_2 q_{12}|) + \frac{|t|^8}{\vartheta^8 n^6}(\mathbb{E}\ell_1\ell_2 q_{12})(\mathbb{E}|\ell_1^2\ell_2^2 q_{12}|).$$

After some rearrangement, the first term in (A.10) becomes

$$\widetilde{\chi}_{L+Q}(t) := \exp\left(-\frac{t^2}{2}\right)P(t) = \exp\left(-\frac{t^2}{2}\right)\left[1 + \sum_{j=2}^{9}\left(\frac{\iota t}{\vartheta}\right)^j \gamma_j\right],$$

where

$$P(t) := 1 + \frac{(\iota t)^2}{2}\left(\frac{\sigma^2}{\vartheta^2} - 1\right) + \frac{(\iota t)^3}{6\vartheta^3 n^2}\left[\mathbb{E}\ell_1^3 + 6\mathbb{E}\ell_1\ell_2 q_{12}\right] + \frac{(\iota t)^4}{4\vartheta^2}\left(\frac{\sigma^2}{\vartheta^2} - 1\right)\binom{n}{2}^{-1}\sigma_q^2$$

$$+ \frac{(\iota t)^5}{12\vartheta^5 n^2}\left[\binom{n}{2}^{-1}(\mathbb{E}\ell_1^3)\sigma_q^2 + 6\vartheta^2\left(\frac{\sigma_\ell^2}{\vartheta^2 n} - 1\right)\mathbb{E}\ell_1\ell_2 q_{12}\right]$$

$$+ \frac{(\iota t)^6}{6\vartheta^6 n^4}\left[(\mathbb{E}\ell_1^3)\mathbb{E}\ell_1\ell_2 q_{12} + 12\binom{n}{2}^{-2}\binom{n}{4}[\mathbb{E}\ell_1\ell_2 q_{12}]^2\right]$$

$$+ \frac{1}{4}\frac{(\iota t)^8}{\vartheta^6 n^4}\left(\frac{\sigma_\ell^2}{\vartheta^2 n} - 1\right)\binom{n}{2}^{-2}\binom{n}{4}[\mathbb{E}\ell_1\ell_2 q_{12}]^2$$

$$+ \frac{1}{12}\frac{(\iota t)^9}{\vartheta^9 n^6}\binom{n}{2}^{-2}\binom{n}{4}\mathbb{E}\ell_1^3[\mathbb{E}\ell_1\ell_2 q_{12}]^2.$$

Since $\widetilde{\chi}_{L+Q}(t) = \exp(-\iota t b)\chi_G(t)$, we have under Assumption (a) and (b)

$$|\chi_F(t) - \chi_G(t)| = O\left(\exp\left(-\frac{t^2}{4}\right)R(t) + \left(\frac{|t|}{\vartheta n^2}\right)^{2+\delta}\Pi_{2+\delta}(n)\right).$$

Therefore,

$$
\begin{aligned}
I_1 &= O\left(\int_{|t|\leq \log n} |t|^{-1}\exp(-t^2/4)R(t)\mathrm{d}t + \frac{\Pi_{2+\delta}(n)}{(\vartheta n^2)^{2+\delta}}\int_{|t|\leq \log n}|t|^{1+\delta}\mathrm{d}t\right)\\
&= O\left(R(1) + \left(\frac{\log n}{\vartheta n^2}\right)^{2+\delta}\Pi_{2+\delta}(n)\right).
\end{aligned}
$$

Bound for I_2

For $1 \leq m < n$, define $Q_m := \frac{1}{\vartheta}\binom{n}{2}^{-1}\sum_{i=1}^{m}\sum_{j=i+1}^{n}q_{ij}$. Using (A.5) we can write

$$
|\chi_F(t)| = |\chi_{L+Q}(t)| \leq \left|\mathbb{E}\exp(\iota t(L+Q-Q_m))\sum_{k=0}^{2}\frac{(itQ_m)^k}{k!})\right| + |t|^{2+\delta}\mathbb{E}|Q_m|^{2+\delta}.
$$

Exploiting the fact that $Q - Q_m$ is only a function of $X_{m+1}, \ldots, X_n$, we have

$$
|\mathbb{E}\exp(\iota t(L+Q-Q_m))| \leq |\psi\left(\tfrac{t}{\vartheta n}\right)|^m.
$$

For the second term

$$
\begin{aligned}
|\mathbb{E}\exp(\iota t(T-Q_m)Q_m)| &= \frac{1}{\vartheta}\binom{n}{2}^{-1}\left|\sum_{i=1}^{m}\sum_{j=i+1}^{n}\mathbb{E}\exp(\iota t(L+Q-Q_m)q_{ij}\right|\\
&\lesssim \frac{1}{\vartheta n^2}|\psi\left(\tfrac{t}{\vartheta n}\right)|^{m-2}mn\mathbb{E}|q_{12}|.
\end{aligned}
$$

Similarly, using the fact that

$$
\left[\vartheta\binom{n}{2}Q_m\right]^2 = \sum_{i=1}^{m}\sum_{j=i+1}^{n}q_{ij}^2, + \sum_{i=1}^{m}\sum_{j=i+1}^{n}\sum_{\substack{k=1,\\k\neq i,j}}^{m}q_{ij}q_{jk} + \sum_{i=1}^{m}\sum_{j=i+1}^{n}\sum_{\substack{k=1,\\k\neq i,j}}^{m}\sum_{\substack{l=k+1,\\l\neq i,j}}^{n}q_{ij}q_{kl},
$$

we conclude for $k \in \{0, 1, 2\}$,

$$
|\mathbb{E}\exp(\iota t(T-Q_m))Q_m^k| \lesssim |\psi\left(\tfrac{t}{\vartheta n}\right)|^{m-2k}\left(\frac{mn}{\vartheta}\binom{n}{2}^{-1}\right)^k\mathbb{E}|q_{12}|^k.
$$

Finally, using the fact that $\vartheta = O(\sigma_\ell/\sqrt{n})$ by Assumption (b) and combining the last displays,

$$
|\chi_F(t)| \lesssim \sum_{k=0}^{2}\left(\frac{|t|m}{\sqrt{n}}\right)^k|\psi\left(\tfrac{t}{\vartheta n}\right)|^{m-2k}\mathbb{E}\left|\frac{q_{12}}{\sigma_\ell}\right|^k + |t|^{2+\delta}\mathbb{E}|Q_m|^{2+\delta}, \tag{A.11}
$$

for $1 \leq m < n$ and $\delta \in [0, 1]$.

By the triangle inequality followed by (A.5), we have

$$\left|\psi\left(\tfrac{t}{\vartheta n}\right)\right| - \left|1 - \tfrac{t^2}{2(\vartheta n)^2}\sigma_\ell^2\right| \le \left|\psi\left(\tfrac{t}{\vartheta n}\right) - \left(1 - \tfrac{t^2}{2(\vartheta n)^2}\sigma_\ell^2\right)\right| \le \tfrac{1}{6}\tfrac{|t|^3}{(\vartheta n)^3}\mathbb{E}|\ell_1|^3.$$

For $|t| \le \tfrac{\sqrt{2}\vartheta n}{\sigma_\ell}$ we have $\left|1 - \tfrac{t^2}{2(\vartheta n)^2}\sigma_\ell^2\right| = 1 - \tfrac{t^2}{2(\vartheta n)^2}\sigma_\ell^2$ hence

$$\left|\psi\left(\tfrac{t}{\vartheta n}\right)\right| \le 1 - \tfrac{t^2}{2(\vartheta n)^2}\sigma_\ell^2 + \tfrac{1}{6}\tfrac{|t|^3}{(\vartheta n)^3}\mathbb{E}|\ell_1|^3; \qquad |t| \le \tfrac{\sqrt{2}\vartheta n}{\sigma_\ell}.$$

Assumption (b) together with (A.2) implies that $\tfrac{\sigma_\ell}{\sqrt{n}\vartheta} \to 1$ as $n \to \infty$. Then, we can find a $N_1 \in \mathbb{N}$ such that $\sqrt{5/6} \le \tfrac{\sigma_\ell}{\sqrt{n}\vartheta} \le (6/5)^{1/3}$ for $n \ge N_1$. Also, Assumption (a) implies the existence of a constant $C > 0$ and $N_2 \in \mathbb{N}$ such that $\mathbb{E}|\ell_1/\sigma_\ell|^3 \le C$ for $n \ge N_2$. Then, for $|t| \le c\sqrt{n}$ where $c := (\sqrt{2}/(6/5)^{1/3}) \wedge (5/(12C))$ and $n \ge N_0 := N_1 \vee N_2$, we have

$$\left|\psi\left(\tfrac{t}{\vartheta n}\right)\right| \le 1 - \frac{t^2}{n}\left[\frac{1}{2}\left(\frac{\sigma_\ell}{\vartheta\sqrt{n}}\right)^2 - \frac{|t|\mathbb{E}|\ell_1/\sigma_\ell|^3}{6\sqrt{n}}\left(\frac{\sigma_\ell}{\vartheta\sqrt{n}}\right)^3\right] \le 1 - \frac{t^2}{3n} \le \exp(-\tfrac{t^2}{3n}). \qquad (\text{A.12})$$

For $\log n < |t| \le c\sqrt{n}$, set $m = [\tfrac{15n\log n}{t^2}] + 1 = O(n)$, then plug in (A.12) to conclude that $\left|\psi\left(\tfrac{t}{\vartheta n}\right)\right|^{m-2k} \lesssim \exp(-\tfrac{t^2 m}{3n}) \lesssim n^{-5}$. Combining this last bound with (A.11), we obtain

$$|\chi_F(t)| \lesssim \sum_{k=0}^{2} \frac{|t|^k}{n^{5-k}}\mathbb{E}\left|\tfrac{q_{12}}{\sigma_\ell}\right|^k + |t|^{2+\delta}\mathbb{E}|Q_m|^{2+\delta}, \qquad (\text{A.13})$$

for $|t| \le c\sqrt{n}$ and $n \ge N_0$. Then,

$$I_2 \lesssim \sum_{k=0}^{2} \frac{1}{n^{5-k-k/2}}\frac{\mathbb{E}|q_{12}|^k}{\sigma_\ell^k} + \left(\frac{\sqrt{n}\log n}{n}\right)^{2+\delta}\left(\tfrac{\sigma_q}{\sigma_\ell}\right)^{2+\delta}\log n$$

$$\lesssim \frac{1}{n}\frac{\sigma_q^2}{n\sigma_\ell^2} + \log n \left(\frac{(\log n)\sigma_q^2}{n\sigma_\ell^2}\right)^{\frac{2+\delta}{2}}.$$

Therefore, since $\tfrac{\sigma_q^2}{n\sigma_\ell^2} = o(1)$ by Assumption (b), we conclude

$$I_2 = o(n^{-1}) + O\left(\left[\frac{(\log n)^{\frac{4+\delta}{2+\delta}}\sigma_q^2}{n\sigma_\ell^2}\right]^{\frac{2+\delta}{2}}\right).$$

Bound for I_3 and I_4

Under Assumption (c), for sufficient large n, we may find a $b > 0$ such that for $|t| > c\sqrt{n}$

$$\left|\psi\left(\tfrac{t}{\vartheta n}\right)\right| \le 1 - b < \exp(-b),$$

where $c > 0$ is define just before (A.12). Set $m = [\frac{4\log n}{b}] + 1$, then $nm \lesssim n\log n$ and $|\psi\left(\frac{t}{\vartheta n}\right)^{m-s}| \lesssim n^{-4}$ for sufficient large n and $s \in \{1, 3, 4, 5\}$. Use these upper bounds on (A.11) to conclude that

$$|\chi_F(t)| \lesssim n^{-4}\left(1 + |t|\log n\mathbb{E}|q_{12}/\sigma_\ell| + t^2(\log n)^2\frac{\sigma_q^2}{\sigma_\ell^2}\right) + |t|^{2+\delta}(n\log n)^{1+\delta/2}\mathbb{E}|q_{12}|^{2+\delta}, \qquad (A.14)$$

for sufficient large n and $|t| > c\sqrt{n}$. Then,

$$I_3 = o(n^{-1/2}) + O\left((n\log n)^{1+\delta/2}\mathbb{E}|q_{12}|^{2+\delta}\int_{c\sqrt{n}\leq|t|\leq\sqrt{n}\log n}|t|^{1+\delta}dt\right)$$

$$= o(n^{-1/2}) + O\left(\left[\frac{\log n}{n}\right]^{2+\delta}\Pi_{2+\delta}(\log n)\right).$$

Finally,

$$I_4 = \int_{|t|>\log n}|t|^{-1}\exp(-\frac{t^2}{2})\left|1 + \sum_{j=2}^{9}\left(\frac{it}{\vartheta}\right)^j\gamma_j\right|dt$$

$$\leq C\int_{t>\log n}t^{-1}\exp(-\frac{t^2}{2})dt + \sum_{j=2}^{9}\frac{|\gamma_j|}{\vartheta^j}\int_{t>\log n}t^{j-1}\exp(-\frac{t^2}{2})dt,$$

where the first integral is $o(n^{-1})$ and the second is $o(1)$. Therefore,

$$I_4 = o\left(n^{-1} + \sum_{j=2}^{9}\frac{|\gamma_j|}{\vartheta^j}\right).$$

The proof is complete. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

A.2 Auxiliary Lemmas

Lemma A.1. *Let $n \geq 2$, $1 \leq l \leq m < n$ and $p \geq 2$ then*

$$\mathbb{E}\left|\sum_{i=l}^{m}\sum_{j=i+1}^{n}q_{ij}\right|^p \leq C_p(n-l)^{p/2}\max_{l<j\leq n}\mathbb{E}\left|\sum_{i=l}^{(m\wedge j)-1}q_{ij}\right|^p \leq K_p\left[(n-l)(m-l)\right]^{p/2}\mathbb{E}|q_{12}|^p,$$

where $C_p := \left[8(p-1)(1 \vee 2^{p-3})\right]^p$ and K_p is a constant only depending on p.

Proof. The double summation on the left-hand side can be written as $\sum_{j=l+1}^{n}\xi_j$ where $\xi_j := \sum_{i=l}^{(m\wedge j)-1}q_{ij}$. Notice that $\{\xi_j, \mathcal{F}_j\}$ is m.d.s when $\mathcal{F}_j$ is the σ-algebra generated by $\{X_1, \ldots, X_j\}$ for $j \geq 1$ and $\mathcal{F}_0$ is trivial. Then by Dharmadhikari *et al.* (1968) followed by a trivial bound

$$\mathbb{E}\left|\sum_{i=l}^{m}\sum_{j=i+1}^{n}q_{ij}\right|^p = \mathbb{E}\left|\sum_{j=l+1}^{n}\xi_j\right|^p \leq C_p(n-l)^{p/2-1}\sum_{j=l+1}^{n}\mathbb{E}|\xi_j|^p \leq C_p(n-l)^{p/2}\max_{l<j\leq n}\mathbb{E}|\xi_j|^p.$$

$\square$

Lemma A.2. *For $p \in [2, \infty)$ there exist a constant C_p only depending on p such that for $\mathcal{S} \subseteq \{(i,j) : 1 \leq i < j \leq n\}$*

$$\mathbb{E} \left| \sum_{\mathcal{S}} q_{ij} \right|^p \leq C_p \left\{ |\mathcal{S}|^{p/2} \left[\mathbb{E}q_{12}^2 \right]^{p/2} \vee s \mathbb{E} \left[(\mathbb{E}(q_{12}^2 | Z_1))^{p/2} \right] \vee |\mathcal{S}| \mathbb{E}|q_{12}|^p \right\},$$

where $|\mathcal{S}|$ denotes the cardinality of the set $\mathcal{S}$, $s := s_i \vee s_j$ with $s_i := \sum_{(i,\cdot) \in \mathcal{S}} \left(\sum_{(\cdot,j) \in \mathcal{S}} 1 \right)^{p/2}$ and $s_j := \sum_{(\cdot,j) \in \mathcal{S}} \left(\sum_{(i,\cdot) \in \mathcal{S}} 1 \right)^{p/2}$.

Proof. Combining Proposition 2.1 with expression (2.18) in Giné *et al.* (2000), we obtain the inequality above for the *decoupled version* of q_{ij}, defined as $\widetilde{q}_{ij} := q(Z_i^{(1)}, Z_j^{(2)})$ where $Z_i^{(j)} : 1 \leq i \leq n$, $1 \leq j \leq 2$ are i.i.d. Finally, we can apply the decoupling inequalities in de la Peña and Montgomery-Smith (1995) to obtain the result at the expense of increasing the constant without altering the order of the upper bound. For further details, see section 2.5 in Giné *et al.* (2000). $\square$

B Proof of Theorem 1 (Standardized Edgeworth Expansion)

We apply Corollary A.1 with $u(Z_i, Z_j) = \nu' U_{ij}$ in (2.1) and $\delta = 1$. We assume throughout that Assumptions 1 and 2 hold. Condition (a) in Theorem A.1 is verified by direct calculations as in Cattaneo *et al.* (2010, 2014a,b). Condition (b) in Theorem A.1 is verified because (A.2) gives $\sigma^2 = \frac{1}{n} \mathbb{V}[\nu' L_i] + \binom{n}{2}^{-1} \mathbb{V}[\nu' Q_{i,j}]$, which implies

$$\sigma_\ell^2 = \nu' \Sigma \nu + O(h^P) \qquad \text{and} \qquad \sigma_q^2 = \frac{1}{h^{d+2}} [\nu' \Delta \nu + h^2 \nu' \mathcal{V} \nu] + o(h^{-d}),$$

with $\mathcal{V}$ given in Cattaneo *et al.* (2010). These results imply $\sigma_q^2 = o(n\sigma_\ell^2)$ if (and only if) $nh^{d+2} \to \infty$. Therefore, we take $\vartheta \asymp \sigma \asymp 1/\sqrt{n}$. Condition (c) in Theorem A.1 holds by assumption.

The additional condition $B \to 0$ in Corollary A.1 holds if (and only if, when $\beta \neq 0$) $\sqrt{n}h^P \to 0$. To see this, using integration by parts, $\mathbb{E}[U_{12}|Z_1] = \int_{\mathbb{R}^d} \nu' \dot{e}(X_1 + uh) K(u) du - Y_1 \int_{\mathbb{R}^d} \nu' \dot{f}(X_1 + uh) K(u) du$. Then, repeated Taylor series expansions and integration by parts give $\mathbb{E}[u(Z_1, Z_2)|Z_1] = \delta(Z_1) + h^P (-1)^P \sum_{[k]=P} \frac{\mu_k}{k!} \delta^{(1+k)}(z) + o(h^P)$. In turn, this result implies that $\mathbb{E}[u(Z_1, Z_2)] = \theta_\nu + h^P \beta + o(h^P)$. As a consequence, $B = (\mathbb{E}[\widehat{\theta}_\nu] - \theta_\nu)/\vartheta = h^P \beta/\vartheta + o(\sqrt{n}h^P)$. See Cattaneo *et al.* (2010, 2014a,b) for details.

Law of iterated expectations, integration by parts, and Taylor series expansions give

$$\mathbb{E}[\ell_1^3] = \kappa_1 + O(h^P).$$

Proceeding analogously, because $\mathbb{E}[\ell_1 \ell_2 q_{12}] = \mathbb{E}[\ell_2 \mathbb{E}[\ell_1 q_{12} | Z_2]] = \mathbb{E}[\ell_2 \ell_1 U_{12}]$ and $\mathbb{E}[\ell_2 \ell_1 U_{12}] = 4\mathbb{E}[\mathbb{E}[U_{12}|Z_1]\mathbb{E}[U_{12}|Z_2]U_{12}] - 8\mathbb{E}[U_{12}]\mathbb{E}[\mathbb{E}[U_{12}|Z_1]^2] + 4\mathbb{E}[U_{12}]^3$, we have $\mathbb{E}[\mathbb{E}[U_{12}|Z_1]\mathbb{E}[U_{12}|Z_2]U_{12}] =$

23

$\mathbb{E}[\delta(Z)\dot{\eta}(Z)] + O(h^P)$, $\mathbb{E}[\mathbb{E}[U_{12}|Z_1]^2] = \mathbb{E}[\delta(Z_1)^2] + O(h^P)$, and $\mathbb{E}[U_{12}] = \theta + O(h^P)$ and $\mathbb{E}[U_{12}]^3 = \theta^3 + O(h^P)$. Collecting these results, we verify

$$\mathbb{E}[\ell_1\ell_2 q_{12}] = \kappa_2 + O(h^P).$$

These results imply $\gamma_3 = O(n^{-2})$, $\gamma_4 = O(n^{-3}h^{-d-2})$, $\gamma_5 = O(n^{-2})$, $\gamma_6 = O(n^{-1})$, $\gamma_8 = O(n^{-3})$, $\gamma_9 = O(n^{-7/2})$.

It remains to bound $\mathcal{E}$. First, by standard results $\mathbb{E}|q_{12}|^3 = O(h^{2d+3})$, so

$$\Pi_3(m) = O\left(\frac{(mn)^{3/2}}{h^{(3/2)(d+2)}} \vee \frac{mn^{3/2}}{h^{(3/2)(d+2)}} \vee \frac{mn}{h^{2d+3}}\right) = O\left(\frac{mn}{h^{d+2}}\right)^{3/2}.$$

Second, using the results in Cattaneo *et al.* (2014b, Supplemental Appendix), we have $\mathbb{E}|\ell_1^2\ell_2 q_{12}| = O(h^{-2d/3-1})$, $\mathbb{E}|\ell_1^2\ell_2^2 q_{12}| = O(h^{-2d/3-1})$, $\mathbb{E}|\ell_1 q_{12}^2| = O(h^{-d-2})$, $\mathbb{E}|\ell_1\ell_2\ell_3 q_{13}q_{23}| = O(h^{-4d/3-2})$. Thus, collecting all the bounds, we verify:

$$\mathcal{E} = O\left(\left(\frac{(\log n)^3}{nh^{d+2}}\right)^{3/2} + \frac{h^{d/3+1}}{nh^{d+2}} + \left(\frac{h^{d/9+2/3}}{nh^{d+2}}\right)^{3/2}\right) = o\left(\frac{1}{nh^{d+2}}\right)$$

This completes the proof. $\qquad\qquad\square$

C Proof of Theorem 2 (Studentized Edgeworth Expansion)

For any estimated scale $\widehat{\vartheta}$ and nonrandom centering ϑ, we have

$$\frac{\widehat{\theta}_\nu - \theta_\nu}{\widehat{\vartheta}} = \frac{\widehat{\theta}_\nu - \theta_\nu}{\vartheta}\left[1 - \frac{\widehat{\vartheta}^2 - \vartheta^2}{2\vartheta^2} + \frac{\widehat{\vartheta} + 2\vartheta^2}{2\vartheta^2\widehat{\vartheta}}\frac{(\widehat{\vartheta}^2 - \vartheta^2)^2}{\widehat{\vartheta}^2 + \vartheta^2}\right].$$

Recall that $\widehat{\theta}_\nu$ is a second-order U-statistic satisfying the H-decomposition $(\widehat{\theta}_\nu - \theta_\nu)/\vartheta = B + \bar{L}/\vartheta + \bar{Q}/\vartheta$. Using standard results for Edgeworth expansions (Bhattacharya and Rao, 1976; Hall, 1992),

$$\sup_{t\in\mathbb{R}}\left|\mathbb{P}\left[\frac{\widehat{\theta}_\nu - \theta_\nu}{\widehat{\vartheta}} \leq t\right] - G(t)\right| \leq E + \mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 + O\left(\frac{r_n}{\log n}\right),$$

where

$$E := \sup_{t\in\mathbb{R}}\left|\mathbb{P}\left[\left(1 - \frac{\widehat{\vartheta}^2 - \vartheta^2}{2\vartheta^2}\right)\left(\nu'\bar{L}/\vartheta + \nu'\bar{Q}/\vartheta\right) + B \leq t\right] - G(t)\right|,$$

$B = \nu'(\mathbb{E}[U_{12}] - \theta)/\vartheta$, G denoting a distribution function later to be set to either G_{AL} or G_{SB} as appropriate, and

$$\mathcal{R}_1 := \mathbb{P}\left[\left|\frac{\widehat{\vartheta} + 2\vartheta^2}{2\vartheta^2\widehat{\vartheta}}\right|\frac{(\widehat{\vartheta}^2 - \vartheta^2)^2}{\widehat{\vartheta}^2 + \vartheta^2} > C\frac{r_n}{(\log n)^2}\right],$$

$$\mathcal{R}_2 := \mathbb{P}\left[\left|\frac{\widehat{\theta}_\nu - \theta_\nu}{\vartheta}\right| > C\log n\right],$$

24

$$\mathcal{R}_3 := \mathbb{P}\left[\left| \frac{\widehat{\vartheta} - \vartheta}{\vartheta} B \right| > C \frac{\sqrt{n} h^P}{\log n} \right],$$

with C denoting a generic constant, which can take different values in different places. The term E will give the Edgeworth expansion upon setting $\widehat{\vartheta}$ and ϑ appropriately, while the terms $\mathcal{R}_1$–$\mathcal{R}_3$ capture higher-order remainders.

Variance Estimators

The estimators $\widehat{\vartheta}_{\texttt{AL}}^2$ and $\widehat{\vartheta}_{\texttt{SB}}^2$ are linear combinations of U-statistics as follows:

$$\widehat{\vartheta}_{\texttt{AL}}^2 = \frac{1}{n}\nu'\widehat{\Sigma}\nu = 2\binom{n}{2}^{-1}\bar{W}_1 + \frac{4}{n}\frac{n-2}{n-1}\bar{W}_2 - \frac{4}{n}\widehat{\theta}_\nu^2$$

and

$$\binom{n}{2}^{-1}h^{-d-2}\nu'\widehat{\Delta}\nu = \binom{n}{2}^{-1}\bar{W}_1$$

with

$$\widehat{\theta}_\nu = \binom{n}{2}^{-1}\sum_{i<j}(\nu'U_{ij}), \qquad \bar{W}_1 = \binom{n}{2}^{-1}\sum_{i<j}(\nu'U_{ij})^2,$$

$$\bar{W}_2 = \binom{n}{3}^{-1}\sum_{i<j<k}W_{ijk}, \qquad W_{ijk} = \frac{(\nu'U_{ij})(\nu'U_{ik}) + (\nu'U_{ij})(\nu'U_{h,jk}) + (\nu'U_{ik})(\nu'U_{h,jk})}{3}.$$

See Lemmas 3.1.1 and 3.1.2 in the Supplemental Appendix of Cattaneo *et al.* (2014b) for a proof. Thus, for $\mathfrak{c} \in \mathbb{R}$, we consider the following generic (debiased when $\mathfrak{c} = 1$) Studentization:

$$\widehat{\vartheta}_{\mathfrak{c}}^2 := (2 - \mathfrak{c})\binom{n}{2}^{-1}\bar{W}_1 + \frac{4}{n}[1 + o(n^{-1})]\bar{W}_2 - \frac{4}{n}\widehat{\theta}_\nu^2.$$

In particular, $\widehat{\vartheta}_{\texttt{AL}}^2 = \widehat{\vartheta}_0^2$ and $\widehat{\vartheta}_{\texttt{SB}}^2 = \widehat{\vartheta}_1^2$. The centering considered in the literature is $\widehat{\vartheta}_{\mathfrak{c}}^2$ is

$$\vartheta_{\mathfrak{c}}^2 := \mathfrak{c}\binom{n}{2}^{-1}\mathbb{E}[\bar{W}_1] + \frac{4}{n}\mathbb{E}[\bar{W}_2] - \frac{4}{n}(\mathbb{E}[\widehat{\theta}_\nu])^2,$$

which implies that $\vartheta_0^2 = \vartheta_{\texttt{AL}}^2$ and $\vartheta_1^2 = \vartheta_{\texttt{SB}}^2 + o(n^{-1})$.

The underlying U-statistics have the following mean square convergence rates:

$$\mathbb{E}[(\widehat{\theta}_\nu - \mathbb{E}[\widehat{\theta}_\nu])^2] = O(n^{-1} + n^{-2}h^{-d-2}),$$

$$\mathbb{E}[(\bar{W}_1 - \mathbb{E}[(\nu'U_{12})^2])^2] = O(n^{-1}h^{-2d-4} + n^{-2}h^{-3d-4}),$$

$$\mathbb{E}[(\bar{W}_2 - \mathbb{E}[(\mathbb{E}[\nu'U_{12}|Z_1])^2])^2] = O(n^{-1} + n^{-2}h^{-d-4} + n^{-3}h^{-2d-4}),$$

The proof is given in Cattaneo *et al.* (2014b, Supplemental Appendix): see Lemma 3.1.3 for the first

25

two results, and Lemma 3.1.4 for the third result. (Note that while the statement of those lemmas gives convergence rates in probability, the proof mean square convergence rates.) Therefore,

$$\mathbb{E}\left[(\widehat{\vartheta}_{\mathfrak{c}}^2 - \vartheta_{\mathfrak{c}}^2)^2\right] \leq Cn^{-4}\mathbb{E}[(\bar{W}_1 - \mathbb{E}[(\nu'U_{12})^2])^2] + Cn^{-2}\mathbb{E}[(\bar{W}_2 - \mathbb{E}[(\mathbb{E}[\nu'U_{12}|Z_1])^2])^2]$$
$$+ Cn^{-2}\mathbb{E}[(\bar{U} - \mathbb{E}[\nu'U_{12}])^2]$$
$$= O(n^{-3} + n^{-4}h^{-d-4}).$$

Similar long calculations as in Cattaneo *et al.* (2014b, Supplemental Appendix) show that:

$$\mathbb{E}[(\widehat{\theta}_\nu - \mathbb{E}[\widehat{\theta}_\nu])^4] = O(n^{-2} + n^{-4}h^{-d-4} + n^{-5}h^{-2d-4} + n^{-6}h^{-3d-4}),$$
$$\mathbb{E}[(\bar{W}_1 - \mathbb{E}[(\nu'U_{12})^2])^4] = O(n^{-2}h^{-4d-8} + n^{-4}h^{-6d-8} + n^{-5}h^{-6d-8} + n^{-6}h^{-7d-8}),$$
$$\mathbb{E}[(\bar{W}_2 - \mathbb{E}[(\mathbb{E}[\nu'U_{12}|Z_1])^2])^4] = O(n^{-2} + n^{-4}h^{-d-8} + n^{-5}h^{-2d-8} + n^{-6}h^{-3d-8}),$$

which gives

$$\mathbb{E}\left[(\widehat{\vartheta}_{\mathfrak{c}}^2 - \vartheta_{\mathfrak{c}}^2)^4\right] \leq Cn^{-8}\mathbb{E}[(\bar{W}_1 - \mathbb{E}[(\nu'U_{12})^2])^4] + Cn^{-4}\mathbb{E}[(\bar{W}_2 - \mathbb{E}[(\mathbb{E}[\nu'U_{12}|Z_1])^2])^4]$$
$$+ Cn^{-4}\mathbb{E}[(\bar{U} - \mathbb{E}[\nu'U_{12}])^4]$$
$$= O(n^{-6} + n^{-8}h^{-d-8}).$$

Consequently, for the remainder of the proof we set $\widehat{\vartheta}^2 = \widehat{\vartheta}_{\mathfrak{c}}^2$ and $\vartheta^2 = \vartheta_{\mathfrak{c}}^2$.

Bounds for $\mathcal{R}_1$–$\mathcal{R}_3$

For n large enough, and using Markov inequality,

$$\mathcal{R}_1 \leq \mathbb{P}\left[(\widehat{\vartheta}_{\mathfrak{c}}^2 - \vartheta_{\mathfrak{c}}^2)^2 > \frac{Cr_n}{n^2(\log n)^2}\right] + o(r_n) \leq Cn^4(\log n)^4 r_n^{-2}\mathbb{E}\left[(\widehat{\vartheta}_{\mathfrak{c}}^2 - \vartheta_{\mathfrak{c}}^2)^4\right] + o(r_n)$$
$$= n^5(\log n)^4 O(n^{-6} + n^{-8}h^{-d-8}) + o(r_n) = o(r_n).$$

Using Theorem A.1 and Corollary A.1, it follows that a valid Edgeworth expansion holds for $\frac{\widehat{\theta}_\nu - \theta_\nu}{\vartheta_{\mathfrak{c}}}$, which implies that

$$\mathcal{R}_2 = 1 - \mathbb{P}\left[\frac{\widehat{\theta}_\nu - \theta_\nu}{\vartheta} \leq C\log n\right] + \mathbb{P}\left[\frac{\widehat{\theta}_\nu - \theta_\nu}{\vartheta} \leq -C\log n\right]$$
$$= 1 - \Phi(C\log(n)) + \Phi(-C\log(n)) + C\frac{\phi(\log n)\log n}{nh^{d+2}} + o(r_n) = o(r_n),$$

by properties of the Gaussian distribution.

Finally, Markov inequality implies

$$\mathcal{R}_3 \leq Cn(\log n)^2\mathbb{E}\left[(\widehat{\vartheta}_{\mathfrak{c}}^2 - \vartheta_{\mathfrak{c}}^2)^2\right] = n(\log n)^2 O(n^{-3} + n^{-4}h^{-d-4}) = o(r_n).$$

Therefore, $\mathcal{R}_1 + \mathcal{R}_2 + \mathcal{R}_3 = o(r_n)$.

Expansion for E

We consider $E = \rho(\breve{F}_{\mathfrak{c}}, G_{\mathfrak{c}})$, where

$$\breve{F}_{\mathfrak{c}}(t) := \mathbb{P}\left[\left(1 - \frac{\widehat{\vartheta}_{\mathfrak{c}}^2 - \vartheta_{\mathfrak{c}}^2}{2\vartheta_{\mathfrak{c}}^2}\right)\left(\frac{\nu'\bar{L}}{\vartheta_{\mathfrak{c}}} + \frac{\nu'\bar{Q}}{\vartheta_{\mathfrak{c}}}\right) + B_{\mathfrak{c}} \leq t\right], \qquad B_{\mathfrak{c}} := \frac{\mathbb{E}[\widehat{\theta}_\nu] - \theta_\nu}{\vartheta_{\mathfrak{c}}},$$

and

$$G_{\mathfrak{c}}(t) := \Phi(t) - \phi(t)\left\{\frac{\sqrt{n}h^P\beta}{\nu'\Sigma\nu} - \frac{1-\mathfrak{c}}{nh^{d+2}}\frac{\nu'\Delta\nu}{\nu'\Sigma\nu}t - \frac{1}{\sqrt{n}6(\nu'\Sigma\nu)^3}\left[\kappa_1(2t^2 + 1) + \kappa_2(t^2 + 1)\right]\right\}.$$

Recall that, in particular, $\mathfrak{c} = 0$ corresponds to `AL` implementation and $\mathfrak{c} = 1$ corresponds to `SB` implementation (i.e., $G_{\texttt{AL}}(t) = G_0(t)$ and $G_{\texttt{SB}}(t) = G_1(t)$). Then, applying the smoothing inequality as in Theorem A.1,

$$\rho(\breve{F}_{\mathfrak{c}}, G_{\mathfrak{c}}) \lesssim \breve{I}_1 + \breve{I}_2 + \breve{I}_3 + \breve{I}_4 + \frac{1}{\sqrt{n}\log n},$$

where

$$\breve{I}_1 := \int_{|t|\leq \log n}\left|\frac{\chi_{\breve{F}_{\mathfrak{c}}}(t) - \chi_{G_{\mathfrak{c}}}(t)}{t}\right|\mathrm{d}t, \qquad \breve{I}_2 := \int_{\log n < |t| \leq c\sqrt{n}}\left|\frac{\chi_{\breve{F}_{\mathfrak{c}}}(t)}{t}\right|\mathrm{d}t,$$

$$\breve{I}_3 := \int_{c\sqrt{n} < |t| \leq \sqrt{n}\log n}\left|\frac{\chi_{\breve{F}_{\mathfrak{c}}}(t)}{t}\right|\mathrm{d}t, \qquad \breve{I}_4 := \int_{|t| > \log n}\left|\frac{\chi_{G_{\mathfrak{c}}}(t)}{t}\right|\mathrm{d}t.$$

The last three integrals above can be upper bounded following the same arguments used in the proof of Theorem A.1 to conclude that $\breve{I}_2 + \breve{I}_3 + \breve{I}_4 = o(\sqrt{n}h^P + n^{-1}h^{-d-2} + n^{-1/2})$. The first integral, $\breve{I}_1$, is analyzed by expanding $\chi_{\breve{F}_{\mathfrak{c}}}(t)$ by generalizing the proof of Theorem A.1 to account for the contribution from Studentization to the sampling distribution of the linearized version of the statistic ($\breve{F}_{\mathfrak{c}}$).

First, by (A.5) we write

$$\chi_{\breve{F}_{\mathfrak{c}}}(t) = \exp(\iota t B_{\mathfrak{c}})\mathbb{E}\exp(\iota t\widetilde{U}_{\mathfrak{c}}) = \left[1 + \iota t B_{\mathfrak{c}} + O(t^2 B_{\mathfrak{c}}^2)\right]\mathbb{E}\exp(\iota t\widetilde{U}_{\mathfrak{c}}), \tag{C.1}$$

where

$$\widetilde{U}_{\mathfrak{c}} = \left(1 - \frac{\widehat{\vartheta}_{\mathfrak{c}}^2 - \vartheta_{\mathfrak{c}}^2}{2\vartheta_{\mathfrak{c}}^2}\right)\left(\frac{\nu'\bar{L}}{\vartheta_{\mathfrak{c}}} + \frac{\nu'\bar{Q}}{\vartheta_{\mathfrak{c}}}\right)$$

From (C.8) below, we have

$$-\frac{\widehat{\vartheta}_{\mathfrak{c}}^2 - \vartheta_{\mathfrak{c}}^2}{2\vartheta_{\mathfrak{c}}^2} = H_{\mathfrak{c}} + \mathcal{T}_{\mathfrak{c}} \tag{C.2}$$

with

$$H_{\mathfrak{c}} := -\binom{n}{2}^{-1}\frac{1-\mathfrak{c}}{2\vartheta_{\mathfrak{c}}^2}\mathbb{E}[q_{12}^2] - \frac{1}{2n\vartheta_{\mathfrak{c}}^2}\frac{1}{n}\sum_{i=1}^{n}\left\{\left(\ell_i^2 - \mathbb{E}[\ell_i^2]\right) + 4\mathbb{E}[\ell_i q_{ij}|Z_i]\right\}$$

$$-\frac{2}{n\vartheta_{\mathfrak{c}}^2}\binom{n}{2}^{-1}\sum_{i<j}\mathbb{E}[q_{ij}q_{ik}|Z_j, Z_k],$$

where we define $\ell_i := \nu'L_i$ and $q_{ij} := \nu'Q_{ij}$, and $\mathcal{T}_{\mathfrak{c}} := -\mathcal{V}_{\mathfrak{c}}/(2\vartheta_{\mathfrak{c}}^2)$ with $\mathcal{V}_{\mathfrak{c}}$ is given in (C.8). Next, applying (A.5) repeatedly, we are left with

$$\mathbb{E}\exp(\iota t\widetilde{U}_{\mathfrak{c}}) = \mathbb{E}\exp\left[\iota t\left(\frac{\nu'\bar{L}}{\vartheta_{\mathfrak{c}}} + \frac{\nu'\bar{Q}}{\vartheta_{\mathfrak{c}}}\right)\right] + \iota t\mathbb{E}H_{\mathfrak{c}}\frac{\nu'\bar{L}}{\vartheta_{\mathfrak{c}}}\exp(\iota t\frac{\nu'\bar{L}}{\vartheta_{\mathfrak{c}}}) + O\left(\mathcal{E}_1(t)\right), \tag{C.3}$$

where $\mathcal{E}_1(t) = |t|\mathbb{E}|\mathcal{T}_{\mathfrak{c}}(\nu'\bar{L}) + (H_{\mathfrak{c}} + \mathcal{T}_{\mathfrak{c}})(\nu'\bar{Q})| + t^2(\mathbb{E}(H_{\mathfrak{c}}(\nu'\bar{L}))^2 + \mathbb{E}|H_{\mathfrak{c}}(\nu'\bar{L})(\nu'\bar{Q})|$. The first term was expanded in the proof of Theorem A.1, as it corresponds to the standardized version of the statistic. The second term can be expanded analogously (see, e.g., Appendix B-(a) in Nishiyama and Robinson (2001)):

$$\mathbb{E}\left[H_{\mathfrak{c}}\frac{\nu'\bar{L}}{\vartheta_{\mathfrak{c}}}\exp(\iota t\nu'\bar{L}/\vartheta_{\mathfrak{c}})\right] \tag{C.4}$$

$$= -[\psi\left(\tfrac{t}{n\vartheta_{\mathfrak{c}}}\right)]^{n-1}\left[\frac{1-\mathfrak{c}}{2}\frac{\iota t}{\vartheta_{\mathfrak{c}}^2}\binom{n}{2}^{-1}\mathbb{E}[q_{12}^2] + O\left(\frac{|t|}{n^2h^{d+2}} + \frac{t^2}{n^{3/2}h^{d+2}} + \frac{|t|h^P}{h^{d+2}}\right)\right]$$

$$- [\psi\left(\tfrac{t}{n\vartheta_{\mathfrak{c}}}\right)]^{n-1}\left[\frac{1}{2\vartheta_{\mathfrak{c}}^3 n^2}(\mathbb{E}\ell_1^3 + 4\mathbb{E}\ell_1\ell_2 q_{12}) + O\left(\tfrac{|t|}{n}\right)\right]$$

$$- [\psi\left(\tfrac{t}{n\vartheta_{\mathfrak{c}}}\right)]^{n-2}\left[\frac{(\iota t)^2}{2\vartheta_{\mathfrak{c}}^3 n^2}(\mathbb{E}\ell_1^3 + 4\mathbb{E}\ell_1\ell_2 q_{12}) + O\left(\tfrac{t^2+|t|^3}{n} + \tfrac{t^4}{n^{3/2}}\right)\right]$$

$$- [\psi\left(\tfrac{t}{n\vartheta_{\mathfrak{c}}}\right)]^{n-3}\left[O\left(\tfrac{|t|^3+|t|}{n} + \tfrac{t^2}{n^{3/2}h^{d+2}} + \tfrac{t^6}{n^3 h^{d+2}} + \tfrac{|t|^5}{n^{5/2}h^{d+2}} + \tfrac{t^4+|t|^3}{n^2 h^{d+2}}\right)\right]. \tag{C.5}$$

Combine (A.9), (A.10), (C.3), and (C.4) to obtain

$$\mathbb{E}\exp(\iota t\widetilde{U}_{\mathfrak{c}}) = \exp\left(-\tfrac{t^2}{2}\right)\left[1 + \frac{(\iota t)^2}{2}\left(\frac{\mathbb{E}[\ell_1^2]}{\vartheta_{\mathfrak{c}}^2 n} - 1\right) + \frac{(\iota t)^3}{6\vartheta_{\mathfrak{c}}^3 n^2}\mathbb{E}\ell_1^3 + O(\mathcal{E}_2(t)) + o(\mathcal{E}_3(t))\right]$$

$$\times \left[1 + \frac{(\iota t)^2}{2\vartheta_{\mathfrak{c}}^2}\binom{n}{2}^{-1}\mathbb{E}[q_{12}^2] + \frac{(\iota t)^3}{\vartheta_{\mathfrak{c}}^3 n^2}\mathbb{E}\ell_1\ell_2 q_{12}\right.$$

$$\left. - \frac{1-\mathfrak{c}}{2}\frac{(\iota t)^2}{\vartheta_{\mathfrak{c}}^2}\binom{n}{2}^{-1}\mathbb{E}[q_{12}^2] - \left(\frac{\iota t + (\iota t)^3}{\vartheta_{\mathfrak{c}}^3 n^2}\right)\left(\frac{\mathbb{E}\ell_1^3}{2} + 2\mathbb{E}\ell_1\ell_2 q_{12}\right) + O(\mathcal{E}_4(t))\right]$$

$$+ O(\mathcal{E}_1(t)), \tag{C.6}$$

where $\mathcal{E}_2(t)$ and $\mathcal{E}_3(t)$ are the last two rates appearing in (A.9) respectively. Also, proceeding as in Nishiyama and Robinson (2001),

$$\mathcal{E}_4(t) = o\left(\frac{t^2 + t^{10}}{nh^{d+2}} + \frac{t^2 + t^6}{\sqrt{n}}\right).$$

Combine (C.6) with (C.1) and expand the product to obtain

$$\chi_{\breve{F}_{\mathfrak{c}}}(t) = \exp\left(-\tfrac{t^2}{2}\right)\left[1 + \sum_{j=1}^{3}(\iota t)^j \breve{\gamma}_{\mathfrak{c},j}\right] + O(\mathcal{E}_5(t)),$$

where

$$\breve{\gamma}_{\mathfrak{c},1} := \left(\frac{\beta h^P}{\vartheta_{\mathfrak{c}}} - \frac{\mathbb{E}\ell_1^3/2 + 2\mathbb{E}\ell_1\ell_2 q_{12}}{\vartheta_{\mathfrak{c}}^3 n^2}\right),$$

$$\breve{\gamma}_{\mathfrak{c},2} := -(1-\mathfrak{c})\binom{n}{2}^{-1}\frac{\mathbb{E}q_{12}^2}{2\vartheta_{\mathfrak{c}}^2},$$

$$\breve{\gamma}_{\mathfrak{c},3} := -\frac{1}{6n^2\vartheta_{\mathfrak{c}}^3}(2\mathbb{E}\ell_1^3 + 6\mathbb{E}\ell_1\ell_2 q_{12}),$$

and

$$
\begin{aligned}
\mathcal{E}_5(t) := & \left[e^{-t^2/2}\frac{|t|^3}{\sqrt{n}} + o(n^{-1/2}(t^6 + |t|^3)e^{-t^2/4})\right]\left[\frac{t^2}{n^2 h^{d+2}} + \frac{|t|^3 + |t|}{\sqrt{n}} + \mathcal{E}_4(t)\right] \\
& + e^{-t^2/2}\left[|t|\sqrt{n}h^P + t^2 h^{2P} + |t|\sqrt{n}h^{2P}\right]\left[\frac{t^2}{n^2 h^{d+2}} + \frac{|t|^3 + |t|}{\sqrt{n}} + \mathcal{E}_4(t)\right] \\
& + (|t|\sqrt{n}h^P + t^2 n h^{2P})\left[e^{-t^2/2}\frac{|t|^3}{\sqrt{n}} + o(n^{-1/2}(t^6 + |t|^3)e^{-t^2/4})\right] \\
& + (|t|\sqrt{n}h^P + t^2 n h^{2P})\left[e^{-t^2/2}\frac{|t|^3}{\sqrt{n}} + o(n^{-1/2}(t^6 + |t|^3)e^{-t^2/4})\right]\left[\frac{t^2}{n^2 h^{d+2}} + \frac{|t| + |t|^3}{\sqrt{n}} + \mathcal{E}_4(t)\right] \\
& + (|t| + t^2\sqrt{n}h^P + |t|^3 n h^{2P})\left(\mathbb{E}|\mathcal{T}_{\mathfrak{c}}\bar{L}| + \mathbb{E}|(H_{\mathfrak{c}} + \mathcal{T}_{\mathfrak{c}})\bar{Q}|\right) \\
& + (t^2 + |t|^3\sqrt{n}h^P + t^4 n h^{2P})\left(\mathbb{E}(H_{\mathfrak{c}}\bar{L})^2 + \mathbb{E}|H_{\mathfrak{c}}\bar{L}\bar{Q}|\right).
\end{aligned}
$$

We showed in the proof of Theorem 1 that $\mathbb{E}\ell_1^3 = \kappa_1 + O(h^P) = o(1)$ and $\mathbb{E}\ell_1\ell_2 q_{12} = \kappa_2 + O(h^P) = o(1)$, and hence

$$
\begin{aligned}
\chi_{\breve{F}_{\mathfrak{c}}}(t) = & \exp\left(-\tfrac{t^2}{2}\right)\left[1 + \iota t\left(\frac{\beta h^P}{\vartheta_{\mathfrak{c}}} - \frac{\kappa_1/2 + 2\kappa_2}{\vartheta_{\mathfrak{c}}^3 n^2}\right) - (\iota t)^2\binom{n}{2}^{-1}\frac{\mathbb{E}q_{12}^2}{2\vartheta_{\mathfrak{c}}^2} - \frac{(\iota t)^3}{6n^2\vartheta_{\mathfrak{c}}^3}(2\kappa_1 + 6\kappa_2)\right] \\
& + O(\mathcal{E}_5(t)) + o\left(\exp\left(-\tfrac{t^2}{2}\right)\frac{|t| + |t|^3}{\sqrt{n}}\right).
\end{aligned}
$$

Note that the first term is the characteristic function of G. Finally, we bound the moments appearing in $\mathcal{E}_5(t)$: $\mathbb{E}|\mathcal{T}_{\mathfrak{c}}\bar{L}|$, $\mathbb{E}|(H_{\mathfrak{c}} + \mathcal{T}_{\mathfrak{c}})\bar{Q}|$, $\mathbb{E}(H_{\mathfrak{c}}\bar{L})^2$, and $\mathbb{E}|H_{\mathfrak{c}}\bar{L}\bar{Q}|$. Holder's inequality combined with the theorem assumptions give

$$
\begin{aligned}
\mathbb{E}|\mathcal{T}_{\mathfrak{c}}\bar{L}| &\leq \sqrt{\mathbb{E}|\mathcal{T}_{\mathfrak{c}}|^2\mathbb{E}|\bar{L}|^2} = O(n^{-1}h^{-(d+2)/2}) \\
\mathbb{E}|(H_{\mathfrak{c}} + \mathcal{T}_{\mathfrak{c}})\bar{Q}| &\leq \sqrt{\mathbb{E}|H_{\mathfrak{c}} + \mathcal{T}_{\mathfrak{c}}|^2\mathbb{E}|\bar{Q}|^2} = O\left((n^{-1/2} + n^{-1}h^{-d-2})(n^{-1/2}h^{-d/2-1})\right) \\
\mathbb{E}(H_{\mathfrak{c}}\bar{L})^2 &= O(n^{-1} + n^{-2}h^{-2d-4}) \\
\mathbb{E}|H_{\mathfrak{c}}\bar{L}\bar{Q}| &= O\left((n^{-1/2} + n^{-1}h^{-d-2})(n^{-1/2}h^{-d/2-1})\right).
\end{aligned}
$$

Therefore, if $(\log n)^9/(nh^{d+2}) \to 0$,

$$\check{I}_1 := \int_{|t| \le \log n} \frac{|\chi_{\check{F}_c}(t) - \chi_{G_c}(t)|}{|t|} = o(\sqrt{n}h^P + n^{-1}h^{-d-2} + n^{-1/2}).$$

The proof is finalized. $\square$

C.1 Alternative Decomposition of $\widehat{\vartheta}_c$

Let $u_{ij} = \nu' U_{ij}$ and following Callaert and Veraverbeke (1981) with S_N^2 given in the their main Theorem, we have

$$S_N^2 := \frac{n^2(n-1)}{(n-2)^2}\widehat{\vartheta}_{\mathtt{AL}}^2 = \frac{4(n-1)}{(n-2)^2}\sum_{i=1}^n (\nu'\widehat{L}_i/2)^2$$

$$= \frac{8}{(n-1)(n-2)^2}\left[\sum_{i<j}(u_{ij} - \mathbb{E}u_{12})^2 + \sum_{i=1}^n \sum_{j<k, j \neq i}(u_{ij} - \mathbb{E}u_{12})(u_{ik} - \mathbb{E}u_{12})\right]$$

$$- \frac{4n(n-1)}{(n-2)^2}(\widehat{\theta} - \mathbb{E}u_{12}).$$

Define $g_i := \mathbb{E}[\ell_j q_{ij}|Z_i]$ and use the fact that $u_{ij} - \mathbb{E}u_{12} = \ell_i/2 + \ell_j/2 + q_{ij}$ to further decompose

$$S_N^2 = \frac{1}{n}\sum_{i=1}^n \ell_i^2 + 4g_i - \binom{n}{2}^{-1}\sum_{i<j}^n \ell_i\ell_j + 2\binom{n}{2}^{-1}\sum_{i<j}^n \left[(\ell_i + \ell_j)q_{ij} - g_i - g_j\right]$$

$$- \frac{4}{n}\binom{n-1}{2}^{-1}\sum_{i=1}^n \ell_i \sum_{j<k, j\neq i}^n q_{jk} + \frac{4}{n-2}\binom{n-1}{2}^{-1}\sum_{i=1}^n \sum_{j<k, j\neq i}^n q_{ij}q_{ik}$$

$$- \frac{4n(n-1)}{(n-2)^2}\left[\binom{n}{2}^{-1}\sum_{i<j} q_{ij}\right]^2 + \frac{4n}{(n-2)^2}\binom{n}{2}^{-1}\sum_{i<j} q_{ij}^2. \tag{C.7}$$

The first term, we center on its expectation

$$\frac{1}{n}\sum_{i=1}^n \ell_i^2 + 4g_i = \mathbb{E}[\ell_1^2] + \frac{1}{n}\sum_{i=1}^n \left(\ell_i^2 - \mathbb{E}[\ell_1^2]\right) + 4g_i.$$

Define $\varphi_{ij} := \mathbb{E}[q_{ki}q_{kj}|Z_i, Z_j]$, and for the fifth term we write

$$\frac{4}{n-2}\binom{n-1}{2}^{-1}\sum_{i=1}^n \sum_{j<k, j\neq i}^n q_{ij}q_{ik} = 4\binom{n-1}{2}^{-1}\sum_{i<j}\varphi_{ij} + \frac{4}{n-2}\binom{n-1}{2}^{-1}\sum_{i=1}^n \sum_{j<k, j\neq i}^n (q_{ij}q_{ik} - \varphi_{jk}).$$

Finally, for the last term

$$\frac{4n}{(n-2)^2}\binom{n}{2}^{-1}\sum_{i<j} q_{ij}^2 = \frac{4n}{(n-2)^2}\binom{n}{2}^{-1}\sum_{i<j}\left[q_{ij}^2 - \varphi_{ii} - \varphi_{jj} + \mathbb{E}[q_{12}]^2\right]$$

30

$$+ \frac{8}{(n-2)^2} \sum_{i=1}^{n} \left(\varphi_{ii} - \mathbb{E}[q_{12}^2] \right) + \frac{4n}{(n-2)^2} \mathbb{E}[q_{12}^2]$$

Plug the last three displays back into (C.7) to conclude

$$S_N^2 = \mathbb{E}[\ell_1^2] + \frac{4n}{(n-2)^2} \mathbb{E}[q_{12}^2] + \frac{1}{n} \sum_{i=1}^{n} \left[(\ell_i^2 - \mathbb{E}[\ell_1^2]) + 4g_i \right] + 4 \binom{n-1}{2}^{-1} \sum_{i<j} \varphi_{ij} + \mathcal{S}_{\mathfrak{c}},$$

where $\mathcal{S}_{\mathfrak{c}}$ collection the remaining terms.

Next, we have

$$\binom{n}{2}^{-1} \sum_{i<j} u_{ij}^2 = \binom{n}{2}^{-1} \sum_{i<j} (q_{ij} - \ell_i/2 - \ell_j/2 - \mathbb{E}u_{12})^2$$

$$= \mathbb{E}q_{12}^2 + \binom{n}{2}^{-1} \sum_{i<j} (q_{ij}^2 - \mathbb{E}q_{ij}^2) - \binom{n}{2}^{-1} \sum_{i=1}^{n} (\ell_i/2)^2 - (\mathbb{E}u_{12})^2$$

$$- \binom{n}{2}^{-1} \sum_{i<j} q_{ij}(\ell_i + \ell_j) + 2\mathbb{E}u_{12} \binom{n}{2}^{-1} \sum_{i<j} (q_{ij} + \ell_i/2 + \ell_j/2)$$

$$+ \frac{1}{2} \binom{n}{2}^{-1} \sum_{i<j} \ell_i \ell_j$$

$$= \mathbb{E}q_{12}^2 + \mathcal{Q}_{\mathfrak{c}},$$

where $\mathcal{Q}_{\mathfrak{c}}$ is by definition.

We have

$$\widehat{\vartheta}_{\mathfrak{c}} = \widehat{\vartheta}_{\mathrm{AL}}^2 - \mathfrak{c} \binom{n}{2}^{-1} h^{-d-2} \nu' \widehat{\Delta} \nu$$

$$= \frac{(n-2)^2}{n^2(n-1)} S_N^2 - \mathfrak{c} \binom{n}{2}^{-1} \left[\binom{n}{2}^{-1} \sum_{i<j} u_{ij}^2 \right]$$

$$= (2 - \mathfrak{c}) \binom{n}{2}^{-1} \mathbb{E}[q_{12}^2] + \frac{1+o(1)}{n} \mathbb{E}[\ell_1^2] + \frac{1+o(1)}{n^2} \sum_{i=1}^{n} \left[(\ell_i^2 - \mathbb{E}[\ell_1^2]) + 4g_i \right]$$

$$+ \frac{1+o(1)}{n} 4 \binom{n}{2}^{-1} \sum_{i<j} \varphi_{ij} + \frac{1+o(1)}{n} \mathcal{S}_{\mathfrak{c}} + \mathfrak{c} \binom{n}{2}^{-1} \mathcal{Q}_{\mathfrak{c}},$$

which gives the following simplified expression for the class of Studentizations:

$$\widehat{\vartheta}_{\mathfrak{c}} = (2 - \mathfrak{c}) \binom{n}{2}^{-1} \mathbb{E}[q_{12}^2] + \frac{1}{n} \mathbb{E}[\ell_1^2] + \frac{1}{n^2} \sum_{i=1}^{n} \left[(\ell_i^2 - \mathbb{E}[\ell_1^2]) + 4g_i \right] + \frac{1}{n} 4 \binom{n}{2}^{-1} \sum_{i<j} \varphi_{ij} + \mathcal{V}_{\mathfrak{c}}, \quad \text{(C.8)}$$

where $\mathcal{V}_{\mathfrak{c}}$ is by definition.

31

References

Ahn, H., Ichimura, H., Powell, J. L., and Ruud, P. A. (2018). "Simple Estimators for Invertible Index Models," *Journal of Business & Economic Statistics*, *36*(1), 1–10.

Aradillas-Lopez, A., Honoré, B. E., and Powell, J. L. (2007). "Pairwise Difference Estimation with Nonparametric Control Variables," *International Economic Review*, *48*(4), 1119–1158.

Bhattacharya, R. N. and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions*: John Wiley and Sons.

Blundell, R. W. and Powell, J. L. (2004). "Endogeneity in Semiparametric Binary Response Models," *Review of Economic Studies*, *71*(3), 655–679.

Callaert, H. and Veraverbeke, N. (1981). "The Order of the Normal Approximation for a Studentized U-statistic," *Annals of Statistics*, *9*, 194–200.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2018). "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association*, *113*(522), 767–779.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2022). "Coverage Error Optimal Confidence Intervals for Local Polynomial Regression," *Bernoulli*, *28*(4), 2998–3022.

Cattaneo, M. D., Crump, R. K., and Jansson, M. (2010). "Robust Data-Driven Inference for Density-Weighted Average Derivatives," *Journal of the American Statistical Association*, *105*(491), 1070–1083.

Cattaneo, M. D., Crump, R. K., and Jansson, M. (2013). "Generalized Jackknife Estimators of Weighted Average Derivatives (with Discussions and Rejoinder)," *Journal of the American Statistical Association*, *108*(504), 1243–1268.

Cattaneo, M. D., Crump, R. K., and Jansson, M. (2014a). "Small Bandwidth Asymptotics for Density-Weighted Average Derivatives," *Econometric Theory*, *30*(1), 176–200.

Cattaneo, M. D., Crump, R. K., and Jansson, M. (2014b). "Bootstrapping Density-Weighted Average Derivatives," *Econometric Theory*, *30*(6), 1135–1164.

Cattaneo, M. D. and Jansson, M. (2018). "Kernel-Based Semiparametric Estimators: Small Bandwidth Asymptotics and Bootstrap Consistency," *Econometrica*, *86*(3), 955–995.

Cattaneo, M. D., Jansson, M., and Ma, X. (2019). "Two-step Estimation and Inference with Possibly Many Included Covariates," *Review of Economic Studies*, *86*(3), 210–245.

Cattaneo, M. D., Jansson, M., and Newey, W. K. (2018a). "Alternative Asymptotics and the Partially Linear Model with Many Regressors," *Econometric Theory*, *34*(2), 277–301.

Cattaneo, M. D., Jansson, M., and Newey, W. K. (2018b). "Inference in Linear Regression Models with Many Covariates and Heteroscedasticity," *Journal of the American Statistical Association*, *113*(523), 1350–1361.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2022). "Locally Robust Semiparametric Estimation," *Econometrica*, *90*(4), 1501–1535.

Dharmadhikari, S. W., Fabian, V., and Jogdeo, K. (1968). "Bounds on the Moments of Martingales," *Annals of Mathematical Statistics*, *39*(5), 1719 – 1723.

Efron, B. and Stein, C. (1981). "The Jackknife Estimate of Variance," *Annals of Statistics*, *9*(3), 586–596.

Giné, E., Latała, R., and Zinn, J. (2000). "Exponential and Moment Inequalities for U-Statistics," In E. Giné, D. M. Mason, and J. A. Wellner (eds.) *High Dimensional Probability II*, 13–38, Boston, MA: Birkhäuser Boston.

Graham, B. S., Niu, F., and Powell, J. L. (2023). "Kernel density estimation for undirected dyadic data," *Journal of Econometrics*.

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.

Hoeffding, W. (1948). "A Class of Statistics with Asymptotically Normal Distribution," *Annals of Mathematical Statistics*, *19*(3), 293 – 325.

Honoré, B. E. and Powell, J. L. (1994). "Pairwise Difference Estimators of Censored and Truncated Regression Models," *Journal of Econometrics*, *64*(1-2), 241–278.

Ichimura, H. and Todd, P. E. (2007). "Implementing Nonparametric and Semiparametric Estimators," In J. Heckman and E. Leamer (eds.) *Handbook of Econometrics, Volume VIB*: Elsevier Science B.V. 5370–5468.

Jing, B.-Y. and Wang, Q. (2003). "Edgeworth expansion for U-statistics under minimal conditions," *Annals of statistics*, *31*(4), 1376–1391.

Matsushita, Y. and Otsu, T. (2021). "Jackknife empirical likelihood: small bandwidth, sparse network and high-dimensional asymptotics," *Biometrika*, *108*(3), 661–674.

Newey, W. K. (1994). "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, *62*(6), 1349–1382.

Newey, W. K., Hsieh, F., and Robins, J. M. (2004). "Twicing Kernels and a Small Bias Property of Semiparametric Estimators," *Econometrica*, *72*, 947–962.

Newey, W. K. and McFadden, D. L. (1994). "Large Sample Estimation and Hypothesis Testing," In R. F. Engle and D. L. McFadden (eds.) *Handbook of Econometrics, Volume IV*, New York: Elsevier Science B.V. 2111–2245.

Nishiyama, Y. and Robinson, P. M. (2000). "Edgeworth Expansions for Semiparametric Averaged Derivatives," *Econometrica*, *68*, 931–979.

Nishiyama, Y. and Robinson, P. M. (2001). "Studentization in Edgeworth Expansions for Estimates of Semiparametric Index Models," In C. Hsiao, K. Morimune, and J. L. Powell (eds.) *Nonlinear Statistical Modeling: Essays in Honor of Takeshi Amemiya*, New York: Cambridge University Press, 197–240.

Nishiyama, Y. and Robinson, P. M. (2005). "The Bootstrap and the Edgeworth Correction for Semiparametric Averaged Derivatives," *Econometrica*, *73*, 197–240.

de la Peña, V. H. and Montgomery-Smith, S. J. (1995). "Decoupling Inequalities for the Tail Probabilities of Multivariate U-statistics," *Annals of Probability*, *23*(2), 806–816.

Powell, J. L. (1994). "Estimation of Semiparametric Models," In R. F. Engle and D. L. McFadden (eds.) *Handbook of Econometrics, Volume IV*, New York: Elsevier Science B.V. 2443–2521.

Powell, J. L. (2017). "Identification and Asymptotic Approximations: Three Examples of Progress in Econometric Theory," *Journal of Economic Perspectives*, *31*(2), 107–24.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989). "Semiparametric Estimation of Index Coefficients," *Econometrica*, *57*(6), 1403–1430.

Powell, J. L. and Stoker, T. M. (1996). "Optimal Bandwidth Choice for Density-Weighted Averages," *Journal of Econometrics*, *75*, 291–316.

Robinson, P. M. (1995). "The Normal Approximation for Semiparametric Averaged Derivatives," *Econometrica*, *63*, 291–216.

Stoker, T. M. (1986). "Consistent Estimation of Scaled Coefficients," *Econometrica*, *54*, 1461–1481.