TESTING FUNCTIONAL INEQUALITIES

SOKBAE LEE, KYUNGCHUL SONG, AND YOON-JAE WHANG

Abstract. This paper develops tests for inequality constraints of nonparametric regression functions. The test statistics involve a one-sided version of L_p-type functionals of kernel estimators ($1 \leq p < \infty$). Drawing on the approach of Poissonization, this paper establishes that the tests are asymptotically distribution free, admitting asymptotic normal approximation. In particular, the tests using the standard normal critical values have asymptotically correct size and are consistent against general fixed alternatives. Furthermore, we establish conditions under which the tests have nontrivial local power against Pitman local alternatives. Some results from Monte Carlo simulations are presented.

Key words. Conditional moment inequalities, kernel estimation, one-sided test, local power, L_p norm, Poissonization.

JEL Subject Classification. C12, C14.

AMS Subject Classification. 62G10, 62G08, 62G20.

1. Introduction

Suppose that we observe $\{(Y_i', X_i')'\}_{i=1}^n$ that are i.i.d. copies from a random vector, $(Y', X')' \in \mathbf{R}^J \times \mathbf{R}^d$. Write $Y_i = (Y_{1i}, \cdot\cdot\cdot, Y_{Ji})' \in \mathbf{R}^J$ and define $m_j(x) \equiv \mathbf{E}[Y_{ji}|X_i = x]$, $j = 1, 2, \cdot\cdot\cdot, J$. The notation $\equiv$ indicates definition.

This paper focuses on the problem of testing functional inequalities:

$$(1.1)\quad \begin{aligned} &H_0 : m_j(x) \leq 0 \text{ for all } (x, j) \in \mathcal{X} \times \mathcal{J}, \text{ vs.} \\ &H_1 : m_j(x) > 0 \text{ for some } (x, j) \in \mathcal{X} \times \mathcal{J}, \end{aligned}$$

where $\mathcal{X} \subset \mathbf{R}^d$ is the domain of interest and $\mathcal{J} \equiv \{1, \ldots, J\}$. Our testing problem is relevant in various applied settings. For example, in a randomized controlled trial, a researcher observes either an outcome with treatment (W_1) or an outcome without treatment (W_0) along with observable pre-determined characteristics of the subjects (X). Let $D = 1$ if the subject belongs to the treatment group and 0 otherwise. Suppose

Date: 14 August 2012.
We would like to thank an editor, an associate editor, three anonymous referees, David Mason, and Oliver Linton for their helpful comments on earlier versions of this paper. Lee thanks the Economic and Social Research Council for the ESRC Centre for Microdata Methods and Practice (RES-589-28-0001) and the European Research Council for the research grant (ERC-2009-StG-240910-ROMETA). Whang thanks the Korea Research Foundation for the research grant (KRF-2009-327-B00094).

that assignment to treatment is random and independent of X and that the assignment probability $p \equiv P\{D = 1\}$, $0 < p < 1$, is fixed by the experiment design. Then the average treatment effect $\mathbf{E}(W_1 - W_0|X = x)$, conditional on $X = x$, can be written as

$$\mathbf{E}(W_1 - W_0|X = x) = \mathbf{E}\left[\frac{DW}{p} - \frac{(1-D)W}{1-p}\middle| X = x\right],$$

where $W \equiv DW_1 + (1 - D)W_0$. In this setup, it may be of interest to test whether or not $m(x) \equiv \mathbf{E}(W_1 - W_0|X = x) \leq 0$ for all x.

In economic theory, primitive assumptions of economic models generate certain testable implications in the form of functional inequalities. For example, Chiappori, Jullien, Salanié, and Salanié (2006) formulated some testable restrictions in the study of insurance markets. Our tests are applicable for testing their restrictions (e.g. equation (4) of Chiappori, Jullien, Salanié, and Salanié (2006)). Furthermore, our method can be used to test for monotone treatment response (see, e.g. Manski (1997)). For example, testing for a decreasing demand curve for each level of price in treatments and for each value of covariates falls within the framework of this paper.

Our test statistic can also be used to construct confidence regions for a parameter that is partially identified under conditional moment inequalities. See, among many others, Andrews and Shi (2011a,b), Armstrong (2011), Chernozhukov, Lee, and Rosen (2009), Chetverikov (2012), and references therein for inference with conditional moment inequalities.

This paper proposes a one-sided L_p approach in testing nonparametric functional inequalities. While measuring the quality of an estimated nonparametric function by its L_p-distance from the true function has long received attention in the literature (see Devroye and Györfi (1985), for an elegant treatment of the L_1 norm of nonparametric density estimation), the advance of this approach for general nonparametric testing seems to have been rather slow relative to other approaches, perhaps due to its technical complexity.

Csörgő and Horváth (1988) first established a central limit theorem for the L_p-distance of a kernel density estimator from its population counterpart, and Horváth (1991) introduced a Poissonization technique into the analysis of the L_p-distance. Beirlant and Mason (1995) developed a different Poissonization technique and established a central limit theorem for the L_p-distance of kernel density estimators and regressograms from their expected values without assuming smoothness conditions for the nonparametric functions. Giné, Mason and Zaitsev (2003: GMZ, hereafter) employed this technique to prove the weak convergence of an L_1-distance process indexed by kernel functions in kernel density estimators.

This paper builds on the contributions of Beirlant and Mason (1995) and GMZ to develop methods for testing (1.1). In particular, the tests that we propose are studentized versions of one-sided L_p-type functionals. We show that our proposed test statistic is distributed as standard normal under the least favorable case of the null hypothesis. Thus, our tests using the standard normal critical values have asymptotically correct size. We also show that our tests are consistent against general fixed alternatives and carry out local power analysis with Pitman alternatives. For the latter, we establish conditions under which the tests have nontrivial local power against Pitman local alternatives, including some $n^{-1/2}$-converging Pitman sequences.

Our tests have the following desirable properties. First, our tests do not require usual smoothness conditions for nonparametric functions for their asymptotic validity and consistency. This is because we do not need pointwise or uniform consistency of an unknown function to implement our tests. For example, a studentized version of our statistic can be estimated without need for controlling the bias. Second, our tests for (1.1) are distribution free under the least favorable case of the null hypothesis where $m_j(x) = 0$, for all $x \in \mathcal{X}$ and for all $j \in \mathcal{J}$ and at the same time have nontrivial power against some, though not all, $n^{-1/2}$-converging Pitman local alternatives. This is somewhat unexpected, given that nonparametric goodness-of-fit tests that involve random vectors of a multi-dimension and have nontrivial power against $n^{-1/2}$-converging Pitman sequences are not often distribution free. Exceptions are tests that use an innovation martingale approach (see, e.g., Khmaladze (1993), Stute, Thies and Zhu (1998), Bai (2003), and Khmaladze and Koul (2004)) or some tests of independence (or conditional independence) among random variables (see, e.g., Blum, Kiefer, and Rosenblatt (1961), Delgado and Mora (2000) and Song (2009)). Third, the local power calculation of our tests for (1.1) reveals an interesting contrast with other nonparametric tests based on kernel smoothers, e.g. Härdle and Mammen (1993) and Horowitz and Spokoiny (2001), where the latter tests are known to have trivial power against $n^{-1/2}$-converging Pitman local alternatives. Our inequality tests can have nontrivial local powers against $n^{-1/2}$-converging Pitman local alternatives, provided that a certain integral associated with local alternatives is strictly positive. On the other hand, it is shown in Section 4 that our equality tests have trivial power against $n^{-1/2}$-converging Pitman local alternatives. Therefore, the one-sided nature of inequality testing is the source of our different local power results. This finding appears new in the literature to the best of our knowledge.

The remainder of the paper is as follows. Section 2 discusses the related literature. Section 3 provides an informal description of our test statistic for a simple case, and establishes conditions under which our tests have asymptotically valid size when the null hypothesis is true and also are consistent against fixed alternatives. We also obtain local power results for the leading cases when $p = 1$ and $p = 2$. In Section 4, we make comparison with functional equality tests and highlight the main differences between testing inequalities and equalities in terms of local power. In Section 5, we report results of some Monte Carlo simulations that show that our tests perform well in finite samples. The proofs of main theorems are contained in Section 6, along with a roadmap for the proof of the main theorem.

2. Related Literature

In this section, we provide details on the related literature. The literature on hypothesis testing involving nonparametric functions has a long history. Many studies have focused on testing parametric or semiparametric specifications of regression functions against nonparametric alternatives. See, e.g., Bickel and Rosenblatt (1973), Härdle and Mammen (1993), Stute (1997), Delgado and González Manteiga (2000), Horowitz and Spokoiny (2001), and Khmaladze and Koul (2004) among many others. The testing problem in this paper is different from the aforementioned papers, as the focus is on whether certain inequality (or equality) restrictions hold, rather than on whether certain parametric specifications are plausible.

When $J = 1$, our testing problem is also different from testing

$$\begin{aligned} H_0 &: m(x) = 0 \text{ for all } x \in \mathcal{X}, \text{ against} \\ H_1 &: m(x) \geq 0 \text{ for all } x \in \mathcal{X} \text{ with strict inequality for some } x \in \mathcal{X}. \end{aligned}$$

Related to this type of testing problems, see Hall, Huber, and Speckman (1997) and Koul and Schick (1997, 2003) among others. In their setup, the possibility that $m(x) < 0$ for some x is excluded, so that a consistent test can be constructed using a linear functional of $m(x)$. On the other hand, in our setup, negative values of $m(x)$ for some x are allowed under both H_0 and H_1. As a result, a linear functional of $m(x)$ would not be suitable for our purpose.

There also exist some papers that consider the testing problem in (1.1). For example, Hall and Yatchew (2005) and Andrews and Shi (2011a,b) considered functions of the form $u \mapsto \max\{u, 0\}^p$ to develop tests for (1.1). However, their tests are not distribution free, although they achieve local power against some $n^{-1/2}$-converging sequences. See also Hall and van Keilegom (2005) for the use of the one-sided L_p-type functionals for testing for monotone increasing hazard rate. None of the aforementioned papers developed test statistics of one-sided L_p-type functionals with kernel estimators like ours. See some remarks of Ghosal, Sen, and van der Vaart (2000, p.1070) on difficulty in dealing with one-sided L_p-type functionals with kernel estimators.

In view of Bickel and Rosenblatt (1973) who considered both L_2 and sup tests, a one-sided sup test appears to be a natural alternative to the L_p-type tests studied in this paper. For example, Chernozhukov, Lee, and Rosen (2009) considered a sup norm approach in testing inequality constraints of nonparametric functions. Also, it may be of interest to develop sup tests based on a one-sided version of a bootstrap uniform confidence interval of $\hat{g}_n$, similar to Claeskens and van Keilegom (2003). The sup tests typically do not have nontrivial power against any $n^{-1/2}$-converging alternatives, but they may have better power against some "sharp peak" type alternatives (Liero, Läuter and Konakov, 1998).

Testing for inequality is related to testing for monotonicity since a null hypothesis associated inequality (respectively, monotonicity) can also be framed as that of monotonicity (respectively, convexity) of integrated moments. For example, Durot (2003) and Delgado and Escanciano (2011, 2012) used the least concave majorant operator to characterize their null hypotheses and developed tests based on the isotonic regression methods.

Finally, we mention that there exist other applications of the Poissonization method. For example, Anderson, Linton, and Whang (2012) developed methodology for kernel estimation of a polarization measure; Lee and Whang (2009) established asymptotic null distributions for the L_1-type test statistics for conditional treatment effects; and Mason (2009) established both finite sample and asymptotic moment bounds for the L_p risk for kernel density estimators. See also Mason and Polonik (2009) and Biau, Cadre, Mason, and Pelletier (2009) for asymptotic distribution theory in support estimation.

Among all the aforementioned papers, our work is most closely related to Lee and Whang (2009), but differs substantially in several important ways. First, we consider the case of multiple functional inequalities, in contrast to the single inequality case of Lee and Whang (2009). This extension requires different arguments (see, e.g. Lemma A7 in Section 6.2) and is necessary in order to encompass important applications such as testing monotonic treatment response and inference with conditional moment inequalities. Second, we extend the L_1 statistic to the general L_p statistic. Such an extension is not only theoretically challenging because many of the results of GMZ apply only to the L_1 statistic (See, e.g., Lemmas A3 and Lemmas A8 in Section 6.2), but also useful to applied econometricians because the L_p-type test statistics with different values of p generally have different power properties. Third, regularity conditions are weaker in this paper than those in Lee and Whang (2009). In particular, we allow the underlying functions to be non-smooth, which should be useful in some contexts. We believe that none of these extensions are trivial. Therefore, we view these two papers as complements rather than substitutes.

The testing framework in this paper could be easily extended to testing stochastic dominance conditional on covariates in the one-sample case or in the program evaluation setup described in the introduction. For the latter setup, testing conditional stochastic dominance amounts to testing $m(x, y) \equiv \mathbf{E}[1(W_1 \leq y) - 1(W_0 \leq y)|X = x) \leq 0$ for all $(x, y) \in \mathcal{XY}$, where $\mathcal{XY}$ is the domain of the interest and W_1 and W_0, as before, are outcomes for treatment and control groups, respectively. Then a conditional stochastic dominance test can be developed by combining a density weighted kernel estimator of $m(x, y)$ with a one-sided L_p-type functional. However, it is not straightforward to extend our framework to general two-sample cases. This is because the propensity score $P(D = 1|X = x)$ is unknown in general and has to be estimated to implement the test. See, for example, Lee and Whang (2009), Delgado and Escanciano (2011), and Hsu (2011) for testing conditional treatment effects, including testing conditional stochastic dominance, in general two-sample cases.

3. Test Statistics and Asymptotic Properties

3.1. **An Informal Description of Our Test Statistics.** Our tests are based on one-sided L_p-type functionals. For $1 \leq p < \infty$, let $\Lambda_p : \mathbf{R} \mapsto \mathbf{R}$ be such that $\Lambda_p(v) \equiv \max\{v, 0\}^p,\ v \in \mathbf{R}$. Consider the following one-sided L_p-type functionals:

$$\varphi \mapsto \Gamma_j(\varphi) \equiv \int_{\mathcal{X}} \Lambda_p(\varphi(x))w_j(x)dx, \text{ for } j \in \mathcal{J},$$

where $w_j : \mathbf{R}^d \to [0, \infty)$ is a nonnegative weight function. Let f denote the density function of X and define $g_j(x) \equiv m_j(x)f(x)$. To construct a test statistic, define

$$\hat{g}_{jn}(x) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} Y_{ji}K\left(\frac{x - X_i}{h}\right),$$

where $K : \mathbf{R}^d \mapsto \mathbf{R}$ is a kernel function and h a bandwidth parameter satisfying $h \to 0$ as $n \to \infty$. Our test statistic is a suitably studentized of version of $\Gamma_j(\hat{g}_{jn}(x))$'s.

Note that we focus on values of x for which $\hat{g}_{jn}(x) > 0$ through the use of $\Lambda_p(v)$. Thus, we expect that when H_0 is true, a suitably studentized version of $\Gamma_j(\hat{g}_{jn})$ is "not too large" for each $j \in \mathcal{J}$ but that when H_0 is false, it will diverge for some $j \in \mathcal{J}$. This motivates the use of a weighted sum of $\Gamma_j(\hat{g}_{jn})$ as a test statistic. We require that at least one component of X be continuously distributed. If some elements of X are discrete, we can modify the integral in the functional above by using some product measure between the Lebesgue and counting measures.

We show in Section 3.2 that under weak assumptions, there exist nonstochastic sequences $a_{jn} \in \mathbf{R}$, $j \in \mathcal{J}$, and $\sigma_n \in (0, \infty)$ such that as $n \to \infty$,

$$T_n \equiv \frac{1}{\sigma_n} \sum_{j=1}^{J} \left\{ n^{p/2} h^{(p-1)d/2} \Gamma_j(\hat{g}_{jn}) - a_{jn} \right\} \overset{d}{\to} N(0,1), \tag{3.1}$$

under the least favorable case of the null hypothesis, where $m_j(x) = 0$, for all $x \in \mathcal{X}$ and for all $j \in \mathcal{J}$. This is done first by deriving asymptotic results for the Poissonized version of the processes, $\{\hat{g}_{jn}(x) : x \in \mathcal{X}\}$, $j \in \mathcal{J}$, and then by translating them back into those for the original processes through the de-Poissonization lemma of Beirlant and Mason (1995). See Appendix 6.1 for details.

To construct a test statistic, we replace a_{jn} and σ_n by appropriate estimators to obtain a feasible version of T_n, say, $\hat{T}_n$, and show that the limiting distribution remains the same under a stronger bandwidth condition. Hence, we obtain a distribution free and consistent test for the nonparametric functional inequality constraints.

To provide a preview of local power analysis with Pitman alternatives in Section 3.3, suppose that $J = 1$ and $p = 1$, and the form of the local alternatives is $g_1(x) = \varrho_n \delta_1(x)$ for some function $\delta_1(x)$, where ϱ_n is a sequence of real numbers that converges to 0 as $n \to \infty$. Then (1) if $\int_{\mathcal{X}} \delta_1(x) w_1(x) dx > 0$, our test has nontrivial power against sequences of local alternatives with $\varrho_n \propto n^{-1/2}$; (2) if $\int_{\mathcal{X}} \delta_1(x) w_1(x) dx = 0$, our test has nontrivial power only against sequences of local alternatives for which $\varrho_n \to 0$ at a rate slower than $n^{-1/2}$; and (3) if $\int_{\mathcal{X}} \delta_1(x) w_1(x) dx < 0$, our test is locally biased whether or not $\varrho_n \propto n^{-1/2}$, although our test is a consistent test against general fixed alternatives.

An alternative statistic is a max statistic such as $\max_{j \in \mathcal{J}} \left\{ n^{p/2} h^{(p-1)d/2} \Gamma_j(\hat{g}_{jn}) - a_{jn} \right\}$, which we do not pursue in this paper since the "max" version of the test is not typically asymptotically pivotal.

3.2. **Test Statistics and Asymptotic Validity.** Define $\mathcal{S}_j \equiv \{x \in \mathcal{X} : w_j(x) > 0\}$ for each $j \in \mathcal{J}$, and, given $\varepsilon > 0$, let $\mathcal{S}_j^{\varepsilon}$ be an ε-enlargement of $\mathcal{S}_j$, i.e., $\mathcal{S}_j^{\varepsilon} \equiv \{x + a : x \in \mathcal{S}_j,\ a \in [-\varepsilon, \varepsilon]^d\}$. For $1 \leq p < \infty$, let

$$r_{j,p}(x) \equiv \mathbf{E}[|Y_{ji}|^p | X_i = x] f(x). \tag{3.2}$$

We introduce the following assumptions.

Assumption 1: (i) For each $j \in \mathcal{J}$ and for some $\varepsilon > 0$, $r_{j,2}(x)$ is bounded away from zero and $r_{j,2p+2}(x)$ is bounded, both uniformly in $x \in \mathcal{S}_j^{\varepsilon}$.
(ii) For each $j \in \mathcal{J}$, $w_j(\cdot)$ is nonnegative on $\mathcal{X}$ and $0 < \int_{\mathcal{X}} w_j^s(x) dx < \infty$, where $s \in \{1, 2\}$.
(iii) For $\varepsilon > 0$ in (i), $\mathcal{S}_j^{\varepsilon} \subset \mathcal{X}$ for all $j \in \mathcal{J}$.

Assumption 2: $K(u) = \Pi_{s=1}^{d} K_s(u_s)$, $u = (u_1, \cdots, u_d)$, with each $K_s : \mathbf{R} \to \mathbf{R}$, $s = 1, \cdots, d$, satisfying that (a) $K_s(u_s) = 0$ for all $u_s \in \mathbf{R}\backslash[-1/2, 1/2]$, (b) K_s is of bounded variation, and (c) $||K_s||_\infty \equiv \sup_{u_s \in \mathbf{R}} |K_s(u_s)| < \infty$ and $\int K_s(u_s)du_s = 1$.

Assumption 1(i) imposes that $\inf\{r_{j,2}(x) : x \in \mathcal{S}_j^\varepsilon\} > 0$ and $\sup\{r_{j,2p+2}(x) : x \in \mathcal{S}_j^\varepsilon\} < \infty$ for each $j \in \mathcal{J}$. Assumption 1(ii) is a weak condition on the weight function. Nonnegativity is important since we develop a sum statistic over j. Assumption 1(iii) is introduced to avoid the boundary problem of kernel estimators by requiring that w_j have support inside an ε-shrunk subset of $\mathcal{X}$. Note that Assumptions 1(i) and (iii) require that $\mathcal{S}_j$ be a bounded set for each $j \in \mathcal{J}$. The conditions for the kernel function in Assumption 2 are quite flexible, except that the kernel functions have bounded support.

Define for $j, k \in \mathcal{J}$ and $x \in \mathbf{R}^d$,

$$\begin{aligned}
\rho_{jk,n}(x) &\equiv \frac{1}{h^d}\mathbf{E}\left[Y_{ji}Y_{ki}K^2\left(\frac{x - X_i}{h}\right)\right], \\
\rho_{jn}^2(x) &\equiv \frac{1}{h^d}\mathbf{E}\left[Y_{ji}^2 K^2\left(\frac{x - X_i}{h}\right)\right], \\
\rho_{jk}(x) &\equiv \mathbf{E}\left[Y_{ji}Y_{ki}|X_i = x\right] f(x) \int K^2(u)du, \text{ and} \\
\rho_j^2(x) &\equiv \mathbf{E}[Y_{ji}^2|X_i = x] f(x) \int K^2(u)du.
\end{aligned}$$

Let $\mathbb{Z}_1$ and $\mathbb{Z}_2$ denote mutually independent standard normal random variables. We introduce the following quantities:

$$\begin{aligned}
a_{jn} &\equiv h^{-d/2}\int_{\mathcal{X}} \rho_{jn}^p(x)w_j(x)dx \cdot \mathbf{E}\Lambda_p(\mathbb{Z}_1) \text{ and} \\
\sigma_{jk,n} &\equiv \int_{\mathcal{X}} q_{jk,p}(x)\rho_{jn}^p(x)\rho_{kn}^p(x)w_j(x)w_k(x)dx,
\end{aligned} \tag{3.3}$$

where $q_{jk,p}(x) \equiv \int_{[-1,1]^d} Cov(\Lambda_p(\sqrt{1 - t_{jk}^2(x,u)}\mathbb{Z}_1 + t_{jk}(x,u)\mathbb{Z}_2),\ \Lambda_p(\mathbb{Z}_2))du$ and

$$t_{jk}(x,u) \equiv \frac{\rho_{jk}(x)}{\rho_j(x)\rho_k(x)} \cdot \frac{\int K(x)K(x+u)dx}{\int K^2(x)dx}.$$

Let Σ_n be a $J \times J$ matrix whose (j,k)-th entry is given by $\sigma_{jk,n}$. Later we use Σ_n to normalize the test statistic. The scale normalization matrix Σ_n does not depend on x, and this is not because we are assuming conditional homoskedasticity in the null hypothesis, but because Σ_n is constituted by covariances of random quantities that already have x integrated out. We also define Σ to be a $J \times J$ matrix whose (j,k)-th entry is given by σ_{jk}, where

$$\sigma_{jk} \equiv \int_{\mathcal{X}} q_{jk,p}(x)\rho_j^p(x)\rho_k^p(x)w_j(x)w_k(x)dx.$$

As for Σ, we introduce the following assumption.

Assumption 3: Σ is positive definite.

For example, Assumption 3 excludes the case where Y_{ji} and Y_{ki} $(j \neq k)$ are perfectly correlated conditional on $X_i = x$ for almost all x with $w_j \equiv w_k$.

The following theorem is the first main result of this paper.

Theorem 1: *Suppose that Assumptions 1-3 hold and that $h \to 0$ and $n^{-1/2}h^{-d} \to 0$ as $n \to \infty$. Furthermore, assume that $m_j(x) = 0$ for almost all $x \in \mathcal{X}$ and for all $j \in \mathcal{J}$. Then*

$$T_n \equiv \frac{1}{\sigma_n}\sum_{j=1}^{J}\left\{n^{p/2}h^{(p-1)d/2}\Gamma_j(\hat{g}_{jn}) - a_{jn}\right\} \xrightarrow{d} N(0,1),$$

where $\sigma_n^2 \equiv \mathbf{1}'\Sigma_n\mathbf{1}$, and $\mathbf{1}$ is a vector of ones.

Note that when $J = 1$, σ_n^2 takes the simple form of $q_p \int_{\mathcal{X}} \rho_{1n}^{2p}(x)w_1^2(x)dx$, where

$$\begin{aligned} q_p &\equiv \int_{[-1,1]^d} Cov(\Lambda_p(\sqrt{1-t^2(u)}\mathbb{Z}_1 + t(u)\mathbb{Z}_2),\ \Lambda_p(\mathbb{Z}_2))du, \text{ and} \\ t(u) &\equiv \int K(x)K(x+u)dx \Big/ \int K^2(x)dx. \end{aligned}$$

To develop a feasible testing procedure, we construct estimators of a_{jn}'s and σ_n^2 as follows. First, define

$$\begin{aligned} \hat{\rho}_{jk,n}(x) &\equiv \frac{1}{nh^d}\sum_{i=1}^{n} Y_{ji}Y_{ki}K^2\left(\frac{x-X_i}{h}\right), \text{ and} \\ \hat{\rho}_{jn}^2(x) &\equiv \frac{1}{nh^d}\sum_{i=1}^{n} Y_{ji}^2K^2\left(\frac{x-X_i}{h}\right). \end{aligned} \tag{3.4}$$

We estimate a_{jn} and $\sigma_{jk,n}$ by:

$$\begin{aligned} \hat{a}_{jn} &\equiv h^{-d/2}\int_{\mathcal{X}} \hat{\rho}_{jn}^p(x)w_j(x)dx \cdot \mathbf{E}\Lambda_p(\mathbb{Z}_1) \text{ and} \\ \hat{\sigma}_{jk,n} &\equiv \int_{\mathcal{X}} \hat{q}_{jk,p}(x)\hat{\rho}_{jn}^p(x)\hat{\rho}_{kn}^p(x)w_j(x)w_k(x)dx, \end{aligned}$$

where $\hat{q}_{jk,p}(x) \equiv \int_{[-1,1]^d} Cov(\Lambda_p(\sqrt{1-\hat{t}_{jk}^2(x,u)}\mathbb{Z}_1 + \hat{t}_{jk}(x,u)\mathbb{Z}_2),\ \Lambda_p(\mathbb{Z}_2))du$ and

$$\hat{t}_{jk}(x,u) \equiv \frac{\hat{\rho}_{jk,n}(x)}{\hat{\rho}_{jn}(x)\hat{\rho}_{kn}(x)} \cdot \frac{\int K(x) K(x+u)\, dx}{\int K^2(x)\, dx}.$$

Note that $\mathbf{E}\Lambda_1(\mathbb{Z}_1) = 1/\sqrt{2\pi} \approx 0.39894$ and $\mathbf{E}\Lambda_2(\mathbb{Z}_1) = 1/2$. When p is an integer, the covariance expression in $q_{jk,p}(x)$ can be computed using the moment generating function of a truncated multivariate normal distribution (Tallis, 1961). More practically, simulated draws from $\mathbb{Z}_1$ and $\mathbb{Z}_2$ can be used to compute the quantities $\mathbf{E}\Lambda_p(\mathbb{Z}_1)$ and $q_{jk,p}(x)$ for general values of p. The integrals appearing above can be evaluated using methods of numerical integration. We define $\hat{\Sigma}_n$ to be a $J \times J$ matrix whose (j,k)-th entry is given by $\hat{\sigma}_{jk,n}$.

Let $\hat{\sigma}_n^2 \equiv \mathbf{1}'\hat{\Sigma}_n\mathbf{1}$. Our test statistic is taken to be

$$\hat{T}_n \equiv \frac{1}{\hat{\sigma}_n} \sum_{j=1}^{J} \left\{ n^{p/2} h^{(p-1)d/2} \Gamma_j(\hat{g}_{jn}) - \hat{a}_{jn} \right\}. \tag{3.5}$$

Let $z_{1-\alpha} \equiv \Phi^{-1}(1-\alpha)$, where Φ denotes the cumulative distribution function of $N(0,1)$. This paper proposes using the following test:

$$\text{Reject } H_0 \text{ if and only if } \hat{T}_n > z_{1-\alpha}. \tag{3.6}$$

The following theorem shows that the test has an asymptotically valid size.

Theorem 2: *Suppose that Assumptions 1-3 hold and that $h \to 0$ and $n^{-1/2}h^{-3d/2} \to 0$, as $n \to \infty$. Furthermore, assume that the kernel function K in Assumption 2 is nonnegative. Then under the null hypothesis, we have*

$$\lim_{n\to\infty} P\{\hat{T}_n > z_{1-\alpha}\} \leq \alpha,$$

with equality holding if $m_j(x) = 0$ for almost all $x \in \mathcal{X}$ and for all $j \in \mathcal{J}$.

Note that the probability of making an error of rejecting the true null hypothesis is largest when $m_j(x) = 0$ for almost all $x \in \mathcal{X}$ and for all $j \in \mathcal{J}$, namely, when we are in *the least favorable case of the null hypothesis.*

The nonparametric test does not require assumptions for m_j's and f beyond those in Assumption 1(i), even after replacing a_{jn}'s and σ_n^2 by their estimators. In particular, the theory does not require continuity or differentiability of f or m_j's. This is because we do not need to control the bias to implement the test. This result uses the assumption that the kernel function K is nonnegative to control the size of the test. (See the proof of Theorem 2 for details.)

The bandwidth condition for Theorem 2 is stronger than that in Theorem 1. This is mainly due to the treatment of the estimation errors in $\hat{a}_{jn}$ and $\hat{\sigma}_n^2$. For the bandwidth parameter, it suffices to take $h = c_1 n^{-s}$ with $0 < s < 1/(3d)$ for a constant $c_1 > 0$. In general, optimal bandwidth choice for nonparametric testing is different from that for nonparametric estimation as we need to balance the size and power of the test instead of the bias and variance of an estimator. For example, Gao and Gijbels (2008) considered testing a parametric null hypothesis against a nonparametric alternative and derived a bandwidth-selection rule by utilizing an Edgeworth expansion of the asymptotic distribution of the test statistic concerned. The methods of Gao and Gijbels (2008) are not directly applicable to our tests, and it is a challenging problem to develop a theory of optimal bandwidths for our tests. We provide some simulation evidence regarding sensitivity to the choice of h in Section 5.

According to Theorems 1-2, each choice of the weight functions w_j leads to an asymptotically valid test. The actual choice of w_j may reflect the relative importance of individual inequality restrictions. When it is of little practical significance to treat individual inequality restrictions differently, one may choose simply $w_j(x) = 1\{x \in \mathcal{S}\}$ with some common support $\mathcal{S}$. Perhaps more naturally, to avoid undue influences of different scales across Y_{ji}'s, one may use $w_j(x) = \tilde{\sigma}_{jj,n}^{-1/2} \bar{w}(x)$, for some common nonnegative weight function $\bar{w}(x)$, where

$$\tilde{\sigma}_{jj,n} \equiv q_p \int_{\mathcal{X}} \hat{\rho}_{jn}^{2p}(x) \bar{w}^2(x) dx, j \in \mathcal{J},$$

where $\hat{\rho}_{jn}^2(x)$ is given as in (3.4). Then $\tilde{\sigma}_{jj,n}$ is consistent for $\sigma_{jj,n}$ (see the proof of Theorem 2), and just as the estimation error of $\hat{\sigma}_n$ in (3.6) leaves the limiting distribution of T_n under the null hypothesis intact, so does the estimation error of $\tilde{\sigma}_{jj,n}$.

The following result shows the consistency of the test in (3.6) against fixed alternatives.

Theorem 3: *Suppose that Assumptions 1-3 hold and that $h \to 0$ and $n^{-1/2} h^{-3d/2} \to 0$, as $n \to \infty$. If H_1 is true and $\Gamma_j(g_j) > 0$ for some $j \in \mathcal{J}$, then we have*

$$\lim_{n \to \infty} P\{\hat{T}_n > z_{1-\alpha}\} = 1.$$

3.3. **Local Asymptotic Power.** We determine the power of the test in (3.6) against some sequences of local alternatives. Consider the following sequences of local alternatives converging to the null hypothesis at the rate $n^{-1/2}$, respectively:

$$H_\delta : g_j(x) = n^{-1/2} \delta_j(x), \text{ for each } j \in \mathcal{J}, \tag{3.7}$$

where $\delta_j(\cdot)$'s are bounded real functions on $\mathbf{R}^d$.

The following theorem establishes a representation of the local asymptotic power functions, when $p \in \{1, 2\}$. For simplicity of notation, let us introduce the following definition: for $s \in \{1, 2\}$, $z \in \{-1, 0, 1\}$, a given weight function vector $w \equiv (w_1, \cdots, w_J)$, and the direction $\delta = (\delta_1, \cdots, \delta_J)'$, let $\eta_{s,z}(w, \delta) \equiv \sum_{j=1}^J \int_{\mathcal{X}} \delta_j^s(x)\rho_j^z(x)w_j(x)dx$, and let $\sigma^2 \equiv \mathbf{1}'\Sigma\mathbf{1}$.

Theorem 4: *Suppose that Assumptions 1-3 hold and that $h \to 0$ and $n^{-1/2}h^{-3d/2} \to 0$, as $n \to \infty$.*
(i) *If $p = 1$, then, under H_δ, we have*

$$\lim_{n\to\infty} P\{\hat{T}_n > z_{1-\alpha}\} = 1 - \Phi(z_{1-\alpha} - \eta_{1,0}(w, \delta)/2\sigma).$$

(ii) *If $p = 2$,* then, *under H_δ, we have*

$$\lim_{n\to\infty} P\{\hat{T}_n > z_{1-\alpha}\} = 1 - \Phi(z_{1-\alpha} - \eta_{1,1}(w, \delta)/(\sigma\sqrt{\pi/2})).$$

Theorem 4 gives explicit local asymptotic power functions under H_δ, when $p = 1$ and $p = 2$. The local power of the test is greater than the size α, whenever the "noncentrality parameter" ($\eta_{1,0}(w, \delta)/2\sigma$ in the case of $p = 1$ and $\eta_{1,1}(w, \delta)/(\sigma\sqrt{\pi/2})$ in the case of $p = 2$) is strictly positive. For example, when $J = 1$ and $p = 1$ (or $p = 2$), the test is asymptotically locally strictly unbiased as long as $\mu_\delta \equiv \int_{\mathcal{X}} \delta_1(x)w_1(x)dx > 0$ (or $\int_{\mathcal{X}} \delta_1(x)\rho_1(x)w_1(x)dx > 0$). Notice that μ_δ can be strictly positive even if $\delta_1(x)$ takes negative values for some $x \in \mathcal{X}$. Therefore, our test has nontrivial local power against some, though not all, $n^{-1/2}$-local alternatives.

On the other hand, if the noncentrality parameter is zero, the test still has nontrivial power against local alternatives converging to the null at the $n^{-1/2}h^{-d/4}$ rate, which is slower than $n^{-1/2}$. To show this, consider the following local alternatives:

$$H_\delta^* : g_j(x) = n^{-1/2}h^{-d/4}\delta_j(x), \text{ for each } j \in \mathcal{J},$$

where $\delta_j(\cdot)$'s are bounded real functions as before. Theorem 4* gives the local asymptotic power functions against H_δ^*.

Theorem 4*: *Suppose that Assumptions 1-3 hold and that $h \to 0$ and $n^{-1/2}h^{-3d/2} \to 0$, as $n \to \infty$.*

(i) *If* $p = 1$ *and* $\eta_{1,0}(w, \delta) = 0$, *then, under* H_δ^*, *we have*

$$\lim_{n\to\infty} P\{\hat{T}_n > z_{1-\alpha}\} = 1 - \Phi(z_{1-\alpha} - \eta_{2,-1}(w, \delta)/\sqrt{8\pi}\sigma).$$

(ii) *If* $p = 2$ *and* $\eta_{1,1}(w, \delta) = 0$, *then, under* H_δ^*, *we have*

$$\lim_{n\to\infty} P\{\hat{T}_n > z_{1-\alpha}\} = 1 - \Phi(z_{1-\alpha} - \eta_{2,0}(w, \delta)/2\sigma).$$

If $\eta_{1,0}(w, \delta) = 0$ in the case of $p = 1$ or $\eta_{1,1}(w, \delta) = 0$ in the case of $p = 2$, then the local power of the test is greater than the size α because the new noncentrality parameter in Theorem 4* is strictly positive. For example, when $J = 1$, we have $\eta_{2,-1}(w, \delta) = \int_{\mathcal{X}} \delta_1^2(x)\rho_1^{-1}(x)w_1(x)dx > 0$ (and $\eta_{2,0}(w, \delta) = \int_{\mathcal{X}} \delta_1^2(x)w_1(x)dx > 0$) for all δ_1. Therefore, when $\eta_{1,0}(w, \delta) = 0$ or $\eta_{1,1}(w, \delta) = 0$, Theorem 4* implies that our test is strictly locally unbiased against the $n^{-1/2}h^{-d/4}$ local alternatives H_δ^*, though it has only trivial local power ($= \alpha$) against the $n^{-1/2}$ local alternatives H_δ.

To explain the results of Theorems 4 and 4* more intuitively, consider the test statistic T_n with $J = 1,\ p = 2$ and $d = 1$. For simplicity, take $w(\cdot) = 1$. Let $\sigma \equiv q_2 \int_{\mathcal{X}} \rho_1^4(x)dx$ and $a_n \equiv h^{-1/2} \int_{\mathcal{X}} \rho_1^2(x)dx \cdot \mathbf{E}\Lambda_2(\mathbb{Z}_1)$. Let the alternative hypothesis be given by

$$H_\delta^* : g(x) = n^{-1/2}h^{-b}\delta_1(x),$$

where $b = 0$ or $1/4$. Consider the statistic $\hat{T}_n$ with $\hat{\sigma}_n$ and $\hat{a}_n$ replaced by their population analogues σ_n and a_n, respectively, i.e.,

$$\begin{aligned} T_n &\equiv \frac{1}{\sigma_n}\left\{nh^{1/2}\int_{\mathcal{X}} \Lambda_2\left(\hat{g}_n(x)\right)dx - a_n\right\} \\ &= \frac{nh^{1/2}}{\sigma_n}\left\{\int_{\mathcal{X}} \Lambda_2\left(\hat{g}_n(x)\right)dx - \int_{\mathcal{X}} E\Lambda_2\left(\hat{g}_n(x)\right)dx\right\} \\ &\quad + \frac{nh^{1/2}}{\sigma_n}\left\{\int_{\mathcal{X}} E\Lambda_2\left(\hat{g}_n(x)\right)dx - \frac{a_n}{nh^{1/2}}\right\}. \end{aligned} \tag{3.8}$$

It is easy to see that T_n has the same asymptotic distribution as $\hat{T}_n$ under the local alternative hypothesis. The first term on the right hand side of (3.8) converges in distribution to the standard normal distribution by the arguments similar to those used to prove Theorem 1. Consider the second term in (3.8). We can approximate it by

$$\frac{1}{\sigma_n}\left\{nh^{1/2}\int_{\mathcal{X}} E\Lambda_2\left(\hat{g}_n(x)\right)dx - a_n\right\}$$

$$= \frac{1}{\sigma_n}\left\{\int_{\mathcal{X}} E\Lambda_2\left(h^{-1/4}\rho_1(x)\frac{\sqrt{nh}\left[\hat{g}_n(x) - E\hat{g}_n(x)\right]}{\rho_1(x)} + n^{1/2}h^{1/4}E\hat{g}_n(x)\right)dx - a_n\right\}$$

$$\simeq \sigma^{-1}\int_{\mathcal{X}} E\Lambda_2\left(h^{-1/4}\rho_1(x)\mathbb{Z}_1 + n^{1/2}h^{1/4}E\hat{g}_n(x)\right)dx - \sigma^{-1}a_n \tag{3.9}$$

$$\simeq \sigma^{-1}\int_{\mathcal{X}} \left\{E\Lambda_2\left(h^{-1/4}\rho_1(x)\mathbb{Z}_1 + h^{1/4-b}\delta_1(x)\right) - E\Lambda_2\left(h^{-1/4}\rho_1(x)\mathbb{Z}_1\right)\right\}dx \tag{3.10}$$

$$\simeq h^{-b}\left(\frac{2\phi(0)}{\sigma}\int_{\mathcal{X}}\delta_1(x)\rho_1(x)dx\right) + h^{1/2-2b}\left(\frac{1}{2\sigma}\int_{\mathcal{X}}\delta_1^2(x)dx\right), \tag{3.11}$$

where (3.9) follows from the Poissonization argument, (3.10) holds by $n^{1/2}h^{1/4}E\hat{g}_n(x) = h^{1/4-b}\int\delta_1(x-uh)K(u)du \simeq h^{1/4-b}\delta_1(x)$, and (3.11) uses a Taylor expansion $\mathbf{E}\Lambda_2(\gamma\mathbb{Z}_1+\mu) - \mathbf{E}\Lambda_2(\gamma\mathbb{Z}_1) \simeq 2\phi(0)\mu\gamma + \Phi(0)\mu^2$ with $\gamma = h^{-1/4}\rho_1(x)$ and $\mu = h^{1/4-b}\delta_1(x)$, where $\phi(\cdot)$ and $\Phi(\cdot)$, respectively, denote the pdf and cdf of the standard normal distribution. This approximation tells us that if $\int_{\mathcal{X}}\delta_1(x)\rho_1(x)dx > 0$, we can take $b = 0$ so that it can achieve nontrivial power against $n^{-1/2}$ alternatives, while if $\int_{\mathcal{X}}\delta_1(x)\rho_1(x)dx = 0$, then we should take $b = 1/4$ so that it has nontrivial local power against $n^{-1/2}h^{-1/4}$ local alternatives. Notice that, in the latter case, $\int_{\mathcal{X}}\delta_1^2(x)dx$ is always positive.

It would also be interesting to compare local power properties of our test with that of Andrews and Shi (2011a). Unlike our test, the test of Andrews and Shi (2011a, Theorem 4(b)) does not require $\int_{\mathcal{X}}\delta_1(x)\rho_1(x)dx > 0$, but excludes some $n^{-1/2}$-local alternatives. An analytical and unambiguous comparison between the two approaches is not straightforward, because the test of Andrews and Shi (2011a) is not asymptotically distribution free, meaning that the local power function may depend on the underlying data generating process in a complicated way. However, we do compare the two approaches in our simulation studies.

When $J = 1$, thanks to Theorem 4, we can compute an optimal weight function that maximizes the local power against a given direction δ. See Stute (1997) for related results of optimal directional tests, and Tripathi and Kitamura (1997) for results of optimal directional and average tests based on smoothed empirical likelihoods.

Define $\sigma_p^2(w_1) \equiv q_p\int_{\mathcal{X}}\rho_{1n}^{2p}(x)w_1^2(x)dx$ for $J = 1$. The optimal weight function (denoted by w_p^*) is taken to be a maximizer of the drift term $\eta_{1,0}(w_1,\delta_1)/\sigma_1(w_1)$ (in the case of $p = 1$) or $\eta_{1,1}(w_1,\delta_1)/\sigma_2(w_1)$ (in the case of $p = 2$) with respect to w_1 under the constraint that $w_1 \geq 0$ and $\int_{\mathcal{X}} w_1(x)\rho^{2p}(x)dx = 1$. The latter condition is for a scale normalization. Let $\delta_1^+ = \max\{\delta_1, 0\}$. Since ρ_1 and w_1 are nonnegative, the Cauchy-Schwarz inequality suggests that the optimal weight function is given by

$$w_p^*(x) = \begin{cases} \frac{\delta_1^+(x)\rho_1^{-2}(x)}{\sqrt{\int_{\mathcal{X}}(\delta_1^+)^2(x)\rho_1^{-2}(x)dx}}, & \text{if } p = 1, \text{ and} \\ \frac{\delta_1^+(x)\rho_1^{-3}(x)}{\sqrt{\int_{\mathcal{X}}(\delta_1^+)^2(x)\rho_1^{-2}(x)dx}}, & \text{if } p = 2. \end{cases} \tag{3.12}$$

To satisfy Assumption 1(iii), we assume that the support of δ_1 is contained in an ε-shrunk subset of $\mathcal{X}$. With this choice of an optimal weight function, the local power function becomes:

$$\begin{aligned} &1 - \Phi\left(z_{1-\alpha} - \frac{\sqrt{\int_{\mathcal{X}}(\delta_1^+)^2(x)\rho_1^{-2}(x)dx}}{2\sqrt{q_1}}\right), \quad \text{if } p = 1, \text{ and} \\ &1 - \Phi\left(z_{1-\alpha} - \frac{\sqrt{\int_{\mathcal{X}}(\delta_1^+)^2(x)\rho_1^{-2}(x)dx}}{\sqrt{q_2\pi/2}}\right), \quad \text{if } p = 2. \end{aligned}$$

4. Comparison with Testing Functional Equalities

It is straightforward to follow the proofs of Theorems 1-3 to develop a test for equality restrictions:

$$\begin{aligned} H_0 &: m_j(x) = 0 \text{ for all } (x,j) \in \mathcal{X} \times \mathcal{J}, \text{ vs.} \\ H_1 &: m_j(x) \neq 0 \text{ for some } (x,j) \in \mathcal{X} \times \mathcal{J}. \end{aligned} \tag{4.1}$$

For this test, we redefine $\Lambda_p(v) = |v|^p$ and, using this, redefine $\hat{T}_n$ in (3.5) and σ^2. Then under the null hypothesis,

$$\hat{T}_n \xrightarrow{d} N(0,1).$$

Therefore, we can take a critical value in the same way as before. The asymptotic validity of this test under the null hypothesis in (4.1) follows under precisely the same conditions as in Theorem 2. However, the convergence rates of the inequality tests and the equality tests under local alternatives are different, as we shall see now.

Consider the local alternatives converging to the null hypothesis at the rate $n^{-1/2}h^{-d/4}$:

$$H_\delta^* : g_j(x) = n^{-1/2}h^{-d/4}\delta_j(x), \text{ for each } j \in \mathcal{J}, \tag{4.2}$$

where $\delta_j(\cdot)$'s are again bounded real functions on $\mathbf{R}^d$. The following theorem establishes the local asymptotic power functions of the test based on $\hat{T}_n$.

Theorem 5: *Suppose that Assumptions 1-3 hold and that $h \to 0$ and $n^{-1/2}h^{-3d/2} \to 0$, as $n \to \infty$.*

(i) *If* $p = 1$, *then under* H_δ^*, *we have*

$$\lim_{n\to\infty} P\{\hat{T}_n > z_{1-\alpha}\} = 1 - \Phi(z_{1-\alpha} - \eta_{2,-1}(w,\delta)/(\sqrt{2\pi}\sigma)).$$

(ii) *If* $p = 2$, *then under* H_δ^*, *we have*

$$\lim_{n\to\infty} P\{\hat{T}_n > z_{1-\alpha}\} = 1 - \Phi(z_{1-\alpha} - \eta_{2,0}(w,\delta)/\sigma).$$

Theorem 5 shows that the equality tests (on (4.1)), in contrast to the inequality tests (on (1.1)), have nontrivial local power against alternatives converging to the null at rate $n^{-1/2}h^{-d/4}$, which is slower than $n^{-1/2}$. This phenomenon of different convergence rates arises because Λ_p is symmetric around zero in the case of equality tests, and it is not in the case of inequality tests. To see this closely, observe that in the case of $p = 1$, the power comparison between the equality test and the inequality test is reduced to comparison between $\mathbf{E}|\mathbb{Z}_1 + \mu| - \mathbf{E}|\mathbb{Z}_1|$ and $\mathbf{E}\max\{\mathbb{Z}_1 + \mu, 0\} - \mathbf{E}\max\{\mathbb{Z}_1, 0\}$ for μ close to zero, where $\mathbb{Z}_1$ follows a standard normal distribution with ϕ denoting its density. Note that we can approximate $\mathbf{E}|\mathbb{Z}_1 + \mu| - \mathbf{E}|\mathbb{Z}_1|$ by $\{\phi''(0) + 2\phi(0)\}\mu^2$ for μ close to zero, and approximate $\mathbf{E}\max\{\mathbb{Z}_1 + \mu, 0\} - \mathbf{E}\max\{\mathbb{Z}_1, 0\}$ by $\Phi(0)\mu$ for μ close to zero. The smaller scale μ^2 in the former case arises because the leading term in the expansion of $\mathbf{E}|\mathbb{Z}_1 + \mu| - \mathbf{E}|\mathbb{Z}_1|$ around $\mu = 0$ disappears due to the symmetry of the absolute value function $|\cdot|$. Therefore, the different rate of convergence arises due to our symmetric treatment of the alternative hypotheses (positive or negative) in the equality test, in contrast to the asymmetric treatment in the inequality test.

Since $\eta_{2,-1}(w,\delta)$ and $\eta_{2,0}(w,\delta)$ are always nonnegative, the equality tests are *locally asymptotically unbiased* against any local alternatives. In contrast, the terms $\eta_{1,0}(w,\delta)$ and $\eta_{1,1}(w,\delta)$ in the local asymptotic power functions of the inequality tests in Theorem 4 can take negative values for some local alternatives, implying that the inequality tests might be asymptotically biased against such local alternatives. This feature is not due to the form of our proposed inequality test, but is rather a common feature in testing moment inequalities. It is because the null hypothesis is given by a composite hypothesis and most of the powerful tests are not similar on the boundary and hence biased against some local alternatives. In principle, one can construct a test that is asymptotically similar on the boundary, but such a test has typically poor power. See Andrews (2011) for details.

The test in Theorem 5 shares some features common in nonparametric tests that are known to detect some smooth local alternatives that have narrow peaks as the sample size increases. See e.g. Fan and Li (2000) and references therein. To see this closely, consider a sequence of non-Pitman local alternatives of type:

$$H^*_{\delta_n} : g_j(x) = \gamma_n \delta_{j,n}(x), \text{ for each } j \in \mathcal{J},$$

where γ_n is a deterministic sequence and $\delta_{j,n}(x)$ is now allowed to change over n. For example, one may consider $\delta_{j,n}(x)$ to be a function with a single peak that becomes sharper as n becomes large, e.g. $\delta_{j,n}(x) = L_j((x - x_0)/\zeta_n)$, where $L_j(\cdot)$ is a bounded function, $x_0 \in \mathbf{R}^d$ is a fixed point, and $\zeta_n \to 0$ as $n \to \infty$. By using the same arguments as in the proof of Theorem 5, we can show that the two-sided version of our test has nontrivial power against such local alternatives provided $\lim_{n\to\infty} nh^{d/2}\gamma_n^2\eta_{2,-1}(w, \delta_{j,n}) \neq 0$ (for $p = 1$) or $\lim_{n\to\infty} nh^{d/2}\gamma_n^2\eta_{2,0}(w, \delta_{j,n}) \neq 0$ (for $p = 2$). However, since our main interest lies in testing functional inequalities, we will not pursue further local power properties of the equality test. On the other hand, it would also be interesting to see whether it would give an adaptive, rate-optimal test to take the supremum of our two-sided version of our test over a set of bandwidths, as in Horowitz and Spokoiny (2001). However, the latter study is beyond of the scope of this paper.

As in Section 3.3, when $J = 1$, an optimal directional test under (4.2) can also be obtained by following the arguments leading up to (3.12) so that

$$w_p^*(x) = \begin{cases} \frac{\delta_1^2(x)\rho_1^{-3}(x)}{\sqrt{\int_{\mathcal{X}} \delta_1^4(x)\rho_1^{-4}(x)dx}}, & \text{if } p = 1, \text{ and} \\ \frac{\delta_1^2(x)\rho_1^{-4}(x)}{\sqrt{\int_{\mathcal{X}} \delta_1^4(x)\rho_1^{-4}(x)dx}}, & \text{if } p = 2. \end{cases}$$

Similarly as before, let the support of δ_1 be contained in an ε-shrunk subset of $\mathcal{X}$. The optimal weight function yields the following local power functions:

$$\begin{aligned} &1 - \Phi\left(z_{1-\alpha} - \frac{\sqrt{\int \delta_1^4(x)\rho_1^{-4}(x)dx}}{\sqrt{2\pi}\bar{q}_1}\right), \quad \text{if } p = 1, \text{ and} \\ &1 - \Phi\left(z_{1-\alpha} - \frac{\sqrt{\int \delta_1^4(x)\rho_1^{-4}(x)dx}}{\sqrt{\bar{q}_2}}\right), \quad \text{if } p = 2, \end{aligned}$$

where $\bar{q}_p \equiv \int_{[-1,1]^d} Cov(|\sqrt{1 - t^2(u)}\mathbb{Z}_1 + t(u)\mathbb{Z}_2|^p, \ |\mathbb{Z}_2|^p)du$, for $p \in \{1, 2\}$.

5. Monte Carlo Experiments

This section reports the finite-sample performance of the one-sided L_1- and L_2-type tests from a Monte Carlo study. In the experiments, n observations of a pair of random variables (Y, X) were generated from $Y = m(X) + \sigma(X)U$, where $X \sim$ Unif$[0, 1]$ and $U \sim N(0, 1)$ and X and U are independent. In all the experiments, we set $\mathcal{X} = [0.05, 0.95]$.

To evaluate the finite-sample size of the tests, we first set $m(x) \equiv 0$. We call this case DGP0. In addition, we consider the following alternative model

$$m(x) = x(1 - x) - c_m \tag{5.1}$$

where $c_m \in \{0.25, 0.20, 0.15, 0.10, 0.05\}$. We call these 5 cases DGPs 1-5. When $c_m = 0.25$ (DGP1), we have $m(x) < 0$ for all $x \neq 0.5$ and $m(x) = 0$ with $x = 0.5$. Hence, this case corresponds to the "interior" of the null hypothesis. In view of asymptotic theory, we expect the empirical probability of rejecting H_0 to converge to zero as n gets large. When $c_m < 0.25$ (DGPs 2-5), we have $m(x) > 0$ for some x. Therefore, these four cases are considered to see the finite-sample power of our tests. Two different functions of $\sigma(x)$ are considered: $\sigma(x) \equiv 1$ (homoskedastic error) and $\sigma(x) = x$ (heteroskedastic error).

The experiments use sample sizes of $n = 50, 200, 1000$ and the nominal level of $\alpha = 0.05$. We performed 1000 Monte Carlo replications in each experiment. In implementing both L_1 and L_2-type tests, we used $K(u) = (3/2)(1 - (2u)^2)I(|u| \leq 1/2)$ and $h = c_h \times \hat{s}_X \times n^{-1/5}$, where $I(A)$ is the usual indicator function that has value one if A is true and zero otherwise, c_h is a constant and $\hat{s}_X$ is the sample standard deviation of X. To check the sensitivity to the choice of the bandwidth, eight different values of c_h are considered: $\{0.75, 1, 1.25, 1.5, 1.75, 2, 2.25, 2.50\}$. Finally, we considered the uniform weight function: $w(x) = 1$ and the inverse standard error weight function: $w(x) = 1/\rho_n(x)$.

To evaluate the relative performance of our test, we have also implemented one of test statistics proposed by Andrews and Shi (2011a), specifically their Cramér-von Mises-type (CvM) statistic with both plug-in asymptotic (PA/Asy) and asymptotic generalized moment selection (GMS/Asy) critical values. Specifically, countable hypercubes are used as instrument functions, and tuning parameters were chosen, following suggestions as in Section 9 of Andrews and Shi (2011a).

Empirical rejection probabilities are plotted in Figures 1-4. 8 different solid lines in each panel correspond to our test with 8 different bandwidth values. 2 dotted lines correspond to the test of Andrews and Shi (2011a) with PA and GMS critical values. For each case, the test with the GMS critical value gives slightly higher rejection probabilities than that with the PA critical value. When H_0 is true and $m(x) \equiv 0$ (DGP0), the differences between the nominal and empirical rejection probabilities are small. When H_0 is true and $m(x)$ is (5.1) with $c_m = 0.25$ (the interior case DGP1), the empirical rejection probabilities are smaller than the nominal level and become almost zero for $n = 1000$.

When H_0 is false and the correct model is (5.1) with $c_m < 0.25$ (DGPs 2-5), the power of both the L_1 and L_2 tests is increasing as c_m gets smaller. This finding is consistent with asymptotic theory since it is likely that our test will be more powerful when $\int_{\mathcal{X}} m(x)w(x)dx$ is larger. Note that in DGPs 3-5, ($c_m = 0.15, 0.10, 0.05$), the rejection probabilities increase as n gets large. This is in line with the asymptotic theory in the preceding sections, for our test is consistent for these values of c_m. However, the rejection probabilities are quite small even with $n = 1000$ for $c_m = 0.20$ (DGP 2). This is not surprising given that our test can be biased, as shown in Section 3.3. To further investigate the issue of bias associated with $\int_{\mathcal{X}} m(x)w(x)dx$, we carried out an additional simulation with $m(x) = \sin(2\pi x)$. It turns out that rejection probabilities were almost one across different values of the bandwidth for both weight functions and for both homoskedastic and heteroskedastic errors. This seems to be consistent with Theorem 4* in Section 3.3. We do not report full details of additional simulation results for brevity.

Simulation results for the CvM statistics are similar to our test statistics. More precisely, in Figure 1 (the homoskedasticity case), the L_1 test with both weight functions seems to be more powerful than Andrews and Shi's test, whereas in Figure 4, their test appears to be more powerful than the L_2 test with the uniform weight. However, for most cases, power performances are comparable between each other. Note further that there is little difference between PA and GMS critical values for the CvM statistic of Andrews and Shi (2011a). This is due to the fact that $m(x)$ is either flat or has a maximum at a single point. We note that the results are not very sensitive to the bandwidth choice for our tests. Finally, regarding the choice of the weight function, we would like to recommend the inverse standard error weight since it seems to perform better than the uniform weight in simulations.

6. Proofs

This section begins with a roadmap for the proof, where the roles of technical lemmas and main difficulties are explained. Then we state the lemmas and present the proofs of the theorems.

6.1. **The Roadmap for the Proof of Theorem 1.** The proof of Theorem 1 follows the structure of the proof of the finite-dimensional convergence in Theorem 1.1 of GMZ.

Under the condition of Theorem 1 that $m_j(x) = 0$ for almost all $x \in \mathcal{X}$ and for all $j \in \mathcal{J}$, we can show that $\mathbf{E}\hat{g}_{jn}(x) = 0$ for almost all x in the support of w_j from some large n on. This means that by letting $v_{jn}(x) \equiv \hat{g}_{jn}(x) - \mathbf{E}\hat{g}_{jn}(x)$ and $\zeta_n(A) \equiv \sum_{j=1}^J \int_A \Lambda_p(v_{jn}(x))w_j(x)dx$ with some $A \subset \mathcal{X}$, we can write T_n as

$$\begin{aligned}(6.1)\qquad & \frac{n^{p/2}h^{(p-1)d/2}}{\sigma_n}\left\{\zeta_n(\mathcal{X}\backslash A) - \mathbf{E}\zeta_n(\mathcal{X}\backslash A)\right\} \\ & + \frac{n^{p/2}h^{(p-1)d/2}}{\sigma_n}\left\{\zeta_n(A) - \mathbf{E}\zeta_n(A)\right\}.\end{aligned}$$

The main part of the proof of Theorem 1 establishes asymptotic normality for the second term and asymptotic negligibility for the first term when A is chosen to nearly cover $\mathcal{X}$.

The proof of asymptotic normality employs the Poissonization method of GMZ which prevents us from choosing A to cover $\mathcal{X}$ entirely. This makes the proof intricate. The asymptotic arguments for both terms of (6.1) require that σ_n is an asymptotically stable quantity. Hence we begin by dealing with σ_n.

Step 1: In Lemma A7, we show that given appropriate $A \subset \mathcal{X}$, $\sigma_n(A) \to \sigma(A) > 0$ as $n \to \infty$, for some $\sigma(A) > 0$, where $\sigma_n(A)$ is σ_n except that the integral domains of $\sigma_{jk,n}$ are restricted to A. To prove the convergence, we choose the domain A to be such that nonparametric functions $\rho_{jk,n}$ that constitute σ_n are continuous and uniformly convergent on this domain. That we can choose such A to be large enough is ensured by Lemma A1. The proof is lengthy, the main step being the approximation of covariances of Poissonized sums. For this approximation, we use a type of a Berry-Esseen bound for sums of independent random variables due to Sweeting (1977). This bound is restated in Lemma A2. Since the bound involves various moments of random quantities, we prepare these moment bounds in Lemmas A4 and A5.

Step 2: We establish that the second term in (6.1) is asymptotically standard normal when A nearly covers $\mathcal{X}$. First, we use Lemma A6 to show that the second component in (6.1) is asymptotically equivalent to

$$\frac{n^{p/2}h^{(p-1)d/2}}{\sigma_n}\left\{\zeta_n(A) - \mathbf{E}\zeta_N(A)\right\}, \tag{6.2}$$

where $\zeta_N(A) \equiv \sum_{j=1}^J \int_A \Lambda_p(v_{jN}(x))w_j(x)dx$, $v_{jN}(x) \equiv \hat{g}_{jN}(x) - \mathbf{E}\hat{g}_{jn}(x)$,

$$\hat{g}_{jN}(x) \equiv \frac{1}{nh^d}\sum_{i=1}^{N} Y_{ji}K\left(\frac{x-X_i}{h}\right), \tag{6.3}$$

and N is a Poisson random variable with mean n and independent of all the other random variables. Then consider

$$S_n(A) \equiv \frac{n^{p/2}h^{(p-1)d/2}\{\zeta_N(A) - \mathbf{E}\zeta_N(A)\}}{\sigma_n(A)}, \tag{6.4}$$

where $\sigma_n^2(A) \equiv \sum_{j=1}^J \sum_{k=1}^J \sigma_{jk,n}(A)$ and $\sigma_{jk,n}(A)$ is $\sigma_{jk,n}$ with the integral domain restricted to A. Note that the numerator of $S_n(A)$ is based on the Poissonized version $v_{jN}(x)$ so that when we cut the integral in $\zeta_N(A)$ into integrals on small disjoint domains and sum them, this latter sum behaves like a sum of independent random variables. In Lemma A9, we construct this sum and apply the CLT to obtain asymptotic normality for $S_n(A)$. Then in Lemma A10, using the de-Poissonization lemma of Beirlant and Mason (1995), we deduce that the conditional distribution of $S_n(A)$ given $N = n$ converges to a standard normal distribution. (This lemma requires the set $\mathcal{X}\backslash A$ to stay nonempty.) Since this conditional distribution is nothing but the distribution of (6.2), we conclude that the second term in (6.1) is asymptotically standard normal. However, this sequence of arguments so far presumes that σ_n is an asymptotically right scale, which means that σ_n should be based on the Poissonized version $v_{jN}(x)$ not on the original one $v_{jn}(x)$.

Step 3: It remains to deal with the first term in (6.1). Since $\sigma_n(A)$ is close to $\sigma(A) > 0$ by Step 1 for large samples, it suffices to show that the quantity

$$n^{p/2}h^{(p-1)d/2}\left\{\zeta_n(\mathcal{X}\backslash A) - \mathbf{E}\zeta_n(\mathcal{X}\backslash A)\right\}$$

is asymptotically negligible for large n and large set A. This is accomplished by Lemma A8, which again uses moment bounds of Lemmas A4 and A5. Since w_j is square integrable, if we can take $A \subset \mathcal{X}$ such that $\int_{\mathcal{X}/A} w_j^2(x)dx$ is small, the asymptotic negligibility of the first component in (6.1) follows by Lemma A8. Lemma 8 extends to Lemma 6.2 of GMZ from $p = 1$ to $p \geq 1$. This generalization is necessary since the majorization inequality of Pinelis (1994) used in GMZ is not directly applicable in the general case with $p \geq 1$.

Step 4: Finally, we approximate $\mathbf{E}\zeta_n(A)$ in the second component in (6.1) by an estimable quantity, $\sum_{j\in\mathcal{J}} a_{jn}$ in Theorem 1. This step is done through Lemma A6. The lemma is adapted from Lemma 6.3 of GMZ, but unlike their case of L_1-norm, our case involves the one-sided L_p-norm with $p \geq 1$. For this modification, we use the algebraic inequality of Lemma A3. This closes the proof of Theorem 1.

6.2. **Technical Lemmas and the Proof of Theorem 1.** We begin with technical lemmas. The lemmas are ordered so that lemmas that come later rely on their preceding lemmas.

The first statement of the lemma below is a special case of Theorem 2(b) of Stein (1970) on pages 62 and 63. The second statement is an extension of Lemma 6.1 of GMZ.

Lemma A1: *Let $J(\cdot) : \mathbf{R}^d \to \mathbf{R}$ be a Lebesgue integrable bounded function and $H : \mathbf{R}^d \to \mathbf{R}$ be a bounded function with compact support S. Then, for almost every $y \in \mathbf{R}^d$,*

$$\int_{\mathbf{R}^d} J(x)H_h(y-x)\,dx \to J(y)\int_S H(x)dx, \text{ as } h \to 0,$$

where $H_h(x) \equiv H(x/h)/h^d$.

Furthermore, suppose that $\bar{J} \equiv \int |J(z)|dz > 0$. Then for all $0 < \varepsilon < \bar{J} \equiv \int |J(z)|dz$, there exist $M > 0$, $v > 0$ and a Borel set B of finite Lebesgue measure $m(B)$ such *that* $B \subset [-M+v, M-v]^d$, $\alpha \equiv \int_{\mathbf{R}^d \backslash [-M,M]^d} |J(z)|dz > 0$, $\int_B |J(z)|dz > \bar{J} - \varepsilon$, J *is continuous on* B, *and*

$$\sup_{y \in B} \left| \int_{\mathbf{R}^d} J(x) H_h(y-x)\, dx - J(y) \int_S H(x) dx \right| \to 0, \text{ as } h \to 0.$$

Proof: The first statement is a special case of Theorem 2(b) of Stein (1970) on pages 62 and 63. The second statement can be proved following the proof of Lemma 6.1 of GMZ. Since J is Lebesgue integrable, the integral $\int_{\mathbf{R}^d \backslash [-M,M]^d} |J(z)|dz$ is continuous in M and converges to zero as $M \to \infty$. We can find $M > 0$ and $v > 0$ such that

$$\int_{\mathbf{R}^d \backslash [-M,M]^d} |J(z)|dz = \varepsilon/8 \text{ and } \int_{\mathbf{R}^d \backslash [-M+v,M-v]^d} |J(z)|dz = \varepsilon/4.$$

The construction of the desired set $B \subset [-M+v, M-v]^d$ can be done using the arguments in the proof of Lemma 6.1 of GMZ. ■

The following result is a special case of Theorem 1 of Sweeting (1977) with $g(x) = \min(x, 1)$ (in his notation). See also Fact 6.1 of GMZ and Fact 4 of Mason (2009) for applications of Theorem 1 of Sweeting (1977).

Lemma A2 (Sweeting (1977)): *Let* $\mathbb{Z} \in \mathbf{R}^k$ *be a mean zero normal random vector with covariance matrix* I *and* $\{W_i\}_{i=1}^n$ *is a set of i.i.d. random vectors in* $\mathbf{R}^k$ *such that* $\mathbf{E}W_i = 0$, $\mathbf{E}W_i W_i' = I$, *and* $\mathbf{E}||W_i||^r < \infty$, $r \geq 3$. *Then for any Borel measurable function* $\varphi : \mathbf{R}^k \to \mathbf{R}$ *such that*

$$\sup_{x \in \mathbf{R}^k} \frac{|\varphi(x) - \varphi(0)|}{1 + ||x||^r \min(||x||, 1)} < \infty,$$

we have

$$\begin{aligned} & \left| \mathbf{E}\left[\varphi\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n W_i \right) \right] - \mathbf{E}\left[\varphi(\mathbb{Z}) \right] \right| \\ \leq \;& c_1 \left(\sup_{x \in \mathbf{R}^k} \frac{|\varphi(x) - \varphi(0)|}{1 + ||x||^r \min(||x||, 1)} \right) \left\{ \frac{1}{\sqrt{n}} \mathbf{E}||W_i||^3 + \frac{1}{n^{(r-2)/2}} \mathbf{E}||W_i||^r \right\} \\ & + c_2 \mathbf{E}\left[\omega_\varphi \left(\mathbb{Z}; \frac{c_3}{\sqrt{n}} \mathbf{E}||W_i||^3 \right) \right], \end{aligned}$$

where c_1, c_2 *and* c_3 *are positive constants that depend only on* k *and* r *and*

$$\omega_\varphi(x; \varepsilon) \equiv \sup \left\{ |\varphi(x) - \varphi(y)| : y \in \mathbf{R}^k, ||x - y|| \leq \varepsilon \right\}.$$

The following algebraic inequality is used frequently throughout the proofs.

LEMMA A3: *For any* $a, b \in \mathbf{R}$, *let* $a_+ = \max(a, 0)$ *and* $b_+ = \max(b, 0)$. *Furthermore, for any real* $a \geq 0$, *if* $a = 0$, *we define* $\lceil a \rceil = 1$, *and if* $a > 0$, *we define* $\lceil a \rceil$ *to be the smallest integer greater than or equal to* a. *Then for any* $p \geq 1$,

$$\begin{aligned}\max\left\{|a_+^p - b_+^p|, ||a|^p - |b|^p|\right\} &\leq 2p|a-b|\left(\sum_{k=0}^{\lceil p-1 \rceil} \frac{\lceil p-1 \rceil!}{k!}|a-b|^{\lceil p-1 \rceil - k}|b|^k\right)^{(p-1)/\lceil p-1 \rceil} \\ &\leq C\sum_{k=0}^{\lceil p-1 \rceil} |a-b|^{p - \frac{(p-1)k}{\lceil p-1 \rceil}}|b|^k,\end{aligned}$$

for some $C > 0$ *that depends only on* p.

PROOF : First, we show the inequality for the case where p is a positive integer. We prove first that $||a|^p - |b|^p|$ has the desired bound. Note that in this case of p being a positive integer, the bound takes the following form:

$$2\sum_{k=0}^{p-1} \frac{p!}{k!}|a-b|^{p-k}|b|^k.$$

When $p = 1$, the bound is trivially obtained. Suppose now that the inequality holds for a positive integer q. First, note that using the mean-value theorem, convexity of the function $f(x) = |x|^q$ for $q \geq 1$, and the triangular inequality,

$$\begin{aligned}||a|^{q+1} - |b|^{q+1}| &\leq (q+1)|a-b|\sup_{\alpha \in [0,1]}\left(\alpha|a| + (1-\alpha)|b|\right)^q \\ &\leq (q+1)|a-b|\sup_{\alpha \in [0,1]}\left(\alpha|a|^q + (1-\alpha)|b|^q\right) \\ &\leq (q+1)|a-b|\left(||a|^q - |b|^q| + 2|b|^q\right).\end{aligned}$$

As for $||a|^q - |b|^q|$, we apply the inequality to bound the last term by

$$\begin{aligned}&(q+1)|a-b|\left(2\sum_{k=0}^{q-1}\frac{q!}{k!}|a-b|^{q-k}|b|^k + |b|^q\right) \\ &= 2\sum_{k=0}^{q}\frac{(q+1)!}{k!}|a-b|^{q-k+1}|b|^k.\end{aligned}$$

Therefore, by the principle of mathematical induction, the desired bound in the case of p being a positive integer follows.

Certainly, we obtain the same bound for $|a_+^p - b_+^p|$ when $p = 1$. When $p > 1$, we observe that by the mean-value theorem,

$$\begin{aligned}|a_+^p - b_+^p| &\leq p|a-b|\left(|a|^{p-1} + |b|^{p-1}\right) \\ &\leq p|a-b|\left(||a|^{p-1} - |b|^{p-1}| + 2|b|^{p-1}\right).\end{aligned}$$

By applying the previous inequality to $||a|^{p-1} - |b|^{p-1}|$, we obtain the desired bound for $|a_+^p - b_+^p|$ when p is a positive integer.

Since the bound holds for any positive integer p, let us consider the case where p is a real number strictly larger than 1. Again, we first show that $||a|^p - |b|^p|$ has the desired bound. Using the mean-value theorem as before and the fact that $|a + b| \leq 2^{1-1/s}(|a|^s + |b|^s)^{1/s}$ for all $s \in [1, \infty)$ and all $a, b \in \mathbf{R}$, we find that for $u \equiv \lceil p - 1 \rceil$,

$$\begin{aligned} ||a|^p - |b|^p| &\leq p\,|a-b|\,(|a|^{p-1} + |b|^{p-1}) \\ &\leq p\,|a-b|\,2^{1-(p-1)/u}(|a|^u + |b|^u)^{(p-1)/u} \\ &\leq p\,|a-b|\,2^{1-(p-1)/u}(||a|^u - |b|^u| + 2|b|^u)^{(p-1)/u}. \end{aligned}$$

Since u is a positive integer, using the previous bound, we bound the right-hand side by

$$p\,|a-b|\,2^{1-(p-1)/u}\left(2\sum_{k=0}^{u-1}\frac{u!}{k!}|a-b|^{u-k}|b|^k + 2|b|^u\right)^{(p-1)/u}.$$

Consolidating the sum in the parentheses, we obtain the wanted bound.

As for the second inequality, observe that

$$\begin{aligned} &2p|a-b|\left(\sum_{k=0}^{\lceil p-1\rceil}\frac{\lceil p-1\rceil!}{k!}|a-b|^{\lceil p-1\rceil-k}|b|^k\right)^{(p-1)/\lceil p-1\rceil} \\ \leq\;& C\max_{k\in\{0,1,\cdots,\lceil p-1\rceil\}}|a-b|^{p-k\{(p-1)/\lceil p-1\rceil\}}|b|^k \leq C\sum_{k=0}^{\lceil p-1\rceil}|a-b|^{p-k\{(p-1)/\lceil p-1\rceil\}}|b|^k, \end{aligned}$$

for some $C > 0$ that depends only on p. We can obtain the same bound for $|a_+^p - b_+^p|$ by noting that $|a_+^p - b_+^p| \leq p\,|a-b|\,(|a|^{p-1} + |b|^{p-1})$ and following the same arguments afterwards as before. ■

Define for $j \in \mathcal{J}$,

$$k_{jn,r}(x) \equiv h^{-d}\mathbf{E}\left[\left|Y_{ji}K\left(\frac{x - X_i}{h}\right)\right|^r\right],\ r \geq 1. \tag{6.5}$$

Lemma A4: *Suppose that Assumptions 1(i)(iii) and 2 hold and $h \to 0$ as $n \to \infty$. Then for $\varepsilon > 0$ in Assumption 1(i), there exist positive integer n_0 and constants $c_1, c_2 > 0$ such that for all $n \geq n_0$, all $r \in [1, 2p+2]$, and all $j \in \mathcal{J}$,*

$$0 < c_1 \leq \inf_{x\in\mathcal{S}_j^{\varepsilon/2}}\rho_{jn}^2(x) \text{ and}$$

$$\sup_{x\in\mathcal{S}_j^{\varepsilon/2}}k_{jn,r}(x) \leq c_2 < \infty.$$

PROOF: Since $h \to 0$ as $n \to \infty$, we apply change of variables to find that from large n on,

$$\begin{aligned} \inf_{x \in \mathcal{S}_j^{\varepsilon/2}} \rho_{jn}^2(x) &= \inf_{x \in \mathcal{S}_j^{\varepsilon/2}} \frac{1}{h^d} \mathbf{E}\left[Y_{ji}^2 K^2 \left(\frac{x - X_i}{h} \right) \right] \\ &\geq \inf_{x \in \mathcal{S}_j^{\varepsilon}} \mathbf{E}\left[Y_{ji}^2 | X = x \right] f(x) \int_{[-1/2,1/2]^d} K^2(u)\, du > c_1, \end{aligned}$$

for some $c_1 > 0$ by Assumptions 1(i) and 2. Similarly, from some large n on,

$$\sup_{x \in \mathcal{S}_j^{\varepsilon/2}} k_{jn,r}(x) \leq \sup_{x \in \mathcal{S}_j^{\varepsilon}} \mathbf{E}\left[|Y_{ji}|^r | X = x \right] f(x) \int |K(u)|^r du < \infty,$$

by Assumptions 1(i) and 2. ■

Define for each $j \in \mathcal{J}$,

$$\hat{g}_{jN}(x) \equiv \frac{1}{nh^d} \sum_{i=1}^{N} Y_{ji} K\left(\frac{x - X_i}{h} \right),\ x \in \mathcal{X},$$

where N is a Poisson random variable that is common across $j \in \mathcal{J}$, has mean n, and is independent of $\{(Y_{ji}, X_i) : j \in \mathcal{J}\}_{i=1}^{\infty}$. Let for each $j \in \mathcal{J}$,

$$v_{jn}(x) \equiv \hat{g}_{jn}(x) - \mathbf{E}\hat{g}_{jn}(x), \text{ and } v_{jN}(x) \equiv \hat{g}_{jN}(x) - \mathbf{E}\hat{g}_{jn}(x).$$

We define, for each $j \in \mathcal{J}$,

$$\begin{aligned} \xi_{jn}(x) &\equiv \frac{\sqrt{nh^d}\, v_{jN}(x)}{\rho_{jn}(x)} \text{ and} \\ V_{jn}(x) &\equiv \frac{\sum_{i \leq N_1} \{ Y_{ji} K((x - X_i)/h) - \mathbf{E}(Y_{ji} K((x - X_i)/h)) \}}{\sqrt{\mathbf{E}[Y_{ji}^2 K^2((x - X_i)/h)]}}, \end{aligned} \tag{6.6}$$

where N_1 denotes a Poisson random variable with mean 1 that is independent of $\{(Y_{ji}, X_i) : j \in \mathcal{J}\}_{i=1}^{\infty}$. Then, $Var(V_{jn}(x)) = 1$. Let $V_{jn}^{(i)}(x),\ i = 1, \cdots, n,$ be i.i.d. copies of $V_{jn}(x)$ so that

$$\xi_{jn}(x) \overset{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} V_{jn}^{(i)}(x). \tag{6.7}$$

LEMMA A5: *Suppose that Assumptions 1(i)(iii) and 2 hold and $h \to 0$ as $n \to \infty$ and*

$$\limsup_{n \to \infty} n^{-r/2+1} h^{(1-r/2)d} < C,$$

for some constant $C > 0$ and for $r \in [2, 2p+2]$. Then, for $\varepsilon > 0$ in Assumption 1(iii),

$$\sup_{x \in \mathcal{S}_j^{\varepsilon/2}} \mathbf{E}\left[|V_{jn}(x)|^r\right] \leq C_1 h^{(1-r/2)d} \text{ and } \sup_{x \in \mathcal{S}_j^{\varepsilon/2}} \mathbf{E}\left[|\xi_{jn}(x)|^r\right] \leq C_2,$$

where C_1 and C_2 are constants that depend only on r.

Proof : For all $x \in \mathcal{S}_j^{\varepsilon/2}$, $\mathbf{E}[V_{jn}^2(x)] = 1$. Recall the definition of $k_{jn,r}(x)$ in (6.5). Then for some $C_0, C_1 > 0$,

$$\sup_{x \in \mathcal{S}_j^{\varepsilon/2}} \mathbf{E}\left|V_{jn}(x)\right|^r \leq C_0 \sup_{x \in \mathcal{S}_j^{\varepsilon/2}} \frac{h^d k_{jn,r}(x)}{h^{rd/2}\rho_{jn}^r(x)} \leq C_1 h^{(1-r/2)d}, \tag{6.8}$$

by Lemma A4, completing the proof of the first statement.

As for the second statement, using (6.7) and applying Rosenthal's inequality (e.g. (2.3) of GMZ), we deduce that for positive constants C_3, C_4 and C_5 that depend only on r,

$$\begin{aligned} \sup_{x \in \mathcal{S}_j^{\varepsilon/2}} \mathbf{E}\left[|\xi_{jn}(x)|^r\right] &\leq C_3 \sup_{x \in \mathcal{S}_j^{\varepsilon/2}} \max\{(\mathbf{E}V_{jn}^2(x))^{r/2}, n^{-r/2+1}\mathbf{E}|V_{jn}(x)|^r\} \\ &\leq C_4 \max\left\{1, C_5 n^{-r/2+1} h^{(1-r/2)d}\right\} \end{aligned}$$

by (6.8). By the condition that $\limsup_{n\to\infty} n^{-r/2+1} h^{(1-r/2)d} < C$, the desired result follows. ■

The following lemma is adapted from Lemma 6.3 of GMZ. The result is obtained by combining Lemmas A2-A5.

Lemma A6: *Suppose that Assumptions 1 and 2 hold and $h \to 0$ and $n^{-1/2}h^{-d} \to 0$ as $n \to \infty$. Then for any Borel set $A \subset \mathbf{R}^d$ and for any $j \in \mathcal{J}$,*

$$\begin{aligned} \int_A \left\{n^{p/2} h^{(p-1)d/2} \mathbf{E}\Lambda_p(v_{jN}(x)) - h^{-d/2}\rho_{jn}^p(x)\mathbf{E}\Lambda_p(\mathbb{Z}_1)\right\} w_j(x)dx &\to 0, \\ \int_A \left\{n^{p/2} h^{(p-1)d/2} \mathbf{E}\Lambda_p(v_{jn}(x)) - h^{-d/2}\rho_{jn}^p(x)\mathbf{E}\Lambda_p(\mathbb{Z}_1)\right\} w_j(x)dx &\to 0. \end{aligned}$$

Proof : Recall the definition of $\xi_{jn}(x)$ in (6.6) and write

$$\begin{aligned} & n^{p/2} h^{(p-1)d/2} \mathbf{E}\Lambda_p(v_{jN}(x)) - h^{-d/2}\rho_{jn}^p(x)\mathbf{E}\Lambda_p(\mathbb{Z}_1) \\ = \; & h^{-d/2}\rho_{jn}^p(x)\left\{\mathbf{E}\Lambda_p(\xi_{jn}(x)) - \mathbf{E}\Lambda_p(\mathbb{Z}_1)\right\}. \end{aligned}$$

In view of Lemma A4 and Assumption 1(ii), we find that it suffices for the first statement of the lemma to show that

$$\sup_{x\in\mathcal{S}_j}\left|\mathbf{E}\Lambda_p(\xi_{jn}(x))-\mathbf{E}\Lambda_p(\mathbb{Z}_1)\right|=o(h^{d/2}). \tag{6.9}$$

By Lemma A5, $\sup_{x\in\mathcal{S}_j}\mathbf{E}\left|V_{jn}(x)\right|^3\leq Ch^{-d/2}$ for some $C>0$. Using Lemma A2 and taking $r=\max\{p,3\}$ and $V_{jn}^{(i)}(x)=W_i$, and $\Lambda_p(\cdot)=\varphi(\cdot)$, we deduce that

$$\begin{aligned}&\sup_{x\in\mathcal{S}_j}\left|\mathbf{E}\Lambda_p(\xi_{jn}(x))-\mathbf{E}\Lambda_p(\mathbb{Z}_1)\right|\\ \leq\;& C_1n^{-1/2}\sup_{x\in\mathcal{S}_j}\mathbf{E}\left|V_{jn}(x)\right|^3+C_2n^{-(r-2)/2}\sup_{x\in\mathcal{S}_j}\mathbf{E}\left|V_{jn}(x)\right|^r\\ &+C_3\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[\omega_{\Lambda_p}\left(\mathbb{Z}_1;\frac{C_4}{\sqrt{n}}\mathbf{E}\left|V_{jn}(x)\right|^3\right)\right],\end{aligned} \tag{6.10}$$

for some constants $C_s>0$, $s=1,2,3$. The first two terms are $o(h^{d/2})$. As for the last expectation, observe that by Lemma A3,

$$\mathbf{E}\left[\omega_{\Lambda_p}\left(\mathbb{Z}_1;\frac{C_4}{\sqrt{n}}\mathbf{E}\left|V_{jn}(x)\right|^3\right)\right]\leq C\sum_{k=0}^{\lceil p-1\rceil}\left(\frac{C_4}{\sqrt{n}}\mathbf{E}\left|V_{jn}(x)\right|^3\right)^{p-\frac{(p-1)k}{\lceil p-1\rceil}}\mathbf{E}|\mathbb{Z}_1|^k.$$

The last sum is $O(n^{-1/2}h^{-d/2})=o(h^{d/2})$ uniformly over $x\in\mathcal{S}_j$, completing the proof of (6.9).

We consider the second statement. Let $\bar{V}_{jn}^{(k)}(x),\ k=1,\cdots,n$, be i.i.d. copies of

$$\frac{Y_{ji}K\left(\frac{x-X_i}{h}\right)-\mathbf{E}\left(Y_{ji}K\left(\frac{x-X_i}{h}\right)\right)}{\sqrt{\mathbf{E}\left[Y_{ji}^2K^2\left(\frac{x-X_i}{h}\right)\right]-\left(\mathbf{E}\left[Y_{ji}K\left(\frac{x-X_i}{h}\right)\right]\right)^2}}$$

so that $Var(\bar{V}_{jn}^{(k)}(x))=1$. Observe that for some constants $C_1,C_2>0$,

$$\sup_{x\in\mathcal{S}_j}\mathbf{E}\left|\bar{V}_{jn}^{(k)}(x)\right|^3\leq Ch^{-d/2}\sup_{x\in\mathcal{S}_j}\frac{k_{jn,3}(x)}{\left(\rho_{jn}^2(x)-h^db_{jn}^2(x)\right)^{3/2}}\leq C_2h^{-d/2}, \tag{6.11}$$

where $b_{jn}(x)\equiv h^{-d}\mathbf{E}\left[Y_{ji}K\left((x-X_i)/h\right)\right]$. The last inequality follows by Lemma A4. Define

$$\bar{\xi}_{jn}(x)\equiv\frac{\sqrt{nh^d}v_{jn}(x)}{\tilde{\rho}_{jn}(x)},$$

where $\tilde{\rho}_{jn}^2(x)\equiv nh^dVar(v_{jn}(x))$. Then $\bar{\xi}_{jn}(x)\overset{d}{=}\frac{1}{\sqrt{n}}\sum_{k=1}^n\bar{V}_{jn}^{(k)}(x)$. Using Lemma A2 and following the arguments in (6.10) analogously, we deduce that

$$\sup_{x\in\mathcal{S}_j}\left|\mathbf{E}\Lambda_p\left(\bar{\xi}_{jn}(x)\right)-\mathbf{E}\Lambda_p(\mathbb{Z}_1)\right|=o(h^{d/2}).$$

This leads us to conclude that

$$\int_A \left\{ n^{p/2} h^{(p-1)d/2} \mathbf{E}\Lambda_p(v_{jn}(x)) - h^{-d/2} \tilde{\rho}_{jn}^p(x) \mathbf{E}\Lambda_p(\mathbb{Z}_1) \right\} w_j(x) dx = o(1).$$

Now, there exists n_0 such that for all $n > n_0$, $\sup_{x \in \mathcal{S}_j} h^d b_{jn}^2(x) < c_1/2$, where $c_1 > 0$ is the constant in Lemma A3. Observe that for all $n > n_0$,

$$\begin{aligned} & \sup_{x \in \mathcal{S}_j} h^{-d/2} \left| \tilde{\rho}_{jn}^p(x) - \rho_{jn}^p(x) \right| \\ = & \sup_{x \in \mathcal{S}_j} h^{-d/2} \left| (\rho_{jn}^2(x) - h^d b_{jn}^2(x))^{p/2} - (\rho_{jn}^2(x))^{p/2} \right| \\ \leq & \sup_{x \in \mathcal{S}_j} \frac{p}{2} h^{d/2} b_{jn}^2(x) \cdot \max \left\{ \left(\rho_{jn}^2(x) + c_1/2 \right)^{p/2-1}, \left(\rho_{jn}^2(x) - c_1/2 \right)^{p/2-1} \right\}. \end{aligned}$$

By Lemma A4, the last term is $O(h^{d/2}) = o(1)$. This completes the proof. ■

Recall the definition: $\rho_j^2(x) \equiv \mathbf{E}[Y_{ji}^2 | X_i = x] f(x) \int K^2(u) du$. Let

$$\begin{aligned} \sigma_{jk,n}(A) &\equiv n^p h^{(p-1)d} \int_A \int_A Cov\left(\Lambda_p(v_{jN}(x)), \Lambda_p(v_{kN}(z))\right) w_j(x) w_k(z) dx dz, \text{ and} \\ \sigma_{jk}(A) &\equiv \int_A q_{jk,p}(x) \rho_j^p(x) \rho_k^p(x) w_j(x) w_k(x) dx, \end{aligned} \tag{6.12}$$

where we recall the definition:

$$q_{jk,p}(x) \equiv \int_{[-1,1]^d} Cov\left(\Lambda_p(\sqrt{1 - t_{jk}^2(x,u)} \mathbb{Z}_1 + t_{jk}(x,u) \mathbb{Z}_2), \Lambda_p(\mathbb{Z}_2) \right) du.$$

Now, let $(Z_{1n}(x), Z_{2n}(z)) \in \mathbf{R}^2$ be a jointly normal centered random vector whose covariance matrix is the same as that of $(\xi_{jn}(x), \xi_{kn}(z))$ for all $x, z \in \mathbf{R}^d$. We define

$$\tau_{jk,n}(A) \equiv \int_A \int_{[-1,1]^d} g_{jk,n}(x,u) \lambda_{jk,n}(x, x + uh) du dx,$$

where

$$\begin{aligned} \lambda_{jk,n}(x,z) &\equiv \rho_{jn}^p(x) \rho_{kn}^p(z) w_j(x) w_k(z) 1_A(x) 1_A(z), \text{ and} \\ g_{jk,n}(x,u) &\equiv Cov\left(\Lambda_p(Z_{1n}(x)), \Lambda_p(Z_{2n}(x + uh))\right). \end{aligned}$$

The following result generalizes Lemma 6.5 of GMZ from a univariate X to a multivariate X. The truncation arguments in their proof on pages 752 and 753 do not apply in the case of multivariate X. The proof of the following lemma employs a different approach for this part.

Lemma A7: *Suppose that Assumptions 1 and 2 hold and let $h \to 0$ as $n \to \infty$ satisfying* $\limsup_{n\to\infty} n^{-r/2+1} h^{(1-r/2)d} < C$ *for any $r \in [2, 2p+2]$ for some $C > 0$.*

(i) *Suppose that* $A \subset \mathcal{S}_j \cap \mathcal{S}_k$ *is any Borel set. Then*

$$\sigma_{jk,n}(A) = \tau_{jk,n}(A) + o(1).$$

(ii) *Suppose further that* A *has a finite Lebesgue measure,* $\rho_j(\cdot)\rho_k(\cdot)$ *and* $w_j(\cdot)w_k(\cdot)$ *are continuous and bounded on* A, *and*

$$\sup_{x\in A} |\rho_{l,n}(x) - \rho_l(x)| \to 0, \text{ as } n \to \infty, \text{ for } l \in \{j,k\}. \tag{6.13}$$

Then, as $n \to \infty$, $\tau_{jk,n}(A) = \sigma_{jk}(A) + o(1)$, *and hence from* (i),

$$\sigma_{jk,n}(A) \to \sigma_{jk}(A).$$

Proof: (i) By change of variables, we write $\sigma_{jk,n}(A) = \tilde{\tau}_{jk,n}(A)$, where

$$\tilde{\tau}_{jk,n}(A) \equiv \int_A \int_{[-1,1]^d} Cov\left(\Lambda_p(\xi_{jn}(x)), \Lambda_p(\xi_{kn}(x+uh))\right) \lambda_{jk,n}(x, x+uh) du dx.$$

Fix $\varepsilon_1 \in (0,1]$ and let $c(\varepsilon_1) = (1+\varepsilon_1)^2 - 1$. Let η_1 and η_2 be two independent random variables that are independent of $(\{Y_{ji}, X_i : j \in \mathcal{J}\}_{i=1}^{\infty}, N)$, each having a two-point distribution that gives two points, $\{\sqrt{c(\varepsilon_1)}\}$ and $\{-\sqrt{c(\varepsilon_1)}\}$, the equal mass of 1/2, so that $\mathbf{E}\eta_1 = \mathbf{E}\eta_2 = 0$ and $Var(\eta_1) = Var(\eta_2) = c(\varepsilon_1)$. Furthermore, observe that for any $r \geq 1$,

$$\mathbf{E}|\eta_1|^r = \frac{1}{2}|c(\varepsilon_1)|^{r/2} + \frac{1}{2}|c(\varepsilon_1)|^{r/2} \leq C\varepsilon_1^{r/2}, \tag{6.14}$$

for some constant $C > 0$ that depends only on r. Define

$$\xi_{jn,1}^{\eta}(x) \equiv \frac{\xi_{jn}(x) + \eta_1}{1+\varepsilon_1} \text{ and } \xi_{kn,2}^{\eta}(x+uh) \equiv \frac{\xi_{kn}(x+uh) + \eta_2}{1+\varepsilon_1}.$$

Note that $Var(\xi_{jn,1}^{\eta}(x)) = Var(\xi_{kn,2}^{\eta}(x+uh)) = 1$. Let $(Z_{1n}^{\eta}(x), Z_{2n}^{\eta}(x+uh))$ be a jointly normal centered random vector whose covariance matrix is the same as that of $(\xi_{jn,1}^{\eta}(x), \xi_{kn,2}^{\eta}(x+uh))$ for all $(x,u) \in \mathbf{R}^d \times [-1,1]^d$. Define

$$\begin{aligned}
\tilde{\tau}_{jk,n}^{\eta}(A) &\equiv \int_A \int_{[-1,1]^d} Cov\left(\Lambda_p(\xi_{jn,1}^{\eta}(x)), \Lambda_p(\xi_{kn,2}^{\eta}(x+uh))\right) \lambda_{jk,n}(x, x+uh) du dx, \\
\tau_{jk,n}^{\eta}(A) &\equiv \int_A \int_{[-1,1]^d} Cov\left(\Lambda_p(Z_{1n}^{\eta}(x)), \Lambda_p(Z_{2n}^{\eta}(x+uh))\right) \lambda_{jk,n}(x, x+uh) du dx.
\end{aligned}$$

Then first observe that

$$\begin{aligned}
\left|\tilde{\tau}_{jk,n}(A) - \tilde{\tau}_{jk,n}^{\eta}(A)\right| &\leq \int_A \int_{[-1,1]^d} |\Delta_{jk,n,1}^{\eta}(x,u)| \lambda_{jk,n}(x, x+uh) du dx \\
&\quad + \int_A \int_{[-1,1]^d} |\Delta_{jk,n,2}^{\eta}(x,u)| \lambda_{jk,n}(x, x+uh) du dx,
\end{aligned}$$

where

$$\begin{aligned}\Delta^{\eta}_{jk,n,1}(x,u) &\equiv \mathbf{E}\Lambda_p(\xi_{jn}(x))\mathbf{E}\Lambda_p(\xi_{kn}(x+uh))\\ &\quad -\mathbf{E}\Lambda_p(\xi^{\eta}_{jn,1}(x))\mathbf{E}\Lambda_p(\xi^{\eta}_{kn,2}(x+uh)) \text{ and}\\ \Delta^{\eta}_{jk,n,2}(x,u) &\equiv \mathbf{E}\Lambda_p(\xi_{jn}(x))\Lambda_p(\xi_{kn}(x+uh))\\ &\quad -\mathbf{E}\Lambda_p(\xi^{\eta}_{jn,1}(x))\Lambda_p(\xi^{\eta}_{kn,2}(x+uh)).\end{aligned}$$

Since for any $a,b \in \mathbf{R}$, $|a_+^p - b_+^p| \le p|a-b|\left(|a|^{p-1}+|b|^{p-1}\right)$, we bound $|\Delta^{\eta}_{jk,n,2}(x,u)|$ by

$$\begin{aligned}&p\mathbf{E}\left[|\xi_{jn}(x)-\xi^{\eta}_{jn,1}(x)|\left(|\xi_{jn}(x)|^{p-1}+|\xi^{\eta}_{jn,1}(x)|^{p-1}\right)|\xi_{kn}(x+uh)|^p\right]\\ &+p\mathbf{E}\left[|\xi_{kn}(x+uh)-\xi^{\eta}_{kn,2}(x+uh)|\left(|\xi_{kn}(x+uh)|^{p-1}+|\xi^{\eta}_{kn,2}(x+uh)|^{p-1}\right)|\xi^{\eta}_{jn,1}(x)|^p\right]\\ \equiv\ & A_{1n}(x,u)+A_{2n}(x,u), \text{ say.}\end{aligned}$$

As for $A_{1n}(x,u)$,

$$\begin{aligned}A_{1n}(x,u) &\le p\left(\mathbf{E}\left[|\xi_{jn}(x)-\xi^{\eta}_{jn,1}(x)|^2(|\xi_{jn}(x)|^{p-1}+|\xi^{\eta}_{jn,1}(x)|^{p-1})^2\right]\right)^{1/2}\\ &\quad\times\left(\mathbf{E}\left[|\xi_{kn}(x+uh)|^{2p}\right]\right)^{1/2}.\end{aligned}$$

Define

$$s \equiv \begin{cases}(p+1)/(p-1) \text{ if } p>1\\ 2 \text{ if } p=1,\end{cases}$$

and $q \equiv (1-1/s)^{-1}$. Note that by Hölder inequality,

$$\begin{aligned}&\mathbf{E}\left[|\xi_{jn}(x)-\xi^{\eta}_{jn,1}(x)|^2(|\xi_{jn}(x)|^{p-1}+|\xi^{\eta}_{jn,1}(x)|^{p-1})^2\right]\\ \le\ &\left(\mathbf{E}\left[|\xi_{jn}(x)-\xi^{\eta}_{jn,1}(x)|^{2q}\right]\right)^{1/q}\left(\mathbf{E}\left[(|\xi_{jn}(x)|^{p-1}+|\xi^{\eta}_{jn,1}(x)|^{p-1})^{2s}\right]\right)^{1/s}.\end{aligned}$$

Now,

$$\begin{aligned}&\mathbf{E}\left[|\xi_{jn}(x)-\xi^{\eta}_{jn,1}(x)|^{2q}\right]=(1+\varepsilon_1)^{-2q}\mathbf{E}\left[|\varepsilon_1\xi_{jn}(x)-\eta_1|^{2q}\right]\\ \le\ & 2^{2q-1}(1+\varepsilon_1)^{-2q}\left\{\varepsilon_1^{2q}\mathbf{E}\left[|\xi_{jn}(x)|^{2q}\right]+\mathbf{E}\left[|\eta_1|^{2q}\right]\right\}.\end{aligned}$$

Applying Lemma A5 and (6.14) to the last bound, we conclude that

$$\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[|\xi_{jn}(x)-\xi^{\eta}_{jn,1}(x)|^{2q}\right]\le\frac{C_1(\varepsilon_1^{2q}+\varepsilon_1^{q})}{(1+\varepsilon_1)^{2q}}\le C_2\varepsilon_1^q,$$

for some constants $C_1, C_2 > 0$. Using Lemma A5, we can also see that for some constants $C_3, C_4 > 0$,

$$\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[(|\xi_{jn}(x)|^{p-1}+|\xi^{\eta}_{jn,1}(x)|^{p-1})^{2s}\right]\le C_3$$

and from some large n on,

$$\sup_{u\in[-1,1]^d}\sup_{x\in\mathcal{S}_k}\mathbf{E}\left[|\xi_{kn}(x+uh)|^{2p}\right]\leq\sup_{x\in\mathcal{S}_k^{\varepsilon/2}}\mathbf{E}\left[|\xi_{kn}(x)|^{2p}\right]\leq C_4,$$

for $\varepsilon>0$ in Assumption 1(iii). Therefore, for some constant $C>0$,

$$\sup_{u\in[-1,1]^d}\sup_{x\in\mathcal{S}_j\cap\mathcal{S}_k}A_{1n}(x,u)\leq C\sqrt{\varepsilon_1}.$$

Using similar arguments for $A_{2n}(x,u)$, we deduce that for some constant $C>0$,

$$\sup_{u\in[-1,1]^d}\sup_{x\in\mathcal{S}_j\cap\mathcal{S}_k}|\Delta_{jk,n,2}^{\eta}(x,u)|\leq C\sqrt{\varepsilon_1}. \tag{6.15}$$

Let us turn to $\Delta_{jk,n,1}^{\eta}(x,u)$. We bound $|\Delta_{jk,n,1}^{\eta}(x,u)|$ by

$$\begin{aligned}&p\mathbf{E}\left[|\xi_{jn}(x)-\xi_{jn,1}^{\eta}(x)|\left(|\xi_{jn}(x)|^{p-1}+|\xi_{jn,1}^{\eta}(x)|^{p-1}\right)\right]\mathbf{E}\left[|\xi_{kn}(x+uh)|^{p}\right]\\&+p\mathbf{E}\left[|\xi_{kn}(x+uh)-\xi_{kn,2}^{\eta}(x+uh)|\left(|\xi_{kn}(x+uh)|^{p-1}+|\xi_{kn,1}^{\eta}(x+uh)|^{p-1}\right)\right]\mathbf{E}[|\xi_{jn,1}^{\eta}(x)|^{p}].\end{aligned}$$

Using similar arguments for $\Delta_{jk,n,2}^{\eta}(x,u)$, we find that for some constant $C>0$,

$$\sup_{u\in[-1,1]^d}\sup_{x\in\mathcal{S}_j\cap\mathcal{S}_k}|\Delta_{jk,n,1}^{\eta}(x,u)|\leq C\sqrt{\varepsilon_1}. \tag{6.16}$$

By Lemma A4 and Assumption 1(ii), there exist $n_0>0$ and $C_1,C_2>0$ such that for all $n\geq n_0$,

$$\begin{aligned}&\int_A\int_{[-1,1]^d}\lambda_{jk,n}(x,x+uh)dudx\\\leq\ &C_1\int_A\int_{[-1,1]^d}w_j(x)w_k(x+uh)dudx\\\leq\ &C_2\sqrt{\int_A w_j^2(x)dx}\sqrt{\int_A\int_{[-1,1]^d}w_k^2(x+uh)dudx}<\infty.\end{aligned} \tag{6.17}$$

Hence

$$\left|\tilde{\tau}_{jk,n}(A)-\tilde{\tau}_{jk,n}^{\eta}(A)\right|\leq C_5\sqrt{\varepsilon_1}\int_A\int_{[-1,1]^d}\lambda_{jk,n}(x,x+uh)dudx\leq C_6\sqrt{\varepsilon_1},$$

for some constants $C_5>0$ and $C_6>0$.

Since the choice of $\varepsilon_1>0$ was arbitrary, it remains for the proof of Lemma A7(i) to prove that

$$\left|\tilde{\tau}_{jk,n}^{\eta}(A)-\tau_{jk,n}(A)\right|=o(1), \tag{6.18}$$

as $n \to \infty$ and then $\varepsilon_1 \to 0$. For any $x \in \mathcal{S}_j \cap \mathcal{S}_k$,

$$\left(\xi^{\eta}_{jn,1}(x), \xi^{\eta}_{kn,2}(x+uh)\right)' \overset{d}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_n^{(i)}(x,u),$$

where $W_n^{(i)}(x,u)$'s are i.i.d. copies of $W_n(x,u) \equiv (q_{jn}(x), q_{kn}(x+uh))'$ with

$$q_{jn}(x) \equiv \left\{ \frac{\sum_{i \le N_1} Y_{ji} K\left((x-X_i)/h\right) - \mathbf{E}\left[Y_{ji} K\left((x-X_i)/h\right)\right]}{h^{d/2} \rho_{jn}(x)} + \eta_1 \right\} / (1+\varepsilon_1).$$

Using the same arguments as in the proof of Lemma A5, we find that for $j \in \mathcal{J}$,

$$\sup_{x \in \mathcal{S}_j^{\varepsilon}} \mathbf{E}\left[|q_{jn}(x)|^3\right] \le C h^{-d/2}, \text{ for some } C > 0. \tag{6.19}$$

Let Σ_{1n} be the 2×2 covariance matrix of $(\xi^{\eta}_{jn,1}(x), \xi^{\eta}_{kn,2}(x+uh))'$. Define

$$\tilde{\Lambda}_{n,p}(v) \equiv \Lambda_p([\Sigma_{1n}^{1/2} v]_1) \Lambda_p([\Sigma_{1n}^{1/2} v]_2),\ v \in \mathbf{R}^2,$$

where $[a]_j$ of a vector $a \in \mathbf{R}^2$ indicates its j-th entry. There exists some $C > 0$ such that for all n,

$$\begin{aligned} \sup_{v \in \mathbf{R}^2} \frac{\left|\tilde{\Lambda}_{n,p}(v) - \tilde{\Lambda}_{n,p}(0)\right|}{1 + ||v||^{2p+2} \min\{||v||, 1\}} &\le C \text{ and} \\ \int \sup_{u \in \mathbf{R}^2 : ||z-u|| \le \delta} \left|\tilde{\Lambda}_{n,p}(z) - \tilde{\Lambda}_{n,p}(u)\right| d\Phi(z) &\le C\delta \text{ for all } \delta \in (0,1]. \end{aligned} \tag{6.20}$$

The correlation between $\xi^{\eta}_{jn,1}(x)$ and $\xi^{\eta}_{kn,2}(x+uh)$ is equal to

$$\mathbf{E}\left[\xi^{\eta}_{jn,1}(x) \xi^{\eta}_{kn,2}(x+uh)\right] = \frac{\mathbf{E}[\xi_{jn}(x) \xi_{kn}(x+uh)]}{(1+\varepsilon_1)^2} \in \left[-(1+\varepsilon_1)^{-2}, (1+\varepsilon_1)^{-2}\right].$$

Hence, as for $\tilde{W}_n^{(i)}(x,u) \equiv \Sigma_{1n}^{-1/2} W_n^{(i)}(x,u)$, by (6.19),

$$\begin{aligned} (6.21)\ \sup_{x \in \mathcal{S}_j \cap \mathcal{S}_k} & \mathbf{E}||\tilde{W}_n^{(i)}(x,u)||^3 \\ \le\ & C_1 (1 - (\mathbf{E}[\xi^{\eta}_{jn,1}(x) \xi^{\eta}_{kn,2}(x+uh)])^2)^{-3/2} \left\{ \sup\nolimits_{x \in \mathcal{S}_j^{\varepsilon}} \mathbf{E}[|q_{jn}(x)|^3] + \sup\nolimits_{x \in \mathcal{S}_k^{\varepsilon}} \mathbf{E}[|q_{kn}(x)|^3] \right\} \\ \le\ & C_1 (1 - (1+\varepsilon_1)^{-4})^{-3/2} \left\{ \sup\nolimits_{x \in \mathcal{S}_j^{\varepsilon}} \mathbf{E}[|q_{jn}(x)|^3] + \sup\nolimits_{x \in \mathcal{S}_k^{\varepsilon}} \mathbf{E}[|q_{kn}(x)|^3] \right\} \\ \le\ & C_2 (1 - (1+\varepsilon_1)^{-4})^{-3/2} h^{-d/2}, \text{ for some } C_1, C_2 > 0, \end{aligned}$$

so that $n^{-1/2}\sup_{x\in\mathcal{S}_j\cap\mathcal{S}_k}\mathbf{E}||\tilde{W}_n^{(i)}(x,u)||^3 = O(n^{-1/2}h^{-d/2})$. By Lemma A2 and following the arguments in (6.10) analogously,

$$\begin{aligned}&\sup_{x\in\mathcal{S}_j\cap\mathcal{S}_k}\left|\mathbf{E}\tilde{\Lambda}_{n,p}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{W}_n^{(i)}(x,u)\right)-\mathbf{E}\tilde{\Lambda}_{n,p}\left(\tilde{Z}_n^{\eta}(x,u)\right)\right|\\ = & \; O\left(n^{-1/2}h^{-d/2}\right)=o(1),\end{aligned}$$

where $\tilde{Z}_n^{\eta}(x,u)\equiv\Sigma_{1n}^{-1/2}(Z_{1n}^{\eta}(x),Z_{2n}^{\eta}(x+uh))'$. Certainly by (6.14) and Lemma A5,

$$\begin{aligned}&Cov(\Lambda_p(Z_{1n}^{\eta}(x)),\Lambda_p(Z_{2n}^{\eta}(x+uh)))\\ \leq & \; \sqrt{\mathbf{E}|Z_{1n}^{\eta}(x)|^{2p}}\sqrt{\mathbf{E}|Z_{2n}^{\eta}(x+uh)|^{2p}}<C,\end{aligned}$$

for some $C>0$ that does not depend on ε_1. Using (6.17), we apply the dominated convergence theorem to obtain that

$$\left|\tau_{jk,n}^{\eta}(A)-\tilde{\tau}_{jk,n}^{\eta}(A)\right|=o(1) \tag{6.22}$$

as $n\to\infty$ for each $\varepsilon_1>0$.

Finally, note from (6.15) and (6.16) that, for all $x\in A$ and all $u\in[-1,1]^d$,

$$\begin{aligned}&Cov(\Lambda_p(Z_{1n}^{\eta}(x)),\Lambda_p(Z_{2n}^{\eta}(x+uh)))\\ = & \; Cov(\Lambda_p(Z_{1n}(x)),\Lambda_p(Z_{2n}(x+uh)))+o(1),\end{aligned}$$

where the $o(1)$ term is one that converges to zero as $n\to\infty$ and then $\varepsilon_1\to 0$. Therefore, by the dominated convergence theorem,

$$\left|\tau_{jk,n}^{\eta}(A)-\tau_{jk,n}(A)\right|=o(1),$$

as $n\to\infty$ and then $\varepsilon_1\to 0$. In view of (6.22), this completes the proof of (6.18) and, as a consequence, that of (i).

(ii) Define $t_{jk,n}(x,u)\equiv\mathbf{E}(\xi_{jn}(x)\xi_{kn}(x+uh))$,

$$\begin{aligned}e_{jk,n}(x,u) &\equiv \frac{1}{h^d}\mathbf{E}\left[Y_{ji}Y_{ki}K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_i}{h}+u\right)\right] \text{ and}\\ e_{jk}(x,u) &\equiv \rho_{jk}(x)\frac{\int K(z)K(z+u)dz}{\int K^2(u)du}.\end{aligned}$$

By Assumption 1(i), and Lemma A4, for almost every $x\in A$ and for each $u\in[-1,1]^d$,

$$\begin{aligned}t_{jk,n}(x,u) &= \frac{1}{\rho_{jn}(x)\rho_{kn}(x+uh)}\frac{1}{h^d}\mathbf{E}\left[Y_{ji}Y_{ki}K\left(\frac{x-X_i}{h}\right)K\left(\frac{x-X_i}{h}+u\right)\right]\\ &= \frac{e_{jk,n}(x,u)}{\rho_{jn}(x)\rho_{kn}(x+uh)}=\frac{e_{jk}(x,u)}{\rho_j(x)\rho_k(x+uh)}+o(1)=t_{jk}(x,u)+o(1),\end{aligned} \tag{6.23}$$

where we recall that $t_{jk}(x,u) = e_{jk}(x,u)/(\rho_j(x)\rho_k(x))$ by the definition of $t_{jk}(\cdot,\cdot)$.

By (6.13),

$$\tau_{jk,n}(A) = \int_A \int_{[-1,1]^d} g_{jk,n}(x,u)\lambda_{jk}(x,x+uh)dudx + o(1),$$

where $\lambda_{jk}(x,z) \equiv \rho_j^p(x)\rho_k^p(z)w_j(x)w_k(z)1_A(x)1_A(z)$. By (6.23), for almost every $x \in A$ and for each $u \in [-1,1]^d$,

$$g_{jk,n}(x,u) \to g_{jk}(x,u), \text{ as } n \to \infty,$$

where $g_{jk}(x,u) \equiv Cov(\Lambda_p(\sqrt{1-t_{jk}^2(x,u)}\mathbb{Z}_1 + t_{jk}(x,u)\mathbb{Z}_2), \Lambda_p(\mathbb{Z}_2))$. Furthermore, since $\rho_j(\cdot)\rho_k(\cdot)$ and $w_j(\cdot)w_k(\cdot)$ are continuous on A and A has a finite Lebesgue measure, we follow the proof of Lemma 6.4 of GMZ to find that $g_{jk,n}(x,u)\lambda_{jk}(x,x+uh)$ converges in measure to $g_{jk}(x,u)\lambda_{jk}(x,x)$ on $A\times[-1,1]^d$, as $n \to \infty$. Using the bounded convergence theorem, we deduce the desired result. ■

The following lemma is a generalization of Lemma 6.2 of GMZ from $p = 1$ to $p \geq 1$. The proof of GMZ does not carry over to this general case because the majorization inequality of Pinelis (1994) used in GMZ does not apply here. (Note that (4) in Pinelis (1994) does not apply when $p > 1$.)

Lemma A8: *Suppose that Assumptions 1 and 2 hold. Furthermore, assume that as* $n \to \infty$, $h \to 0$, $n^{-1/2}h^{-d} \to 0$. *Then there exists a constant* $C > 0$ *such that for any Borel set* $A \subset \mathbf{R}^d$ *and for all* $j \in \mathcal{J}$,

$$\text{limsup}_{n\to\infty}\mathbf{E}\left[\left|n^{p/2}h^{(p-1)d/2}\int_A \{\Lambda_p(v_{jn}(x)) - \mathbf{E}[\Lambda_p(v_{jn}(x))]\} w_j(x)dx\right|\right]$$
$$\leq \ C\int_A w_j(x)dx + C\sqrt{\int_A w_j^2(x)dx}.$$

Proof : It suffices to show that there exists $C > 0$ such that for any Borel set $A \subset \mathbf{R}^d$,

Step 1: $\mathbf{E}\left[\left|n^{p/2}h^{(p-1)d/2}\int_A (\Lambda_p(v_{jn}(x)) - \Lambda_p(v_{jN}(x)))\, w_j(x)dx\right|\right] \leq C\int_A w_j(x)dx$,

Step 2: $\mathbf{E}\left[\left|n^{p/2}h^{(p-1)d/2}\int_A (\Lambda_p(v_{jN}(x)) - \mathbf{E}[\Lambda_p(v_{jN}(x))])\, w_j(x)dx\right|\right] \leq C\sqrt{\int_A w_j^2(x)dx}$, and

Step 3: $n^{p/2}h^{(p-1)d/2}\left|\int_A (\mathbf{E}\Lambda_p(v_{jN}(x)) - \mathbf{E}[\Lambda_p(v_{jn}(x))])\, w_j(x)dx\right| \to 0$ as $n \to \infty$.

Indeed, by chaining Steps 1, 2 and 3, we obtain the desired result.

PROOF OF STEP 1: For simplicity, let

$$\begin{aligned} u_{jn}^2(x) &\equiv \mathbf{E}\left[Y_{ji}^2 K^2\left(\frac{x-X_i}{h}\right)\right] - \left(\mathbf{E}\left[Y_{ji}K\left(\frac{x-X_i}{h}\right)\right]\right)^2 \text{ and} \\ \bar{V}_{n,ji}(x) &\equiv \frac{1}{u_{jn}(x)}\left\{Y_{ji}K\left(\frac{x-X_i}{h}\right) - \mathbf{E}\left[Y_{ji}K\left(\frac{x-X_i}{h}\right)\right]\right\}. \end{aligned}$$

We write, if $N = n$, $\sum_{i=N+1}^{n} = 0$, and if $N > n$, $\sum_{i=N+1}^{n} = -\sum_{i=n+1}^{N}$. Using this notation, write

$$v_{jn}(x) = \frac{1}{nh^d}\sum_{i=1}^{N}\bar{V}_{n,ji}(x)u_{jn}(x) + \frac{1}{nh^d}\sum_{i=N+1}^{n}\bar{V}_{n,ji}(x)u_{jn}(x).$$

Now, observe that

$$\begin{aligned} \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{N}\bar{V}_{n,ji}(x)u_{jn}(x) &= \frac{1}{\sqrt{nh^d}}\sum_{i=1}^{N}\left\{Y_{ji}K\left(\frac{x-X_i}{h}\right) - \mathbf{E}\left[Y_{ji}K\left(\frac{x-X_i}{h}\right)\right]\right\} \\ &= \sqrt{nh^d}\left\{\hat{g}_{jN}(x) - \frac{1}{h^d}\mathbf{E}\left[Y_{ji}K\left(\frac{x-X_i}{h}\right)\right]\right\} \\ &\quad + \sqrt{nh^d}\left(\frac{n-N}{n}\right)\cdot\frac{1}{h^d}\cdot\mathbf{E}\left[Y_{ji}K\left(\frac{x-X_i}{h}\right)\right] \\ &= \sqrt{nh^d}v_{jN}(x) + \sqrt{nh^d}\left(\frac{n-N}{n}\right)\cdot\frac{1}{h^d}\cdot\mathbf{E}\left[Y_{ji}K\left(\frac{x-X_i}{h}\right)\right]. \end{aligned}$$

Letting

$$\begin{aligned} \eta_{jn}(x) &\equiv \sqrt{n}\left(\frac{n-N}{n}\right)\cdot\frac{1}{h^d}\cdot\mathbf{E}\left[Y_{ji}K\left(\frac{x-X_i}{h}\right)\right] \text{ and} \\ s_{jn}(x) &\equiv \frac{1}{\sqrt{nh^d}}\sum_{i=N+1}^{n}\bar{V}_{n,ji}(x)u_{jn}(x), \end{aligned}$$

we can write

$$\sqrt{nh^d}v_{jn}(x) = \sqrt{nh^d}v_{jN}(x) + (\sqrt{h^d}\eta_{jn}(x) + s_{jn}(x)). \tag{6.24}$$

First, note that for some constant $C > 0$,

$$\sup_{x\in\mathcal{S}_j} u_{jn}^2(x) \le Ch^d, \tag{6.25}$$

from some large n on, by Lemma A4. Recall the definition of $\tilde{\rho}_{jn}(x)$: $\tilde{\rho}_{jn}(x) \equiv \sqrt{nh^d Var(v_{jn}(x))}$ and note that

$$\tilde{\rho}_{jn}^2(x) = \rho_{jn}^2(x) - h^d b_{jn}^2(x) = h^{-d}u_{jn}^2(x).$$

As in the proof of Lemma A5, there exist n_0, $C_1 > 0$ and $C_2 > 0$ such that for all $n \geq n_0$,

$$\begin{aligned} C_1 &> \sup_{x\in\mathcal{S}_j} \sqrt{\rho_{jn}^2(x) - h^d b_{jn}^2(x)} = \sup_{x\in\mathcal{S}_j} \tilde{\rho}_{jn}(x) \\ &\geq \inf_{x\in\mathcal{S}_j} \tilde{\rho}_{jn}(x) \geq \inf_{x\in\mathcal{S}_j} \sqrt{\rho_{jn}^2(x) - h^d b_{jn}^2(x)} > C_2. \end{aligned} \tag{6.26}$$

Using (6.25), (6.24), and (6.26), we deduce that for some C_1, C_2, C_3, and $C_4 > 0$,

$$\begin{aligned} &\left| n^{p/2} h^{(p-1)d/2} \int_A \left(\Lambda_p(v_{jn}(x)) - \Lambda_p(v_{jN}(x))\right) w_j(x)dx \right| \\ \leq\ & C_1 h^{-d/2} \left| \int_A \left(\Lambda_p\left(\sqrt{nh^d} \frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)} \right) - \Lambda_p\left(\sqrt{nh^d} \frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)} \right) \right) w_j(x)dx \right| \\ \leq\ & C_2 \int_A \left| \frac{\eta_{jn}(x)}{\tilde{\rho}_{jn}(x)} \right| \left(\left| \sqrt{nh^d} \frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)} \right|^{p-1} + \left| \sqrt{nh^d} \frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)} \right|^{p-1} \right) w_j(x)dx \\ & + C_3 h^{-d/2} \int_A \left| \frac{s_{jn}(x)}{\tilde{\rho}_{jn}(x)} \right| \left(\left| \sqrt{nh^d} \frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)} \right|^{p-1} + \left| \sqrt{nh^d} \frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)} \right|^{p-1} \right) w_j(x)dx \\ \leq\ & C_4 \int_A |\eta_{jn}(x)| \left(\left| \sqrt{nh^d} \frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)} \right|^{p-1} + \left| \sqrt{nh^d} \frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)} \right|^{p-1} \right) w_j(x)dx \\ & + C_3 \int_A \left| \frac{s_{jn}(x)}{u_{jn}(x)} \right| \left(\left| \sqrt{nh^d} \frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)} \right|^{p-1} + \left| \sqrt{nh^d} \frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)} \right|^{p-1} \right) w_j(x)dx \\ =\ & A_{1n} + A_{2n}, \text{ say.} \end{aligned}$$

To deal with A_{1n} and A_{2n}, we first show the following:

Claim 1: $\sup_{x\in\mathcal{S}_j} \mathbf{E}[\eta_{jn}^2(x)] = O(1)$.

Claim 2: $\sup_{x\in\mathcal{S}_j} \mathbf{E}[|s_{jn}(x)/u_{jn}(x)|^2] = o(1)$.

Claim 3: $\sup_{x\in\mathcal{S}_j} \mathbf{E}[|\sqrt{nh^d} v_{jN}(x)/\tilde{\rho}_{jn}(x)|^{2p-2}] = O(1)$.

Proof of Claim 1: By Lemma A4 and the fact that $\mathbf{E}|n^{-1/2}(n-N)|^2 = O(1)$,

$$\sup_{x\in\mathcal{S}_j} \mathbf{E}\left[\eta_{jn}^2(x)\right] \leq \mathbf{E}\left| \sqrt{n}\left(\frac{n-N}{n}\right) \right|^2 \cdot \sup_{x\in\mathcal{S}_j} \left| \frac{1}{h^d} \cdot \mathbf{E}\left[Y_{ji} K\left(\frac{x-X_i}{h}\right) \right] \right|^2 = O(1).$$

Proof of Claim 2: Note that

$$\left| \sqrt{nh^d} \frac{s_{jn}(x)}{u_{jn}(x)} \right| = \left| \sum_{i=n+1}^{N} \bar{V}_{n,ji}(x) \right|. \tag{6.27}$$

Certainly $Var(\bar{V}_{n,ji}(x)) = 1$. As seen in (6.11), $\sup_{x\in\mathcal{S}_j}\mathbf{E}\left|\bar{V}_{n,ji}(x)\right|^3 \le Ch^{-d/2}$ for some $C>0$. Similarly,

$$\sup_{x\in\mathcal{S}_j}\mathbf{E}\left|\bar{V}_{n,ji}(x)\right|^4 \le \frac{h^d k_{jn,4}(x)}{h^{2d}\left(\rho_{jn}^2(x) - h^d b_{jn}^2(x)\right)^2} \le Ch^{-d},$$

for some $C>0$. Hence by Lemma 1(i) of Horváth (1991), for some $C>0$,

$$\begin{aligned}\mathbf{E}\left(\sqrt{nh^d}\frac{s_{jn}(x)}{u_{jn}(x)}\right)^2 \le\ & \mathbf{E}|N-n|\mathbf{E}|\mathbb{Z}_1|^2 \\ & +C\left\{\mathbf{E}\left|N-n\right|^{1/2}\mathbf{E}\left|\bar{V}_{n,ji}(x)\right|^3 + \mathbf{E}\left|\bar{V}_{n,ji}(x)\right|^4\right\}.\end{aligned}$$

Note that $\mathbf{E}|N-n| = O(n^{1/2})$ and $\mathbf{E}\left|N-n\right|^{1/2} = O(n^{1/4})$ (e.g. (2.21) and (2.22) of Horváth (1991)). Therefore, there exists $C>0$ such that

$$\sup_{x\in\mathcal{S}_j}\mathbf{E}\left(\sqrt{nh^d}\frac{s_{jn}(x)}{u_{jn}(x)}\right)^2 \le C\left\{n^{1/2} + n^{1/4}h^{-d/2} + h^{-d}\right\}.$$

Since $n^{-1/2}h^{-d}\to 0$, $\sup_{x\in\mathcal{S}_j}\mathbf{E}[(s_{jn}(x)/u_{jn}(x))^2] = o(1)$.

Proof of Claim 3: By (6.8), Lemmas A3-A4, and (6.26), we have

$$\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[\left|\sqrt{nh^d}\frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)}\right|^{2p-2}\right] = \sup_{x\in\mathcal{S}_j}\left|\frac{\rho_{jn}(x)}{\tilde{\rho}_{jn}(x)}\right|^{2p-2}\mathbf{E}\left(\left|\frac{\sqrt{nh^d}v_{jN}(x)}{\rho_{jn}(x)}\right|^{2p-2}\right) \le C,$$

for some $C>0$. This completes the proof of Claim 3.

Now, using Claims 1-3, we prove Step 1. Let $\mu_j(A) \equiv \int_A w_j(x)dx$. Since $h^{(p-1)d/2} = O(1)$ when $p=1$, and $\sqrt{a+b}\le\sqrt{a}+\sqrt{b}$ for any $a\ge 0$ and $b\ge 0$,

$$\begin{aligned}\mathbf{E}\left[A_{1n}\right] \le\ & C\int_A \mathbf{E}\left[|\eta_{jn}(x)|\left(\left|\sqrt{nh^d}\frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)}\right|^{p-1} + \left|\sqrt{nh^d}\frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)}\right|^{p-1}\right)\right]w_j(x)dx \\ \le\ & C\mu_j(A)\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[|\eta_{jn}(x)|\left(\left|\sqrt{nh^d}\frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)}\right|^{p-1} + \left|\sqrt{nh^d}\frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)}\right|^{p-1}\right)\right] \\ \le\ & 2C\mu_j(A)\times\left(\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[\eta_{jn}^2(x)\right]\right)^{1/2} \\ & \times\left(\left(\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[\left|\sqrt{nh^d}\frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)}\right|^{2p-2}\right]\right)^{1/2} + \left(\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[\left|\sqrt{nh^d}\frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)}\right|^{2p-2}\right]\right)^{1/2}\right).\end{aligned}$$

Certainly, as in the proof of Lemma A5,

$$\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[\left|\sqrt{nh^d}\frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)}\right|^{2p-2}\right]\leq C, \tag{6.28}$$

for some constant $C>0$. Hence using Claims 1 and 3, we conclude that $\mathbf{E}\left[A_{1n}\right]\leq C\mu_j(A)$ for some $C>0$. As for A_{2n}, similarly, we obtain that for some $C>0$,

$$\begin{aligned}\mathbf{E}\left[A_{2n}\right] &\leq C\int_A\mathbf{E}\left[\left|\frac{s_{jn}(x)}{u_{jn}(x)}\right|\left(\left|\sqrt{nh^d}\frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)}\right|^{p-1}+\left|\sqrt{nh^d}\frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)}\right|^{p-1}\right)\right]w_j(x)dx\\ &\leq 2C\mu_j(A)\times\left(\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[\left|\frac{s_{jn}(x)}{u_{jn}(x)}\right|^2\right]\right)^{1/2}\\ &\quad\times\left(\left(\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[\left|\sqrt{nh^d}\frac{v_{jn}(x)}{\tilde{\rho}_{jn}(x)}\right|^{2p-2}\right]\right)^{1/2}+\left(\sup_{x\in\mathcal{S}_j}\mathbf{E}\left[\left|\sqrt{nh^d}\frac{v_{jN}(x)}{\tilde{\rho}_{jn}(x)}\right|^{2p-2}\right]\right)^{1/2}\right).\end{aligned}$$

By Claims 2 and 3 and (6.28), $\mathbf{E}\left[A_{2n}\right]=o(1)$. Hence the proof of Step 1 is completed.

Proof of Step 2: We can follow the proof of Lemma A7(i) to show that

$$\mathbf{E}\left[n^{p/2}h^{(p-1)d/2}\int_A\left(|v_{jN}(x)|^p-\mathbf{E}\left[|v_{jN}(x)|^p\right]\right)w_j(x)dx\right]^2=\kappa_{jn}(A)+o(1),$$

where $\kappa_{jn}(A)\equiv\int_A\int_{[-1,1]^d}r_{jn}(x,u)\lambda_{jn}(x,x+uh)dudx$,

$$\begin{aligned}\lambda_{jn}(x,z) &\equiv \rho_{jn}^p(x)\rho_{jn}^p(z)w_j(x)w_j(z)1_{A\cap\mathcal{S}_j}(x)1_{A\cap\mathcal{S}_j}(z)\text{ and}\\ r_{jn}(x,u) &\equiv Cov\left(|Z_{jn,A}(x)|^p,|Z_{jn,B}(x+uh)|^p\right),\end{aligned}$$

with $(Z_{jn,A}(x),Z_{jn,B}(x+uh))'\in\mathbf{R}^2$ denoting a centered normal random vector whose covariance matrix is equal to that of $(\xi_{jn}(x),\xi_{jn}(x+uh))'$. By Cauchy-Schwarz inequality and Lemma A5,

$$\sup_{x\in\mathcal{S}_j}r_{jn}(x,u)\leq\sup_{x\in\mathcal{S}_j}\sqrt{\mathbf{E}\left|Z_{jn,A}(x)\right|^{2p}\mathbf{E}\left|Z_{jn,B}(x+uh)\right|^{2p}}<\infty.$$

Furthermore, for each $u\in[-1,1]^d$,

$$\int_A\lambda_{jn}(x,x+uh)dx\leq\sqrt{\int_A w_j^2(x)dx}\sqrt{\int_{A+uh}w_j^2(x)dx}.$$

Since $\int_{\mathcal{S}_j^\varepsilon}w_j^2(x)dx<\infty$ for some $\varepsilon>0$ (Assumption 1(ii)), we find that as $h\to 0$, the last term converges to $\int_A w_j^2(x)dx$. We obtain the desired result of Step 2.

Proof of Step 3: The convergence above follows from the proof of Lemma A6. ■

Let $\mathcal{C} \subset \mathbf{R}^d$ be a bounded Borel set such that

$$\alpha \equiv P\left\{X \in \mathbf{R}^d \backslash \mathcal{C}\right\} > 0.$$

For any Borel set $A \subset \mathcal{C}$, let

$$\begin{aligned}\zeta_n(A) &\equiv \sum_{j=1}^{J} \int_A \Lambda_p(v_{jn}(x))w_j(x)dx \text{ and} \\ \zeta_N(A) &\equiv \sum_{j=1}^{J} \int_A \Lambda_p(v_{jN}(x))w_j(x)dx.\end{aligned}$$

We also let $\sigma_n^2(A) \equiv \sum_{j=1}^J \sum_{k=1}^J \sigma_{jk,n}(A)$, and $\sigma^2(A) \equiv \sum_{j=1}^J \sum_{k=1}^J \sigma_{jk}(A)$. We define

$$S_n(A) \equiv \frac{n^{p/2}h^{(p-1)d/2}\{\zeta_N(A) - \mathbf{E}\zeta_N(A)\}}{\sigma_n(A)},$$

where

$$\begin{aligned}U_n &\equiv \frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{N} 1\left\{X_i \in \mathcal{C}\right\} - nP\left\{X \in \mathcal{C}\right\}\right\}, \text{ and} \\ V_n &\equiv \frac{1}{\sqrt{n}}\left\{\sum_{i=1}^{N} 1\{X_i \in \mathbf{R}^d \backslash \mathcal{C}\} - nP\{X \in \mathbf{R}^d \backslash \mathcal{C}\}\right\}.\end{aligned}$$

Lemma A9: *Suppose that Assumptions 1 and 2 hold. Furthermore, assume that as* $n \to \infty$, $h \to 0$, *and* $n^{-1/2}h^{-d} \to 0$. *Let* $A \subset \mathcal{C}$ *be such that* $\sigma^2(A) > 0$, $\alpha \equiv P\{X \in \mathbf{R}^d \backslash \mathcal{C}\} > 0$, $\rho_j(\cdot)$*'s and* $w_j(\cdot)$*'s are continuous and bounded on* A, *and condition in* (6.13) *is satisfied for all* $l = 1, \cdots, J$. *Then,*

$$(S_n(A), U_n) \xrightarrow{d} (\mathbb{Z}_1, \sqrt{1-\alpha}\mathbb{Z}_2).$$

Proof : First, we show that

$$Cov\,(S_n(A), U_n) \to 0. \tag{6.29}$$

Write

$$Cov\,(S_n(A), U_n) = \frac{n^{p/2}h^{(p-1)d/2}}{\sigma_n(A)} \sum_{j=1}^{J} \int_A Cov\,(\Lambda_p(v_{jN}(x)), U_n)\, w_j(x)dx.$$

It suffices for (6.29) to show that

$$Cov\left(n^{p/2}h^{pd/2}\{\zeta_N(A) - \mathbf{E}\zeta_N(A)\}, U_n\right) = o(h^{d/2}), \tag{6.30}$$

since $\sigma_n^2(A) \to \sigma^2(A) \equiv \sum_{j=1}^J \sum_{k=1}^J \sigma_{jk}(A) > 0$ by Lemma A7. For any $x \in \mathcal{S}_j$,

$$\left(\frac{\sqrt{nh^d} v_{jN}(x)}{\rho_{jn}(x)}, \frac{U_n}{\sqrt{P\{X \in \mathcal{C}\}}} \right) \stackrel{d}{=} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n Q_n^{(k)}(x), \frac{1}{\sqrt{n}} \sum_{k=1}^n U^{(k)} \right),$$

where $(Q_n^{(k)}(x), U^{(k)})$'s are i.i.d. copies of $(Q_n(x), U)$ with

$$\begin{aligned} Q_n(x) &\equiv \frac{1}{h^{d/2}\rho_{jn}(x)} \left\{ \sum_{i \le N_1} Y_{ji} K\left(\frac{x - X_i}{h}\right) - \mathbf{E}\left[Y_{ji} K\left(\frac{x - X_i}{h}\right)\right]\right\} \text{ and} \\ U &\equiv \frac{\sum_{i \le N_1} 1\{X_i \in \mathcal{C}\} - P\{X \in \mathcal{C}\}}{\sqrt{P\{X \in \mathcal{C}\}}}. \end{aligned}$$

Uniformly over $x \in \mathcal{S}_j$,

$$r_n(x) \equiv \mathbf{E}[Q_n(x)U] = O(h^{d/2}) = o(1), \tag{6.31}$$

by Lemma A4. Let $(Z_{1n}, Z_{2n})'$ be a centered normal random vector with the same covariance matrix as that of $(Q_n(x), U)'$. Let the 2 by 2 covariance matrix be $\Sigma_{n,2}$.

Since $\frac{1}{\sqrt{n}} \sum_{k=1}^n U^{(k)}$ and Z_{2n} have mean zero, we write

$$\begin{aligned} & Cov\left(\Lambda_p\left(\frac{1}{\sqrt{n}} \sum_{k=1}^n Q_n^{(k)}(x)\right), \frac{1}{\sqrt{n}} \sum_{k=1}^n U^{(k)} \right) - Cov\left(\Lambda_p(Z_{1n}), Z_{2n}\right) \\ = \; & \mathbf{E}\left[\Lambda_p\left(\frac{1}{\sqrt{n}} \sum_{k=1}^n Q_n^{(k)}(x)\right) \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n U^{(k)} \right)\right] - \mathbf{E}\left[\Lambda_p(Z_{1n}) Z_{2n}\right] \equiv A_n(x), \text{ say.} \end{aligned}$$

Define $\bar{\Lambda}_{n,p}(v) \equiv \Lambda_p([\Sigma_{n,2}^{1/2} v]_1)[\Sigma_{n,2}^{1/2} v]_2$, $v \in \mathbf{R}^2$. There exists some $C > 0$ such that for all $n \ge 1$,

$$\begin{aligned} \sup_{v \in \mathbf{R}^2} \frac{\left|\bar{\Lambda}_{n,p}(v) - \bar{\Lambda}_{n,p}(0)\right|}{1 + ||v||^{p+1} \min\{||v||, 1\}} &\le C \text{ and} \\ \int \sup_{u \in \mathbf{R}^2 : ||z-u|| \le \delta} \left|\bar{\Lambda}_{n,p}(z) - \bar{\Lambda}_{n,p}(u)\right| d\Phi(z) &\le C\delta, \text{ for all } \delta \in (0,1]. \end{aligned}$$

Letting $W_n^{(k)}(x) \equiv \Sigma_{n,2}^{-1/2} \cdot (Q_n^{(k)}(x), U^{(k)})'$, observe that using (6.31) and following the arguments in (6.21), from some large n on, for some $C > 0$,

$$\begin{aligned} \mathbf{E}||W_n^{(k)}(x)||^3 &= \mathbf{E}||\Sigma_{n,2}^{-1/2}(Q_n^{(k)}(x), U^{(k)})'||^3 \\ &= \mathbf{E}[\{tr(\Sigma_{n,2}^{-1/2}(Q_n^{(k)}(x), U^{(k)})'(Q_n^{(k)}(x), U^{(k)})\Sigma_{n,2}^{-1/2})\}^{3/2}] \\ &\le C(1 - r_n^2(x))^{-3/2} \mathbf{E}\left[|Q_n(x)|^3 + |U|^3\right] \le Ch^{-d/2}. \end{aligned}$$

Hence, by Lemma A2,

$$\begin{aligned}\sup_{x\in\mathcal{S}_j}|A_n(x)| &= \sup_{x\in\mathcal{S}_j}\left|\mathbf{E}\bar{\Lambda}_{n,p}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n W_n^{(i)}(x)\right)-\mathbf{E}\bar{\Lambda}_{n,p}\left(\tilde{Z}_n\right)\right| \\ &= O\left(n^{-1/2}h^{-d/2}\right)=o(h^{d/2}),\end{aligned}$$

where $\tilde{Z}_n \equiv \Sigma_{n,2}^{-1/2}(Z_{1n},Z_{2n})'$. This completes the proof of (6.30) and hence that of (6.29).

Now, define

$$\Delta_n(x)\equiv n^{p/2}h^{(p-1)d/2}\sum_{j=1}^J\{\Lambda_p(v_{jN}(x))-\mathbf{E}[\Lambda_p(v_{jN}(x))]\}w_j(x).$$

Following Mason and Polonik (2009), we slice the integral $\int_{\mathcal{X}}\Delta_n(x)dx$ into a sum of a 1-dependent random field. Let $\mathcal{C}$ be as given in the lemma. Let $\mathbb{Z}^d$ be the set of d-tupes of integers, and let $\{R_{n,\mathbf{i}}:\mathbf{i}\in\mathbb{Z}^d\}$ be the collection of rectangles in $\mathbf{R}^d$ such that $R_{n,\mathbf{i}}=[a_{n,\mathbf{i}_1},b_{n,\mathbf{i}_1}]\times\cdots\cdot\times[a_{n,\mathbf{i}_d},b_{n,\mathbf{i}_d}]$, where $\mathbf{i}_j$ is the j-th entry of $\mathbf{i}$, and $h\le b_{n,\mathbf{i}_j}-a_{n,\mathbf{i}_j}\le 2h$, for all $j=1,\cdot\cdot\cdot,d$, and two different rectangles $R_{n,\mathbf{i}}$ and $R_{n,\mathbf{j}}$ do not have intersection with nonempty interior, and the union of the rectangles $R_{n,\mathbf{i}}$, $\mathbf{i}\in\mathbb{Z}_n^d$, cover $\mathcal{C}$, from some sufficiently large n on, where $\mathbb{Z}_n^d$ be the set of d-tuples of integers whose absolute values less than or equal to n.

We let $B_{n,\mathbf{i}}=R_{n,\mathbf{i}}\cap\mathcal{C}$ and $\mathcal{I}_n\equiv\{\mathbf{i}\in\mathbb{Z}_n^d:B_{n,\mathbf{i}}\neq\varnothing\}$. Then $B_{n,\mathbf{i}}$ has Lebesgue measure $m(B_{n,\mathbf{i}})$ bounded by C_1h^d and the cardinality of the set $\mathcal{I}_n$ is bounded by C_2h^{-d} for some positive constants C_1 and C_2. Define

$$\begin{aligned}\alpha_{\mathbf{i},n} &\equiv \frac{1}{\sigma_n(A)}\int_{B_{n,\mathbf{i}}\cap A}\Delta_n(x)dx \text{ and} \\ u_{\mathbf{i},n} &\equiv \frac{1}{\sqrt{n}}\left\{\sum_{j=1}^N 1\{X_j\in B_{n,\mathbf{i}}\}-nP\{X_j\in B_{n,\mathbf{i}}\}\right\}.\end{aligned}$$

Then, we can write

$$S_n(A)=\sum_{\mathbf{i}\in\mathcal{I}_n}\alpha_{\mathbf{i},n} \text{ and } U_n=\sum_{\mathbf{i}\in\mathcal{I}_n}u_{\mathbf{i},n}.$$

Certainly $Var(S_n(A))=1$ and it is easy to check that $Var(U_n)=1-\alpha$. Take $\mu_1,\mu_2\in\mathbf{R}$ and let

$$y_{\mathbf{i},n}\equiv\mu_1\alpha_{\mathbf{i},n}+\mu_2u_{\mathbf{i},n}.$$

From (6.29),

$$Var\left(\sum_{\mathbf{i}\in\mathcal{I}_n}y_{\mathbf{i},n}\right)\to\mu_1^2+\mu_2^2(1-\alpha) \text{ as } n\to\infty.$$

Since $\sigma_n^r(A) = \sigma^r(A) + o(1),\ r > 0$, by Lemma A7 and $m(B_{n,\mathbf{i}}) \le Ch^d$ for a constant $C > 0$, we take $r \in (2, (2p+2)/p]$ and bound

$$\begin{aligned} &\sigma_n^r(A) \sum_{\mathbf{i} \in \mathcal{I}_n} \mathbf{E}|\alpha_{\mathbf{i},n}|^r \\ \le\ & C \sup_{x \in A} \mathbf{E}\left|\Delta_n(x)\right|^r \sum_{\mathbf{i} \in \mathcal{I}_n} \left(\int_A \int_A \int_A 1_{B_{n,\mathbf{i}}}(u,v,s) du dv ds \right)^{r/3}, \end{aligned}$$

where $1_B(u,v,s) \equiv 1\{u \in B\}1\{v \in B\}1\{s \in B\}$. Using Jensen's inequality, we have

$$\begin{aligned} \sup_{x \in A} \mathbf{E}\left|\Delta_n(x)\right|^r &\le C_1 n^{rp/2} h^{r(p-1)d/2} \sup_{x \in A} \sum_{j=1}^J \mathbf{E}\left|v_{jN}(x)\right|^{rp} w_j^r(x) \\ &\le C_2 n^{rp/2} h^{r(p-1)d/2} \max_{1 \le j \le J} \sup_{x \in A \cap \mathcal{S}_j} \mathbf{E}\left|v_{jN}(x)\right|^{rp} \end{aligned}$$

for some $C_1, C_2 > 0$. As for the last term, we apply Rosenthal's inequality (see. e.g. Lemma 2.3. of GMZ): for some constant $C > 0$,

$$\begin{aligned} & n^{rp/2} h^{r(p-1)d/2} \sup_{x \in A \cap \mathcal{S}_j} \mathbf{E}\left|v_{jN}(x)\right|^{rp} \\ \le\ & C h^{r(p-1)d/2} \sup_{x \in A \cap \mathcal{S}_j} \left(\frac{1}{h^{2d}} \mathbf{E}\left[Y_{ji}^2 K^2 \left(\frac{x - X_i}{h} \right) \right] \right)^{rp/2} \\ & + C h^{r(p-1)d/2} \sup_{x \in A \cap \mathcal{S}_j} \left(\frac{n}{n^{rp/2} h^{rpd}} \mathbf{E}\left[\left| Y_{ji} K \left(\frac{x - X_i}{h} \right) \right|^{rp} \right] \right). \end{aligned}$$

By Lemma A4, the first term is $O(h^{-rd/2})$ and the last term is $O(n^{1-rp/2}h^{-rdp/2-rd/2+d})$. Hence we find that

$$\begin{aligned} \sum_{\mathbf{i} \in \mathcal{I}_n} \mathbf{E}|\alpha_{\mathbf{i},n}|^r &= \text{Cardinality of } \mathcal{I}_n \times O\left(m(B_{n,\mathbf{i}})^r h^{-rd/2} \{1 + n^{1-rp/2} h^{-rdp/2+d}\} \right) \\ &= O\left(h^{rd/2-d} \{1 + n^{1-rp/2} h^{-rdp/2+d}\} \right) = o(1) \end{aligned}$$

for any $r \in (2, (2p+2)/p]$, because $n^{-1/2}h^{-d} \to 0$. Therefore, as $n \to \infty$,

$$\sum_{\mathbf{i} \in \mathcal{I}_n} \mathbf{E}|\alpha_{\mathbf{i},n}|^r \to 0 \text{ for any } r \in (2, (2p+2)/p].$$

Also, arguing similarly as in (6.56) of GMZ, we can show that $\sum_{\mathbf{i} \in \mathcal{I}_n} \mathbf{E}|u_{\mathbf{i},n}|^r \to 0$ as $n \to \infty$ for any $r \in (2, (2p+2)/p]$. Since X_i's are common across different j's, the sequence $\{y_{\mathbf{i},n}\}_{\mathbf{i} \in \mathcal{I}_n}$ is a 1-dependent random field (see Mason and Polonik (2009)). The desired result of Lemma A9 follows by Theorem 1 of Shergin (1993) and the Cramér-Wold device. ■

LEMMA A10: *Suppose that the conditions of Lemma A9 are satisfied, and let $A \subset \mathbf{R}^d$ be a Borel set in Lemma A9. Then,*

$$\frac{n^{p/2}h^{(p-1)d/2}\left\{\zeta_n(A) - \mathbf{E}\zeta_n(A)\right\}}{\sigma_n(A)} \xrightarrow{d} N(0,1), \text{ as } n \to \infty.$$

PROOF: The conditional distribution of $S_n(A)$ given $N = n$ is equal to that of

$$\frac{n^{p/2}h^{(p-1)d/2}}{\sigma_n(A)} \sum_{j=1}^{J} \int_A \left\{\Lambda_p(v_{jn}(x)) - \mathbf{E}\Lambda_p(v_{jN}(x))\right\} w_j(x)dx.$$

Using Lemma A9 and the de-Poissonization argument of Beirlant and Mason (1995) (see also Lemma 2.4 of GMZ), this conditional distribution converges to $N(0,1)$. Now by Lemma A6, it follows that

$$n^{p/2}h^{(p-1)d/2} \sum_{j=1}^{J} \int_A \left\{\mathbf{E}\Lambda_p(v_{jN}(x)) - \mathbf{E}\Lambda_p(v_{jn}(x))\right\} w_j(x)dx \to 0,$$

as $n \to \infty$. This completes the proof. ■

PROOF OF THEOREM 1 : Fix $\varepsilon > 0$ as in Assumption 1(iii), and take $n_0 > 0$ such that for all $n \geq n_0$,

$$\{x - uh : x \in \mathcal{S}_j, u \in [-1/2, 1/2]^d\} \subset \mathcal{S}_j^{\varepsilon} \subset \mathcal{X} \text{ for all } j \in \mathcal{J}.$$

Since we are considering the least favorable case of the null hypothesis,

$$\mathbf{E}[Y_{ji}K((x - X_i)/h)]/h^d = \int_{[-1/2,1/2]^d} m_j(x - uh)K(u)du = 0, \text{ for almost all } x \in \mathcal{S}_j,$$

for all $n \geq n_0$ and for all $j \in \mathcal{J}$. Therefore, $\hat{g}_{jn}(x) = v_{jn}(x)$ for almost all $x \in \mathcal{S}_j$, $j \in \mathcal{J}$, and for all $n \geq n_0$. From here on, we consider only $n \geq n_0$.

We fix $0 < \varepsilon_l \to 0$ as $l \to \infty$ and take a compact set $\mathcal{W}_l \subset \mathcal{S}_j$ such that for each $j \in \mathcal{J}$, w_j is bounded and continuous on $\mathcal{W}_l$ and for $s \in \{1, 2\}$,

$$\int_{\mathcal{X} \backslash \mathcal{W}_l} w_j^s(x)dx \to 0 \text{ as } l \to \infty. \tag{6.32}$$

We can choose such $\mathcal{W}_l$ following the arguments in the proof of Lemma 6.1 of GMZ because w_j^s is integrable by Assumption 1(ii). Take $M_{l,j}, v_{l,j} > 0$, $j = 1, 2, \cdots, J$, such that for $\mathcal{C}_{l,j} \equiv [-M_{l,j} + v_{l,j}, M_{l,j} - v_{l,j}]^d$,

$$P\left\{X_i \in \mathbf{R}^d \backslash \mathcal{C}_{l,j}\right\} > 0,$$

and for some Borel $A_{l,j} \subset \mathcal{C}_{l,j} \cap \mathcal{W}_l$, $\rho_j(\cdot)$ is bounded and continuous on $A_{l,j}$,

$$\begin{aligned} \sup_{x \in A_{l,j}} |\rho_{jn}(x) - \rho_j(x)| &\to 0, \text{ as } n \to \infty, \text{ and} \\ \int_{\mathcal{W}_l \backslash A_{l,j}} \rho_j(x) w_j^s(x) dx &\to 0, \text{ as } l \to \infty, \text{ for } s \in \{1,2\}. \end{aligned} \tag{6.33}$$

The existence of $M_{l,j}, v_{l,j}$ and ε_l and the sets $A_{l,j}$ are ensured by Lemma A1. By Assumption 1(i), we find that the second convergence in (6.33) implies that $\int_{\mathcal{W}_l \backslash A_{l,j}} w_j^s(x) dx \to 0$ as $l \to \infty$, for $s \in \{1,2\}$. Now, take $A_l \equiv \cap_{j=1}^J A_{l,j}$ and $\mathcal{C}_l \equiv \cap_{j=1}^J \mathcal{C}_{l,j}$, and observe that for $s \in \{1,2\}$,

$$\int_{\mathcal{W}_l \backslash A_l} w_j^s(x) dx \le \sum_{j=1}^J \int_{\mathcal{W}_l \backslash A_{l,j}} w_j^s(x) dx \to 0, \tag{6.34}$$

as $l \to \infty$ for all $j \in \mathcal{J}$.

First, we write

$$\begin{aligned} & \frac{\sum_{j=1}^J \left\{ n^{p/2} h^{(p-1)d/2} \Gamma_j(\hat{g}_{jn}) - a_{jn} \right\}}{\sigma_n} \\ = & \frac{n^{p/2} h^{(p-1)d/2}}{\sigma_n} \left\{ \zeta_n(\mathcal{X} \backslash \mathcal{W}_l) - \mathbf{E}\zeta_n(\mathcal{X} \backslash \mathcal{W}_l) \right\} \\ & + \frac{n^{p/2} h^{(p-1)d/2}}{\sigma_n} \left\{ \zeta_n(\mathcal{W}_l \backslash A_l) - \mathbf{E}\zeta_n(\mathcal{W}_l \backslash A_l) \right\} \\ & + \frac{n^{p/2} h^{(p-1)d/2}}{\sigma_n} \left\{ \zeta_n(A_l) - \mathbf{E}\zeta_n(A_l) \right\}. \end{aligned} \tag{6.35}$$

Since $\mathcal{X} \backslash A_l = (\mathcal{X} \backslash \mathcal{W}_l) \cup (\mathcal{W}_l \backslash A_l)$, by Lemma A8, (6.32), and (6.34),

$$n^{p/2} h^{(p-1)d/2} \left\{ \zeta_n(\mathcal{X} \backslash A_l) - \mathbf{E}\zeta_n(\mathcal{X} \backslash A_l) \right\} \xrightarrow{p} 0, \text{ as } n \to \infty, \text{ and } l \to \infty. \tag{6.36}$$

Furthermore, we write $|\sigma_n^2 - \sigma_n^2(A_l)|$ as

$$\begin{aligned} & \sum_{j=1}^J \sum_{k=1}^J \int_{\mathcal{X}} q_{jk,p}(x) \left(1 - 1_{A_l}(x)\right) \rho_{jn}^p(x) \rho_{kn}^p(x) w_j(x) w_k(x) dx \\ \le & \sum_{j=1}^J \sum_{k=1}^J \sup_{x \in \mathcal{S}_j \cap \mathcal{S}_k} \left| q_{jk,p}(x) \rho_{jn}^p(x) \rho_{kn}^p(x) \right| \int_{\mathcal{X}} \left(1 - 1_{A_l}(x)\right) w_j(x) w_k(x) dx \\ = & \sum_{j=1}^J \sum_{k=1}^J \sup_{x \in \mathcal{S}_j \cap \mathcal{S}_k} \left| q_{jk,p}(x) \rho_{jn}^p(x) \rho_{kn}^p(x) \right| \int_{\mathcal{X} \backslash A_l} w_j(x) w_k(x) dx. \end{aligned}$$

Observe that as $l \to \infty$,

$$\left| \int_{\mathcal{X} \setminus A_l} w_j(x) w_k(x) dx \right|^2 \leq \left(\int_{\mathcal{X} \setminus A_l} w_j^2(x) dx \right) \left(\int_{\mathcal{X} \setminus A_l} w_k^2(x) dx \right) \to 0.$$

From Lemma A4, it follows that

$$\lim_{l \to \infty} \text{limsup}_{n \to \infty} \left| \sigma_n^2 - \sigma_n^2(A_l) \right| = 0. \tag{6.37}$$

Furthermore, since $\sigma_n^2(A_l) \to \sigma^2(A_l)$ as $n \to \infty$ for each l by Lemma A7, and $\sigma^2(A_l) \to \sigma^2 > 0$ as $l \to \infty$, by Assumption 1, it follows that for any $\varepsilon_1 \in (0, \sigma^2)$,

$$\begin{aligned} 0 &< \sigma^2 - \varepsilon_1 \leq \text{liminf}_{n \to \infty} \sigma_n^2 \\ &\leq \text{limsup}_{n \to \infty} \sigma_n^2 \leq \sigma^2 + \varepsilon_1 < \infty. \end{aligned} \tag{6.38}$$

Combining this with (6.36), we find that as $n \to \infty$ and $l \to \infty$,

$$\frac{n^{p/2} h^{(p-1)d/2}}{\sigma_n} \left\{ \zeta_n(\mathcal{X} \setminus A_l) - \mathbf{E} \zeta_n(\mathcal{X} \setminus A_l) \right\} = o_P(1).$$

As for the last term in (6.35), by (6.38) and Lemma A10, as $n \to \infty$ and $l \to \infty$,

$$n^{p/2} h^{(p-1)d/2} \left| \zeta_n(A_l) - \mathbf{E} \zeta_n(A_l) \right| = O_P(1).$$

Therefore, by (6.37),

$$\begin{aligned} & \frac{n^{p/2} h^{(p-1)d/2}}{\sigma_n} \left\{ \zeta_n(A_l) - \mathbf{E} \zeta_n(A_l) \right\} \\ = & \frac{n^{p/2} h^{(p-1)d/2}}{\sigma_n(A_l)} \left\{ \zeta_n(A_l) - \mathbf{E} \zeta_n(A_l) \right\} + o_P(1), \end{aligned}$$

where $o_P(1)$ is a term that vanishes in probability as $n \to \infty$ and $l \to \infty$. For each $l \geq 1$, the last term converges in distribution to $N(0, 1)$ by Lemma A10. Since $\sigma_n^2(A_l) \to \sigma^2$ as $n \to \infty$ and $l \to \infty$, we conclude that

$$\sum_{j=1}^{J} \left\{ n^{p/2} h^{(p-1)d/2} \Gamma_j(\hat{g}_{jn}) - a_{jn} \right\} \xrightarrow{d} N\left(0, \sigma^2\right).$$

■

6.3. **Proofs of Other Theorems.** We now give proofs of other theorems in the paper.
Proof of Theorem 2 : We first show that for each $j \in \mathcal{J}$,

$$\begin{aligned} \hat{a}_{jn} &= a_{jn} + O_P(n^{-1/2} h^{-3d/2}) \text{ and} \\ \hat{\sigma}_n^2 &= \sigma_n^2 + O_P(n^{-1/2} h^{-3d/2}). \end{aligned} \tag{6.39}$$

For this, we show that for all $j, k = 1, \cdots, J$,

$$\sup_{x \in \mathcal{S}_j \cap \mathcal{S}_k} |\hat{\rho}_{jk,n}(x) - \rho_{jk,n}(x)| = O_P\left(n^{-1/2}h^{-d}\right). \tag{6.40}$$

Write $\sup_{x \in \mathcal{S}_j \cap \mathcal{S}_k} |\hat{\rho}_{jk,n}(x) - \rho_{jk,n}(x)|$ as

$$\sup_{x \in \mathcal{S}_j \cap \mathcal{S}_k} \left| \frac{1}{nh^d} \sum_{i=1}^{n} \left\{ Y_{ji}Y_{ki}K^2\left(\frac{x - X_i}{h}\right) - \mathbf{E}\left[Y_{ji}Y_{ki}K^2\left(\frac{x - X_i}{h}\right)\right] \right\} \right|.$$

Let $\varphi_{n,x}(y_1, y_2, z) \equiv y_1 y_2 K^2((x - z)/h)$ and $\mathcal{K}_n \equiv \{\varphi_{n,x}(\cdot, \cdot, \cdot) : x \in \mathcal{S}_j \cap \mathcal{S}_k\}$. We define $N(\varepsilon, \mathcal{K}_n, L_2(Q))$ to be a covering number of $\mathcal{K}_n$ with respect to $L_2(Q)$, i.e., the smallest number of maps $\varphi_j,\ j = 1, \cdots, N_1$, such that for all $\varphi \in \mathcal{K}_n$, there exists φ_j such that $\int (\varphi_j - \varphi)^2 dQ \leq \varepsilon^2$. By Assumption 2(b), Lemma 2.6.16 of van der Vaart and Wellner (1996), and Lemma A.1 of Ghosal, Sen and van der Vaart (2000), we find that for some $C > 0$,

$$\sup_Q \log N(\varepsilon, \mathcal{K}_n, L_2(Q)) \leq C \log \varepsilon,$$

where the supremum is over all discrete probability measures. We take $\bar{\varphi}_n(y_1, y_2, z) \equiv y_1 y_2 ||K||_\infty^2$ to be the envelope of $\mathcal{K}_n$. By Theorem 2.14.1 of van der Vaart and Wellner (1996), we deduce that

$$n^{1/2}h^d \mathbf{E}\left[\sup_{x \in \mathcal{S}_j \cap \mathcal{S}_k} |\hat{\rho}_{jk,n}(x) - \rho_{jk,n}(x)|\right] \leq C,$$

for some positive constant C. This yields (6.40). In view of the definitions of $\hat{a}_{jn}$ and $\hat{\sigma}_n^2$, and Lemma A4, this completes the proof of (6.39).

Since $g_j(x) \leq 0$ for all $x \in \mathcal{X}$ under the null hypothesis and K is nonnegative,

$$\begin{aligned} \sup_{x \in \mathcal{S}_j} \mathbf{E}\hat{g}_{jn}(x) &= \sup_{x \in \mathcal{S}_j} \int g_j(x - uh) K(u)\, du \leq \int \sup_{x \in \mathcal{S}_j} g_j(x - uh) K(u)\, du \\ &\leq \int \sup_{x \in \mathcal{X}} g_j(x) K(u)\, du = \sup_{x \in \mathcal{X}} g_j(x) \leq 0, \end{aligned}$$

from some large n on. The second inequality follows from Assumption 1(iii). Therefore,

$$\int_{\mathcal{X}} \Lambda_p(\hat{g}_{jn}(x)) w_j(x) dx \leq \int_{\mathcal{X}} \Lambda_p(\hat{g}_{jn}(x) - \mathbf{E}\hat{g}_{jn}(x)) w_j(x) dx.$$

Hence by using this and (6.39), we bound $P\{\hat{T}_n > z_{1-\alpha}\}$ by

$$P\left\{ \frac{1}{\sigma_n} \sum_{j=1}^{J} \left\{ n^{p/2} h^{(p-1)d/2} \int_{\mathcal{X}} \Lambda_p(\hat{g}_{jn}(x) - \mathbf{E}\hat{g}_{jn}(x)) w_j(x) dx - a_{jn} \right\} > z_{1-\alpha} \right\} + o(1).$$

By Theorem 1, the leading probability converges to α as $n \to \infty$, delivering the desired result. ■

Proof of Theorem 3: Fix j such that $\Gamma_j(g_j) > 0$. We focus on the case with $p > 1$. The proof in the case with $p = 1$ is simpler and hence omitted. Using the triangular inequality, we bound $|\Gamma_j(\hat{g}_{jn}) - \Gamma_j(g_j)|$ by

$$\left|\int_{\mathcal{X}} \{\Lambda_p(\hat{g}_{jn}(x)) - \Lambda_p(\mathbf{E}\hat{g}_{jn}(x))\} w_j(x)dx\right| + \left|\int_{\mathcal{X}} \{\Lambda_p(\mathbf{E}\hat{g}_{jn}(x)) - \Lambda_p(g_j(x))\} w_j(x)dx\right|.$$

There exists n_0 such that for all $n \geq n_0$, $\sup_{x\in\mathcal{S}_j} |\mathbf{E}\hat{g}_{jn}(x)| < \infty$ by Lemma A4. Also, note that $\sup_{x\in\mathcal{S}_j}|g_j(x)| < \infty$ by Assumption 1(i). Hence, applying Lemma A3, from some large n on, for some $C_1, C_2 > 0$,

$$\begin{aligned} |\Gamma_j(\hat{g}_{jn}) - \Gamma_j(g_j)| \;\leq\; & C_1 \sum_{k=0}^{\lceil p-1\rceil} \int_{\mathcal{X}} |\hat{g}_{jn}(x) - \mathbf{E}\hat{g}_{jn}(x)|^{p-kz} w_j(x)dx \\ & + C_2 \sum_{k=0}^{\lceil p-1\rceil} \int_{\mathcal{X}} |\mathbf{E}\hat{g}_{jn}(x)) - g_j(x)|^{p-kz} w_j(x)dx, \end{aligned}$$

where $z = (p-1)/\lceil p-1\rceil$. Observe that $0 \leq z \leq 1$.

As for the second integral, take $\varepsilon > 0$ and a compact set $D \subset \mathbf{R}^d$ such that $\int_{\mathcal{X}\setminus D} w_j(x)dx < \varepsilon$ and g_j is continuous on D. Such a set D exists by Lemma A1. Since D is compact, g_j is in fact uniformly continuous on D. By change of variables,

$$\begin{aligned} \mathbf{E}\hat{g}_{jn}(x) - g_j(x) &= \int_{[-1/2,1/2]^d} \{g_j(x-uh)K(u) - g_j(x)\} du \\ &= \int_{[-1/2,1/2]^d} \{g_j(x-uh) - g_j(x)\} K(u)du \end{aligned}$$

and obtain that for $k = 0, 1, \cdots, p-1$,

$$\begin{aligned} & \int_{\mathcal{X}} |\mathbf{E}\hat{g}_{jn}(x) - g_j(x)|^{p-kz} w_j(x)dx \\ = \; & \int_{D} |\mathbf{E}\hat{g}_{jn}(x) - g_j(x)|^{p-kz} w_j(x)dx + \int_{\mathcal{X}\setminus D} |\mathbf{E}\hat{g}_{jn}(x) - g_j(x)|^{p-kz} w_j(x)dx \\ \leq \; & C_3 \sup_{u\in[-1/2,1/2]^d} \sup_{x\in D\cap\mathcal{S}_j} |g_j(x-uh) - g_j(x)|^{p-kz} \\ & + C_4 \int_{\mathcal{X}\setminus D} \int_{[-1/2,1/2]^d} |g_j(x-uh) - g_j(x)|^{p-kz} w_j(x)dudx, \end{aligned}$$

for some positive constants C_3 and C_4. Note that the constant C_4 involves $||K||_\infty$. The first term is $o(1)$ as $h \to 0$, because g_j is uniformly continuous on D. By Assumption

1(i), the last term is bounded by

$$C_5 \int_{\mathcal{X}\setminus D} w_j(x)dx < C_6\varepsilon, \text{ for some } C_5, C_6 > 0,$$

for some large n on. Since the choice of ε was arbitrary, we conclude that as $n \to \infty$,

$$|\Gamma_j(\hat{g}_{jn}) - \Gamma_j(g_j)| \leq C_1 \int_{\mathcal{X}} |\hat{g}_{jn}(x) - \mathbf{E}\hat{g}_{jn}(x)|^{p-kz} w_j(x)dx + o(1).$$

As for the leading integral, from the result of Theorem 1 (replacing $\Lambda_p(\cdot)$ there by $|\cdot|^{p-kz}$), we find that

$$\int_{\mathcal{X}} |\hat{g}_{jn}(x) - \mathbf{E}\hat{g}_{jn}(x)|^{p-kz} w_j(x)dx = O_P(n^{-(p-kz)/2}h^{-(p-kz-1)d/2-d/2}).$$

Since $n^{-1/2}h^{-d/2} \to 0$ by the condition of the theorem, we conclude that $\Gamma_j(\hat{g}_{jn}) \xrightarrow{p} \Gamma_j(g_j)$. Using the similar argument, we can also show that

$$\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2 \text{ and } \hat{a}_{jn} = O_P(h^{-d/2}) \text{ for all } j \in \mathcal{J},$$

where $\sigma^2 = \mathbf{1}'\Sigma\mathbf{1} > 0$. Hence

$$\hat{\sigma}_n^{-1}\{\Gamma_j(\hat{g}_{jn}) - n^{-p/2}h^{-pd/2}h^{d/2}\hat{a}_{jn}\} \xrightarrow{p} \sigma^{-1}\Gamma_j(g_j) > 0.$$

Therefore,

$$P\{\hat{T}_n > z_{1-\alpha}\} \geq P\left\{\sigma^{-1}\Gamma_j(g_j) > 0\right\} + o(1) \to 1,$$

where the inequality holds by the fact that $n^{-1/2}h^{-d/2} \to 0$ and $\hat{a}_{jn} = O_P(h^{-d/2})$. ■

Lemma A11: *Suppose that Assumptions 1-3 hold, $n^{-1/2}h^{-d} \to 0$, and that $\sqrt{n}g_j(\cdot) = \delta_j(\cdot)$, $j \in \mathcal{J}$, for real bounded functions δ_j, $j \in \mathcal{J}$, for each n. Then,*

$$\frac{1}{\sigma_n}\sum_{j=1}^{J}\left\{n^{p/2}h^{(p-1)d/2}\Gamma_{j,\delta}(\hat{g}_{jn}) - \tilde{a}_{jn}\right\} \xrightarrow{d} N(0,1),$$

where $\tilde{a}_{jn} \equiv \int \mathbf{E}\Lambda_p(h^{-d/(2p)}\rho_{jn}(x)\mathbb{Z}_1 + h^{d(p-1)/(2p)}\delta_{jn}(x))w_j(x)dx$ and $\delta_{jn}(x) \equiv \int \delta_j(x - uh)K(u)du$.

Proof: By change of variables,

$$\sqrt{n}\mathbf{E}\hat{g}_{jn}(x) = \sqrt{n}\int g_j(x-uh)K(u)du = \int \delta_j(x-uh)K(u)du.$$

Since δ_j is bounded, $\sup_{x\in\mathcal{S}_j}\sqrt{n}\,|\mathbf{E}\hat{g}_{jn}(x)| = O(1)$. Hence

$$\frac{\sqrt{nh^d}\hat{g}_{jN}(x)}{\rho_{jn}(x)} = \xi_{jn}(x) + \frac{\sqrt{nh^d}\mathbf{E}\hat{g}_{jn}(x)}{\rho_{jn}(x)} = \xi_{jn}(x) + O(h^{d/2}), \tag{6.41}$$

under the local alternatives. Using this and following the proof of Lemma A7, we find that under the local alternatives, $\sigma_{jk,n} \to \sigma_{jk}$. Also, as in the proof of Theorem 1, we use (6.41) and deduce that

$$\frac{1}{\sigma_n}\sum_{j=1}^{J} n^{p/2}h^{(p-1)d/2}\left\{\Gamma_j(\hat{g}_{jn}) - \mathbf{E}\Gamma_j(\hat{g}_{jn})\right\} \overset{d}{\to} N(0,1). \tag{6.42}$$

Now, as for $n^{p/2}h^{(p-1)d/2}\sigma_n^{-1}\mathbf{E}\Gamma_j(\hat{g}_{jn})$, We first note that

$$\begin{aligned} n^{p/2}h^{(p-1)d/2}\Gamma_j(\hat{g}_{jn}) &= h^{-d/2}\Gamma_j(n^{1/2}h^{d/2}\{\hat{g}_{jn} - \mathbf{E}\hat{g}_{jn}\} + n^{1/2}h^{d/2}\mathbf{E}\hat{g}_{jn}) \\ &= \Gamma_j(h^{-d/(2p)}\rho_{jn}(x)\xi_{jn}(x) + h^{(p-1)d/(2p)}\delta_{jn}(x)). \end{aligned}$$

We follow the proof of Lemma A4 and Lemma A6 (applying Lemma A2 with $\Lambda_p(v)$ in Lemma A6 replaced by $\Lambda_p(v + h^{d(p-1)/(2p)}\delta_{jn}(x)/\rho_{jn}(x))$) to deduce that

$$\int \left\{n^{p/2}h^{(p-1)d/2}\mathbf{E}\Lambda_p(\hat{g}_{jn}(x)) - \mathbf{E}\Lambda_p(\bar{Z}_{jn}(x))\right\} w_j(x)dx \to 0,$$

where $\bar{Z}_{jn}(x) \equiv h^{-d/(2p)}\rho_{jn}(x)\mathbb{Z}_1 + h^{d(p-1)/(2p)}\delta_{jn}(x)$. ■

Proof of Theorem 4: Under the local alternatives, by (6.39) and (6.42),

$$\begin{aligned} & P\{\hat{T}_n > z_{1-\alpha}\} \\ = \;& P\{\hat{\sigma}_n^{-1}\Sigma_{j=1}^{J}\{n^{p/2}h^{(p-1)d/2}\Gamma_j(\hat{g}_{jn}) - \hat{a}_{jn}\} > z_{1-\alpha}\} \\ = \;& P\{\sigma^{-1}\Sigma_{j=1}^{J}\{n^{p/2}h^{(p-1)d/2}\Gamma_j(\hat{g}_{jn}) - \tilde{a}_{jn} + \tilde{a}_{jn} - \hat{a}_{jn}\} > z_{1-\alpha}\} + o(1) \\ = \;& P\{\mathbb{Z}_1 + \sigma^{-1}\Sigma_{j=1}^{J}\{\tilde{a}_{jn} - a_{jn}\}) > z_{1-\alpha}\} + o(1). \end{aligned} \tag{6.43}$$

Fix $\varepsilon > 0$ and take a compact set $A_\varepsilon \subset \mathcal{S}_j$ such that $\int_{\mathcal{S}_j\setminus A_\varepsilon} w_j(x)dx < \varepsilon$. Furthermore, without loss of generality, let A_ε be a set on which $\delta_j(\cdot)$ and $\rho_j(\cdot)$ are uniformly continuous. Then for any $\varepsilon_1 > 0$, there exists $\lambda > 0$ such that $\sup_{z\in\mathbf{R}^d:||x-z||<\lambda}|\delta_j(z)-\delta_j(x)| \le \varepsilon_1$ uniformly over $x \in A_\varepsilon$. Hence from some large n on,

$$\sup_{x\in A_\varepsilon}|\delta_{jn}(x) - \delta_j(x)| \le \int_{[-1/2,1/2]^d}\sup_{x\in A_\varepsilon}|\delta_j(x-uh) - \delta_j(x)|\,K(u)du \le \varepsilon_1.$$

Since the choice of ε_1 was arbitrary, we conclude that $|\delta_{jn}(x)-\delta_j(x)| \to 0$ uniformly over $x \in A_\varepsilon$. Similarly, we also conclude that $|\rho_{jn}(x) - \rho_j(x)| \to 0$ uniformly over $x \in A_\varepsilon$. Using these facts, we analyze $\sigma^{-1}\Sigma_{j=1}^{J}\{\tilde{a}_{jn} - a_{jn}\}$ for each case of $p \in \{1, 2\}$.

(i) Suppose $p = 1$. For $\gamma > 0$ and $\mu \in \mathbf{R}$,

$$\begin{aligned}
\mathbf{E}\max\{\gamma\mathbb{Z}_1 + \mu, 0\} &= \mathbf{E}[\gamma\mathbb{Z}_1 + \mu|\gamma\mathbb{Z}_1 + \mu > 0]P\{\gamma\mathbb{Z}_1 + \mu > 0\} \\
&= \{\mu + \gamma\phi(-\mu/\gamma)/(1 - \Phi(-\mu/\gamma))\}(1 - \Phi(-\mu/\gamma)) \\
&= \mu(1 - \Phi(-\mu/\gamma)) + \gamma\phi(-\mu/\gamma) \\
&= \mu\Phi(\mu/\gamma) + \gamma\phi(\mu/\gamma).
\end{aligned}$$

Taking $\gamma_{jn} \equiv h^{-d/2}\rho_{jn}(x)$, we have

$$\begin{aligned}
&\mathbf{E}\max\{\gamma_{jn}\mathbb{Z}_1 + \delta_{jn}(x), 0\} - \mathbf{E}\max\{\gamma_{jn}\mathbb{Z}_1, 0\} \\
&= \delta_{jn}(x)\Phi(\delta_{jn}(x)/\gamma_{jn}) + \gamma_{jn}\phi(\delta_{jn}(x)/\gamma_{jn}) - \gamma_{jn}\phi(0) \\
&= \delta_{jn}(x)\Phi(0) + O(h^{d/2}),
\end{aligned}$$

uniformly in $x \in \mathcal{S}_j$. Therefore, we can write $\lim_{n\to\infty}\{\tilde{a}_{jn} - a_{jn}\}$ as

$$\begin{aligned}
&\lim_{n\to\infty}\int_{\mathcal{X}} \mathbf{E}[\Lambda_1(h^{-d/2}\rho_{jn}(x)\mathbb{Z}_1 + \delta_{jn}(x)) - \Lambda_1(h^{-d/2}\rho_{jn}(x)\mathbb{Z}_1)]w_j(x)dx \\
&= \frac{1}{2}\int_{A_\varepsilon} \delta_j(x)w_j(x)dx + \frac{1}{2}\lim_{n\to\infty}\int_{\mathcal{X}\setminus A_\varepsilon} \delta_{jn}(x)w_j(x)dx.
\end{aligned}$$

Since δ_{jn} is uniformly bounded, there exists $C > 0$ such that the last integral is bounded by $C\varepsilon$. Since the choice of $\varepsilon > 0$ was arbitrary, in view of (6.43), this gives the desired result.

(ii) Suppose $p = 2$. For $\gamma > 0$ and $\mu \in \mathbf{R}$,

$$\begin{aligned}
\mathbf{E}\max\{\gamma\mathbb{Z}_1 + \mu, 0\}^2 &= \mathbf{E}[(\gamma\mathbb{Z}_1 + \mu)^2|\gamma\mathbb{Z}_1 + \mu > 0]P\{\gamma\mathbb{Z}_1 + \mu > 0\} \\
&= (\mu^2 + \gamma^2)\Phi(\mu/\gamma) + \mu\gamma\phi(\mu/\gamma).
\end{aligned}$$

Taking $\gamma_{jn} \equiv h^{-d/4}\rho_{jn}(x)$ and $\mu_{jn} = h^{d/4}\delta_{jn}(x)$, we have

$$\begin{aligned}
&\mathbf{E}\max\{\gamma_{jn}\mathbb{Z}_1 + \mu_{jn}, 0\}^2 - \mathbf{E}\max\{\gamma_{jn}\mathbb{Z}_1, 0\}^2 \\
&= \gamma_{jn}^2\{\Phi(\mu_{jn}/\gamma_{jn}) - \Phi(0)\} + \mu_{jn}^2\Phi(\mu_{jn}/\gamma_{jn}) + \mu_{jn}\gamma_{jn}\phi(\mu_{jn}/\gamma_{jn}) \\
&= \{\mu_{jn}\gamma_{jn}\phi(0) + O(h^{d/2})\} + O(h^{d/2}) + \{\mu_{jn}\gamma_{jn}\phi(0) + O(h^d)\} \\
&= 2\phi(0)\delta_{jn}(x)\rho_{jn}(x) + O(h^{d/2}), \text{ uniformly in } x \in \mathcal{S}_j.
\end{aligned}$$

Hence we write $\lim_{n\to\infty}\{\tilde{a}_{jn}-a_{jn}\}$ as

$$\begin{aligned}
&\lim_{n\to\infty}\int_{\mathcal{S}_j}\mathbf{E}[\Lambda_2(h^{-d/4}\rho_{jn}(x)\mathbb{Z}_1+h^{d/4}\delta_{jn}(x))-\Lambda_2(h^{-d/4}\rho_{jn}(x)\mathbb{Z}_1)]w_j(x)dx\\
=\;& 2\phi(0)\lim_{n\to\infty}\int_{\mathcal{S}_j}\delta_{jn}(x)\rho_{jn}(x)w_j(x)dx+O(h^{d/2})\\
=\;& \sqrt{\frac{2}{\pi}}\lim_{n\to\infty}\int_{A_\varepsilon}\delta_{jn}(x)\rho_{jn}(x)w_j(x)dx+\sqrt{\frac{2}{\pi}}\lim_{n\to\infty}\int_{\mathcal{S}_j\backslash A_\varepsilon}\delta_{jn}(x)\rho_{jn}(x)w_j(x)dx+O(h^{d/2}).
\end{aligned}$$

The second term is bounded by $C\varepsilon$ for some $C>0$, because $\delta_{jn}\rho_{jn}$ is bounded. Since the choice of $\varepsilon>0$ was arbitrary and

$$\int_{A_\varepsilon}\delta_{jn}(x)\rho_{jn}(x)w_j(x)dx\to\int_{A_\varepsilon}\delta_j(x)\rho_j(x)w_j(x)dx,\text{ as } n\to\infty,$$

in view of (6.43), this gives the desired result. ■

Proof of Theorem 4*: Let $A_\varepsilon\subset\mathcal{S}_j$ be defined as in the proof of Theorem 4.
(i) Suppose $p=1$. Under H_δ^*, take $\gamma\equiv h^{-d/2}\rho_{jn}(x)$ and $\mu=h^{-d/4}\delta_{jn}(x)$ to get

$$\begin{aligned}
&\mathbf{E}\max\{\gamma\mathbb{Z}_1+\mu,0\}-\mathbf{E}\max\{\gamma\mathbb{Z}_1,0\}\\
=\;& h^{-d/4}\delta_{jn}(x)\Phi(h^{d/4}\delta_{jn}(x)/\rho_{jn}(x))+h^{-d/2}\rho_{jn}(x)\left[\phi(h^{d/4}\delta_{jn}(x)/\rho_{jn}(x))-\phi(0)\right]\\
=\;& h^{-d/4}\delta_{jn}(x)\Phi(0)+\frac{1}{2}\phi(0)\left[\delta_{jn}^2(x)/\rho_{jn}(x)\right]+O(h^{d/4}),
\end{aligned}$$

uniformly in $x\in\mathcal{S}_j$. Therefore, if $\eta_{1,0}(w,\delta)=0$ under H_δ^*, we can write $\lim_{n\to\infty}\{\tilde{a}_{jn}-a_{jn}\}$ as

$$\begin{aligned}
&\lim_{n\to\infty}\int_{\mathcal{X}}\mathbf{E}[\Lambda_1(h^{-d/2}\rho_{jn}(x)\mathbb{Z}_1+h^{-d/4}\delta_{jn}(x))-\Lambda_1(h^{-d/2}\rho_{jn}(x)\mathbb{Z}_1)]w_j(x)dx\\
=\;& \frac{1}{2}\phi(0)\lim_{n\to\infty}\int_{\mathcal{S}_j}\left[\delta_{jn}^2(x)/\rho_{jn}(x)\right]w_j(x)dx\\
=\;& \frac{1}{2}\phi(0)\int_{A_\varepsilon}\left[\delta_j^2(x)/\rho_j(x)\right]w_j(x)dx+\frac{1}{2}\phi(0)\lim_{n\to\infty}\int_{\mathcal{S}_j\backslash A_\varepsilon}\left[\delta_{jn}^2(x)/\rho_{jn}(x)\right]w_j(x)dx.
\end{aligned}$$

Since δ_{jn}^2/ρ_{jn} is uniformly bounded and the choice of $\varepsilon>0$ is arbitrary, we get the desired result.

(ii) Suppose $p = 2$. Under H_δ^*, we take $\gamma \equiv h^{-d/4}\rho_{jn}(x)$ and $\mu = \delta_{jn}(x)$, so that, by a Taylor expansion,

$$\begin{aligned}
& \mathbf{E}\max\{\gamma\mathbb{Z}_1 + \mu, 0\}^2 - \mathbf{E}\max\{\gamma\mathbb{Z}_1, 0\}^2 \\
= \; & \gamma^2\{\Phi(\mu/\gamma) - \Phi(0)\} + \mu^2\Phi(\mu/\gamma) + \mu\gamma\phi(\mu/\gamma) \\
= \; & \left\{\phi(0)\mu\gamma + \frac{1}{6}\phi''(a^*)\frac{\mu^3}{\gamma}\right\} + \left\{\Phi(0)\mu^2 + \phi(a^*)\frac{\mu^3}{\gamma}\right\} + \mu\gamma\left\{\phi(0) + \frac{1}{2}\phi''(a^*)\frac{\mu^3}{\gamma}\right\} \\
= \; & h^{-d/4}\cdot 2\phi(0)\delta_{jn}(x)\rho_{jn}(x) + \frac{1}{2}\delta_{jn}^2(x) + O(h^{d/4}),
\end{aligned}$$

uniformly in $x \in \mathcal{S}_j$, where a^* denotes a term that lies between 0 and μ/γ. Therefore, if $\eta_{1,1}(w,\delta) = 0$ under $H_{2\delta}$, then we can write $\lim_{n\to\infty}\{\tilde{a}_{jn} - a_{jn}\}$ as

$$\begin{aligned}
& \lim_{n\to\infty}\int_{\mathcal{X}} \mathbf{E}[\Lambda_2(h^{-d/2}\rho_{jn}(x)\mathbb{Z}_1 + h^{-d/4}\delta_{jn}(x)) - \Lambda_2(h^{-d/2}\rho_{jn}(x)\mathbb{Z}_1)]w_j(x)dx \\
= \; & \frac{1}{2}\lim_{n\to\infty}\int_{\mathcal{S}_j}\delta_{jn}^2(x)w_j(x)dx \\
= \; & \frac{1}{2}\int_{A_\varepsilon}\delta_j^2(x)w_j(x)dx + \frac{1}{2}\lim_{n\to\infty}\int_{\mathcal{S}_j\backslash A_\varepsilon}\delta_{jn}^2(x)w_j(x)dx.
\end{aligned}$$

Since δ_{jn}^2 is uniformly bounded and the choice of $\varepsilon > 0$ is arbitrary, we get the desired result. ■

Proof of Theorem 5: Similarly as before, we fix $\varepsilon > 0$ and take a compact set $A_\varepsilon \subset \mathcal{S}_j$ such that $\int_{\mathcal{S}_j\backslash A_\varepsilon} w_j(x)dx < \varepsilon$ and $\delta_j(\cdot)$ and $\delta_j(\cdot)\rho_j^{-1}(\cdot)$ are uniformly continuous on A_ε. By change of variables and uniform continuity,

$$\begin{aligned}
\sup_{x\in A_\varepsilon}|\delta_{jn}(x)\rho_{jn}^{-1}(x) - \delta_j(x)\rho_j^{-1}(x)| &\to 0 \text{ and} \\
\sup_{x\in A_\varepsilon}|\delta_{jn}(x) - \delta_j(x)| &\to 0.
\end{aligned}$$

(i) Suppose $p = 1$. For $\gamma > 0$ and $\mu \in \mathbf{R}$,

$$\mathbf{E}\,|\gamma\mathbb{Z}_1 + \mu| = 2\gamma\phi(\mu/\gamma) + 2\mu\,[\Phi(\mu/\gamma) - 1/2]\,.$$

With $\gamma_{jn} \equiv h^{-d/2}\rho_{jn}(x)$ and $\mu_{jn} = h^{-d/4}\delta_{jn}(x)$, we find that uniformly over $x \in \mathcal{S}_j$,

$$\begin{aligned}
& \mathbf{E}|\gamma_{jn}\mathbb{Z}_1 + \mu_{jn}| - \mathbf{E}|\gamma_{jn}\mathbb{Z}_1| \\
= \; & 2\gamma_{jn}[\phi(\mu_{jn}/\gamma_{jn}) - \phi(0)] + 2\mu_{jn}[\Phi(\mu_{jn}/\gamma_{jn}) - 1/2] \\
= \; & [\phi''(0) + 2\phi(0)]\,\delta_{jn}^2(x)\rho_{jn}^{-1}(x) + O(h^{d/4}).
\end{aligned}$$

Therefore, we write $\lim_{n\to\infty}\{\tilde{a}_{jn} - a_{jn}\}$ as

$$\begin{aligned}
&\lim_{n\to\infty}\int_{\mathcal{S}_j}\mathbf{E}[\Lambda_1(h^{-d/2}\rho_{jn}(x)\mathbb{Z}_1 + n^{-d/4}\delta_{jn}(x)) - \Lambda_1(h^{-d/2}\rho_{jn}(x)\mathbb{Z}_1)]w_j(x)dx\\
= \;&\frac{1}{\sqrt{2\pi}}\lim_{n\to\infty}\int_{\mathcal{S}_j}\delta_{jn}^2(x)\rho_{jn}^{-1}(x)w_j(x)dx + O(h^{d/4})\\
= \;&\frac{1}{\sqrt{2\pi}}\int_{A_\varepsilon}\delta_j^2(x)\rho_j^{-1}(x)w_j(x)dx + \frac{1}{\sqrt{2\pi}}\lim_{n\to\infty}\int_{\mathcal{S}_j\setminus A_\varepsilon}\delta_{jn}^2(x)\rho_{jn}^{-1}(x)w_j(x)dx + o(1).
\end{aligned}$$

By Assumption 4 and Lemma A4, $\delta_{jn}^2(x)\rho_{jn}^{-1}(x)$ is bounded uniformly over $x \in \mathcal{S}_j$, enabling us to bound the second integral by $C\varepsilon$ for some $C > 0$. Since ε is arbitrarily chosen, in view of (6.43), this gives the desired result.

(ii) Suppose $p = 2$. We have, for each $x \in \mathcal{S}_j$,

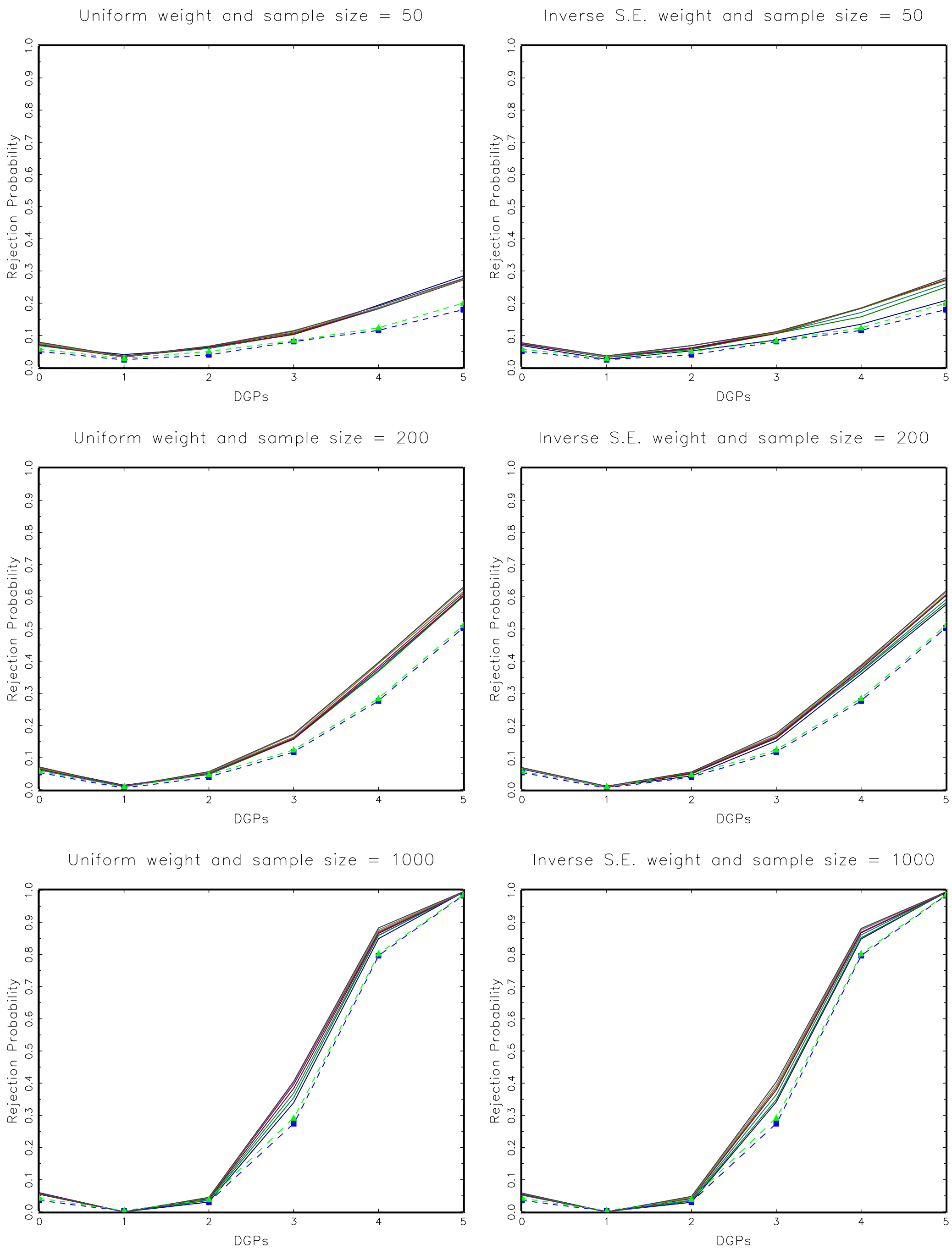
$$\mathbf{E}\{h^{-d/4}\rho_{jn}(x)\mathbb{Z}_1 + \delta_{jn}(x)\}^2 - \mathbf{E}\{h^{-d/4}\rho_{jn}(x)\mathbb{Z}_1\}^2 = \delta_{jn}^2(x).$$

Therefore, we write $\lim_{n\to\infty}\{\tilde{a}_{jn} - a_{jn}\}$ as

$$\begin{aligned}
&\lim_{n\to\infty}\int_{\mathcal{X}}\mathbf{E}[\Lambda_2(h^{-d/4}\rho_{jn}(x)\mathbb{Z}_1 + \delta_{jn}(x)) - \Lambda_2(h^{-d/4}\rho_{jn}(x)\mathbb{Z}_1)]w_j(x)dx\\
= \;&\int_{A_\varepsilon}\delta_j^2(x)w_j(x)dx + \lim_{n\to\infty}\int_{\mathcal{S}_j\setminus A_\varepsilon}\delta_{jn}(x)w_j(x)dx + o(1)
\end{aligned}$$

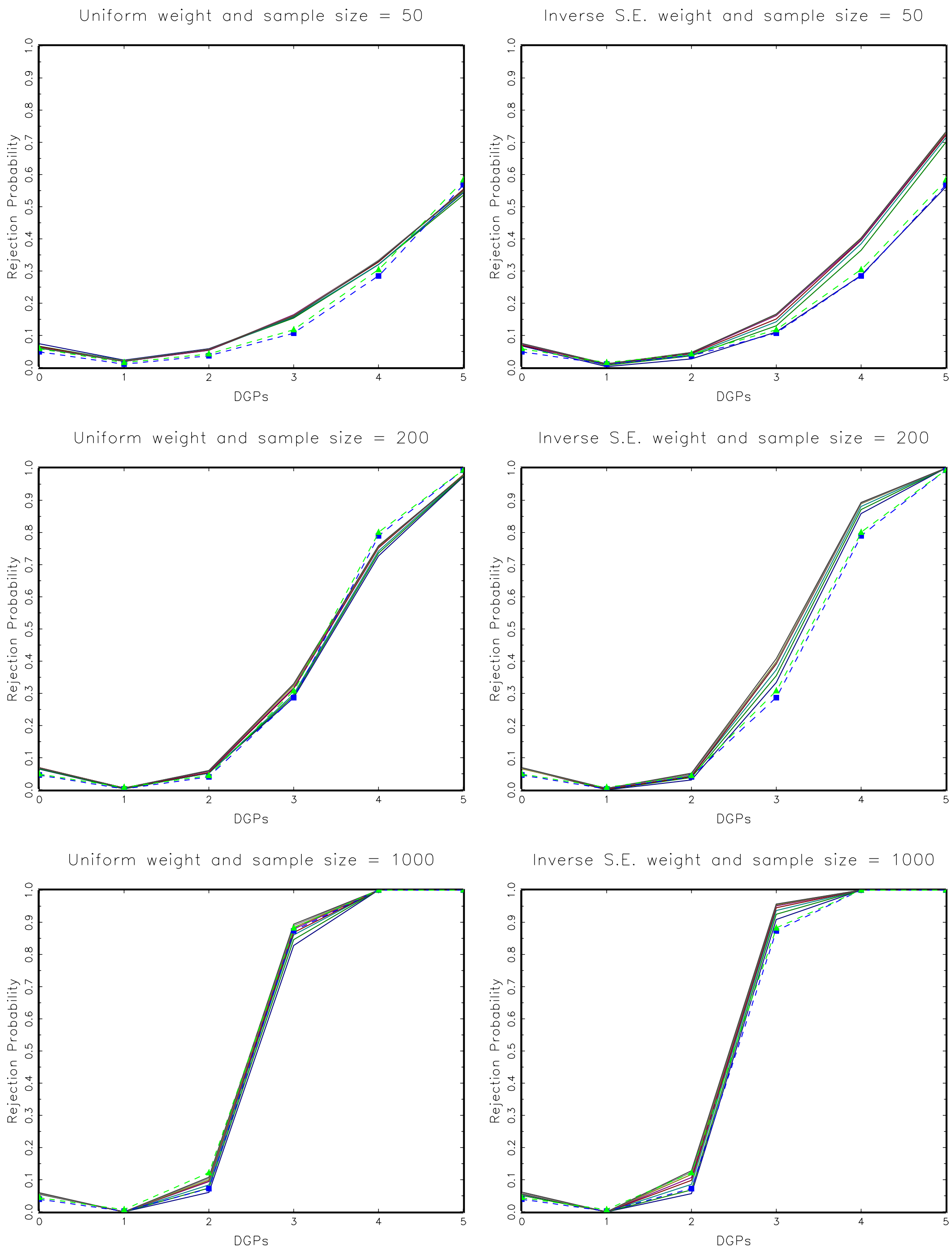
The second integral is bounded by $C\varepsilon$ for some $C > 0$, and in view of (6.43), this gives the desired result. ■

FIGURE 1. Results of Monte Carlo Experiments: L_1 test and $\sigma(x) \equiv 1$



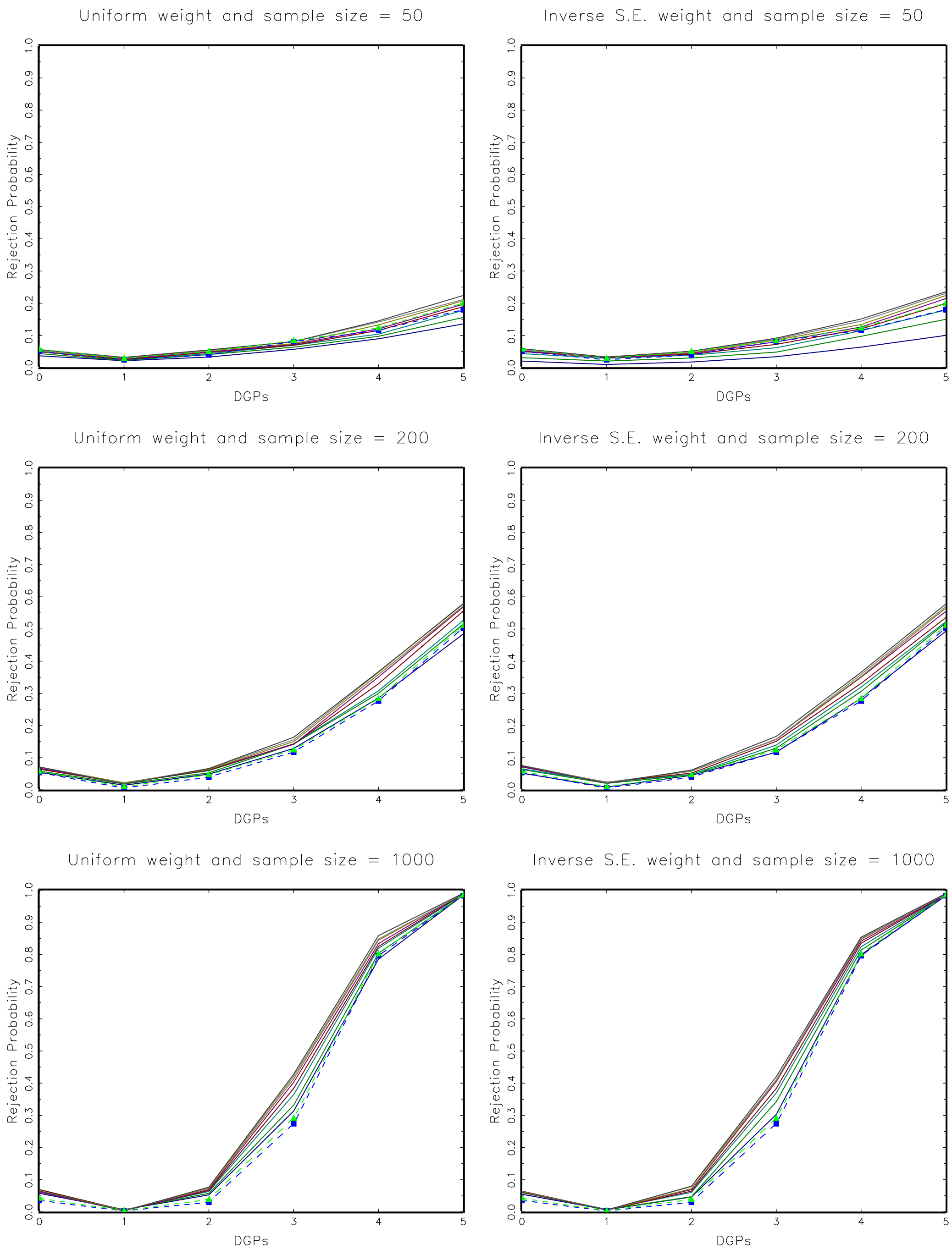
Notes: 8 different solid lines in each panel correspond to our test with 8 different bandwidth values. 2 dotted lines correspond to the test of Andrews and Shi (2011a) with PA and GMS critical values. The nominal level for each test is $\alpha = 0.05$. There are 1000 Monte Carlo replications in each experiment.

FIGURE 2. Results of Monte Carlo Experiments: L_1 test and $\sigma(x) = x$



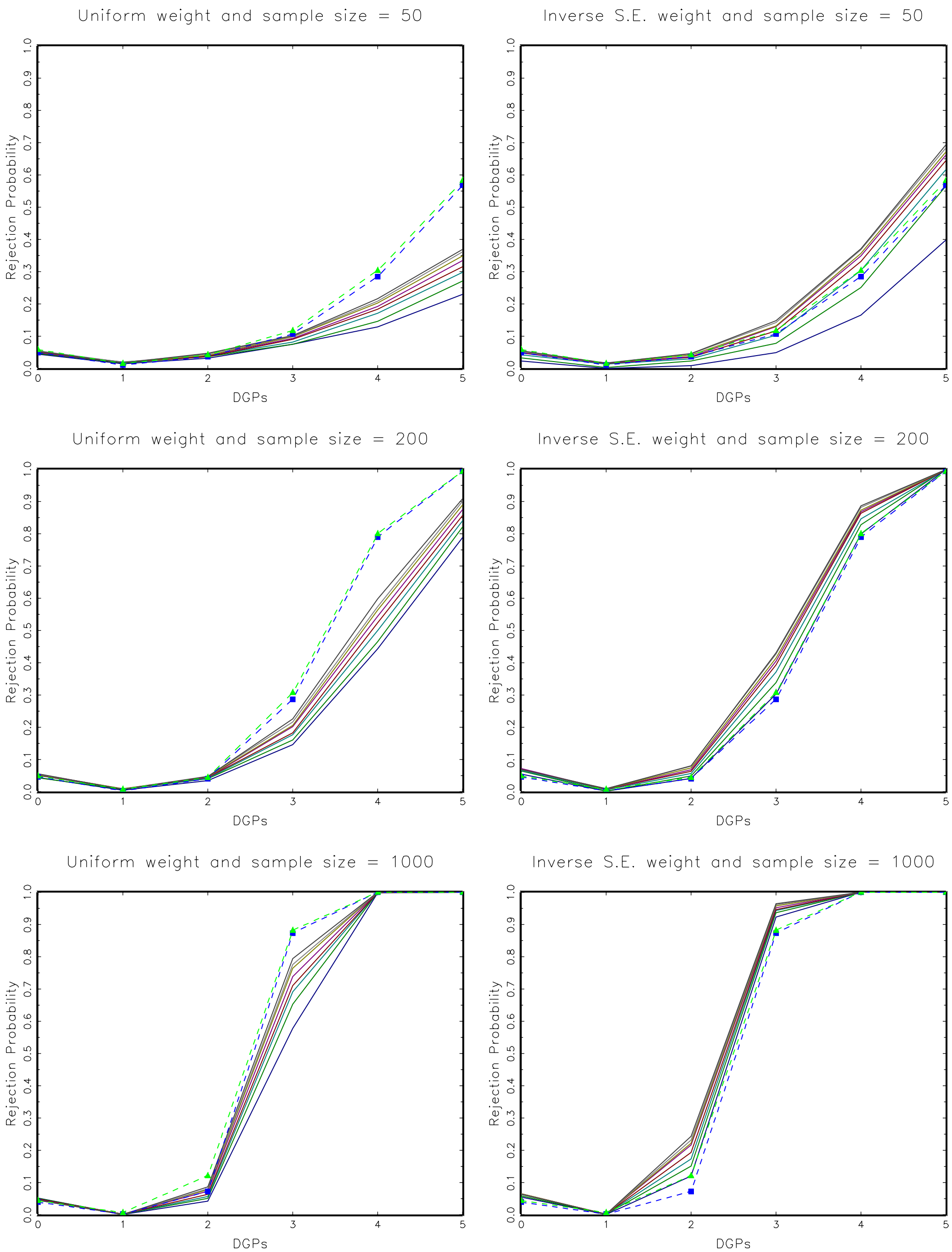
Notes: See notes in Figure 1.

FIGURE 3. Results of Monte Carlo Experiments: L_2 test and $\sigma(x) \equiv 1$



Notes: See notes in Figure 1.

FIGURE 4. Results of Monte Carlo Experiments: L_2 test and $\sigma(x) = x$



Notes: See notes in Figure 1.

References

[1] Anderson, G., O. Linton, and Y.-J. Whang (2012): "Nonparametric estimation and inference about the overlap of two distributions," *Journal of Econometrics*, forthcoming.

[2] Andrews, D. W. K. (2011): "Similar-on-the-boundary tests for moment inequalities exist, but have poor power," *Cowles Foundation Discussion Paper*, No. 1815, available at http://cowles.econ.yale.edu/P/cd/d18a/d1815.pdf.

[3] Andrews, D. W. K. and X. Shi (2011a): "Inference based on conditional moment inequalities," *Cowles Foundation Discussion Paper*, No. 1761R, available at http://cowles.econ.yale.edu/P/cd/d17b/d1761-r.pdf.

[4] Andrews, D. W. K. and X. Shi (2011b): "Nonparametric inference based on conditional moment inequalities," *Cowles Foundation Discussion Paper*, No. 1840, available at http://cowles.econ.yale.edu/P/cd/d18a/d1840.pdf.

[5] Armstrong, T. B. (2011): "Asymptotically Exact Inference in Conditional Moment Inequality Models," Working Paper, Stanford University, available at http://www.stanford.edu/~timothya/.

[6] Bai, J. (2003): "Testing parametric conditional distributions of dynamic models," *Review of Economics and Statistics* 85, 531-549.

[7] Beirlant, J., and D. M. Mason (1995): "On the asymptotic normality of L_p-norms of empirical functionals," *Mathematical Methods of Statistics* 4, 1-19.

[8] Biau, G., B. Cadre, D. M. Mason, and B. Pelletier (2009): "Asymptotic normality in density support estimation," *Electronic Journal of Probability* 14, 2617-2635.

[9] Bickel, P. J. and M. Rosenblatt (1973): "On some global measures of the deviations of density function estimates," *Annals of Statistics* 1, 1071-1095.

[10] Blum, J. R., J. Kiefer, and M. Rosenblatt (1961): "Distribution free tests of independence based on the sample distribution function," *Annals of Mathematical Statistics* 32, 485-498.

[11] Chernozhukov, V., S. Lee, and A. Rosen (2009): "Intersection bounds: estimation and inference," *Cemmap Working Papers*, CWP 19/09, available at http://www.cemmap.ac.uk/wps/cwp1909.pdf.

[12] Denis Chetverikov (2009): "Adaptive test of conditional moment inequalities," *arXiv Working Papers*, arXiv:1201.0167v2, available at http://arxiv.org/abs/1201.0167v2.

[13] Chiappori, P.-A., B. Jullien, B. Salanié, and F. Salanié (2006): "Asymmetric information in insurance: general testable implications," *Rand Journal of Economics* 37, 783-798.

[14] Claeskens, G. and I. van Keilegom (2003): "Bootstrap confidence bands for regression curves and their derivatives," *Annals of Statistics* 31, 1852-1884.

[15] Csörgő, M. and L. Horváth (1988): "Central limit theorems for L_p-norms of density estimators," *Zeitschrits für Wahrscheinlichkeitheorie und verwandte Gebiete* 80, 269-291.

[16] Delgado, M. A. and J. C. Escanciano (2011): "Conditional Stochastic Dominance Testing," Working Paper, Universidad Carlos III de Madrid and Indiana University.

[17] Delgado, M. A. and J. C. Escanciano (2012): "Distribution-free tests of stochastic monotonicity," *Journal of Econometrics*, forthcoming.

[18] Delgado, M. A. and W. González Manteiga (2001): "Significance testing in nonparametric regression based on the bootstrap," *Annals of Statistics* 29, 1469-1507.

[19] Delgado, M. A. and J. Mora (2000): "A nonparametric test for serial independence of regression errors," *Biometrika* 87, 228-234.

[20] Devroye, L. and L. Györfi (1985): *Nonparametric Density Estimation: The L1 View*, Wiley, New York.

[21] DiBenedetto, E. (2001): *Real Analysis*, Birkhäuser, New York.

[22] Durot, C. (2003): "A Kolmogorov-type test for monotonicity of regression, " *Statistics & Probability Letters*, 63, 425-433.

[23] Einav, L., A. Finkelstein, and J. Levin (2010): "Beyond testing: empirical models of insurance markets,"*Annual Review of Economics* 2, 311-336.

[24] Fan, Y. and Q. Li (2000): "Consistent Model Specification Tests: Kernel-Based Tests Versus Bierens' ICM Tests,"*Econometric Theory* 16, 1016-1041.

[25] Gao, J. and I. Gijbels (2008). "Bandwidth selection in nonparametric kernel testing," *Journal of the American Statistical Association* 103(484), 1584-1594.

[26] Ghosal, S., A. Sen and A. W. van der Vaart (2000): "Testing monotonicity of regression," *Annals of Statistics* 28, 1054-1082.

[27] Giné, E., D. M. Mason, and A. Y. Zaitsev (2003): "The L_1-norm density estimator process," *Annals of Probability* 31, 719-768.

[28] Hall, P., Huber,C., and Speckman,P.L. (1997): "Covariate-matched one-sided tests for the difference between functional means, " *J. Amer. Statist. Assoc.* 92, 1074-1083.

[29] Hall, P. and I. Van Keilegom (2005): "Testing for monotone increasing hazard rate, "*Annals of Statistics* 33, 1109-1137.

[30] Hall, P. and Yatchew, A. (2005): "Unified approach to testing functional hypotheses in semiparametric contexts,"*Journal of Econometrics* 127(2), 225-252.

[31] Härdle, W. and E. Mammen (1993): "Comparing nonparametric versus parametric regression fits," *Annals of Statistics* 21, 1926-1947.

[32] Horowitz, J. L. and V. G. Spokoiny (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model against a Nonparametric Alternative, " *Econometrica*, Vol. 69, No. 3, 599-631.

[33] Horváth, L. (1991): "On L_p-norms of multivariate density estimators," *Annals of Statistics* 19, 1933-1949.

[34] Hsu, Y.-C. (2011): "Consistent tests of conditional treatment effects" Working Paper, University of Missouri.

[35] Khmaladze, E. V. (1993): "Goodness of fit problem and scanning innovation martingales," *Annals of Statistics* 21, 798-829.

[36] Khmaladze, E. V. and H. Koul (2004): "Martingale transforms goodness-of-fit tests in regression models," *Annals of Statistics* 32, 995-1034.

[37] Koul, H.L. and Schick,A. (1997): "Testing for the equality of two nonparametric regression curves, " *J. Statist. Plann. Inference* 65, 293-314.

[38] Koul, H.L. and Schick,A. (2003): "Testing for superiority among two regression curves, " *J. Statist. Plann. Inference* 117, 15-33.

[39] Lee, S. and Y.-J. Whang (2009): "Nonparametric tests of conditional treatment effects," *Cemmap Working Papers*, CWP 36/09, available at http://www.cemmap.ac.uk/wps/cwp3609.pdf.

[40] Liero, H., H. Läuter, and V. Konakov (1998): "Nonparametric versus parametric goodness of fit," *Statistics* 31, 115-149.

[41] Manski, C. F. (2003): *Partial Identification of Probability Distributions*, Springer-Verlag, New York.

[42] Mason, D. M. (2009): "Risk bounds for kernel density estimators," *Journal of Mathematical Sciences* 163, 238-261.

[43] Mason, D. M. and W. Polonik (2009): "Asymptotic normality of plug-in level set estimates," *Annals of Applied Probability* 19, 1108-1142.

[44] Pinelis, I. F. (1994): "On a majorization inequality for sums of independent random variables," *Statistics and Probability Letters* 19, 97-99.

[45] Shergin, V. V. (1993): "Central limit theorem for finitely-dependent random variables," *Journal of Mathematical Sciences* 67, 3244-3248.

[46] Song, K. (2009): "Testing conditional independence via Rosenblatt transforms," *Annals of Statistics* 37, 4011-4045.

[47] Stein, E. M. (1970): *Singular Integrals and Differentiability Properties of Functions*, Princeton University Press, Princeton.

[48] Stute, W. (1997): "Nonparametric model checks for regression," *Annals of Statistics* 25, 613-642.

[49] Stute, W., S. Thies, and L. Zhu (1998): "Model checks for regression: an innovation process approach," *Annals of Statistics* 26, 1916-1934.

[50] Sweeting, T. J.(1977): "Speeds of convergence in the multidimensional central limit theorem," *Annals of Probability* 5 28-41.

[51] Tallis, G. M.(1961): "The moment generating function of the truncated multi-normal distribution," *Journal of the Royal Statistical Society. Series B (Methodological)* 23, 223-229.

[52] Tripathi, G. and Y. Kitamura (2003): "Testing conditional moment restrictions," *Annals of Statistics* 31, 2059-2095.

[53] van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*, New York, NY, Springer-Verlag.

Department of Economics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 151-742, Republic of Korea, and Centre for Microdata Methods and Practice, Institute for Fiscal Studies, 7 Ridgmount Street, London, WC1E 7AE, UK.

E-mail address: `sokbae@gmail.com`

Department of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC, V6T 1Z1, Canada

E-mail address: `kysong@mail.ubc.ca`

Department of Economics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 151-742, Republic of Korea.

E-mail address: `whang@snu.ac.kr`