KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE

OF

ECONOMIC RESEARCH

Discussion Paper No.889

"TESTING FOR A GENERAL CLASS OF
FUNCTIONAL INEQUALITIES"

SOKBAE LEE, KYUNGCHUL SONG,
AND YOON-JAE WHANG

February 2014

KYOTO UNIVERSITY

KYOTO, JAPAN

TESTING FOR A GENERAL CLASS OF FUNCTIONAL INEQUALITIES

SOKBAE LEE[1], KYUNGCHUL SONG[2], AND YOON-JAE WHANG[1]

Abstract. In this paper, we propose a general method for testing inequality restrictions on nonparametric functions. Our framework includes many nonparametric testing problems in a unified framework, with a number of possible applications in auction models, game theoretic models, wage inequality, and revealed preferences. Our test involves a one-sided version of L_p functionals of kernel-type estimators ($1 \leq p < \infty$) and is easy to implement in general, mainly due to its recourse to the bootstrap method. The bootstrap procedure is based on nonparametric bootstrap applied to kernel-based test statistics, with estimated "contact sets." We provide regularity conditions under which the bootstrap test is asymptotically valid uniformly over a large class of distributions, including the cases that the limiting distribution of the test statistic is degenerate. Our bootstrap test is shown to exhibit good power properties in Monte Carlo experiments, and we provide a general form of the local power function. As an illustration, we consider testing implications from auction theory, provide primitive conditions for our test, and demonstrate its usefulness by applying our test to real data. We supplement this example with the second empirical illustration in the context of wage inequality.

Key words. Bootstrap, conditional moment inequalities, kernel estimation, local polynomial estimation, L_p norm, nonparametric testing, partial identification, Poissonization, quantile regression, uniform asymptotics

JEL Subject Classification. C12, C14.

[1]Department of Economics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 151-742, Republic of Korea.
[2]Vancouver School of Economics, University of British Columbia, 997 - 1873 East Mall, Vancouver, BC, V6T 1Z1, Canada
E-mail addresses: sokbae@sokbae.ac.kr, kysong@mail.ubc.ca, whang@snu.ac.kr.
Date: 21 February 2014.
We would like to thank Emmanuel Guerre and participants at numerous seminars and conferences for their helpful comments. We also thank Kyeongbae Kim, Koohyun Kwon and Jaewon Lee for capable research assistance. Lee's work was supported by the European Research Council (ERC-2009-StG-240910-ROMETA). Song acknowledges the financial support of Social Sciences and Humanities Research Council of Canada. Whang's work was supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2011-342-B00004). This paper was partly written while Whang was visiting the Institute of Economic Research at Kyoto University, and he thanks them for their hospitality.

1. Introduction

In this paper, we propose a general method for testing inequality restrictions on nonparametric functions. To describe our testing problem, let $v_{\tau,1},\ldots,v_{\tau,J}$ denote nonparametric real-valued functions on $\mathbf{R}^d$ for each index $\tau \in \mathcal{T}$, where $\mathcal{T}$ is a subset of a finite dimensional space. We focus on testing

$$(1.1)\qquad \begin{aligned} &H_0 : \max\{v_{\tau,1}(x),\cdots,v_{\tau,J}(x)\} \le 0 \text{ for all } (x,\tau) \in \mathcal{X}\times\mathcal{T}, \text{ against} \\ &H_1 : \max\{v_{\tau,1}(x),\cdots,v_{\tau,J}(x)\} > 0 \text{ for some } (x,\tau) \in \mathcal{X}\times\mathcal{T}, \end{aligned}$$

where $\mathcal{X}\times\mathcal{T}$ is a domain of interest. We propose a one-sided L_p integrated test statistic based on nonparametric estimators of $v_{\tau,1},\ldots,v_{\tau,J}$. We provide general asymptotic theory for the test statistic and suggest a bootstrap procedure to compute critical values. We establish that our test has correct uniform asymptotic size and is not conservative. We also determine the asymptotic power of our test under fixed alternatives and some local alternatives.

We allow for a general class of nonparametric functions, including, as special cases, conditional mean, quantile, hazard, and distribution functions and their derivatives. For example, $v_{\tau,j}(x) = P(Y_j \le \tau | X = x)$ can be the conditional distribution function of Y_j given $X = x$, or $v_{\tau,j}(x)$ can be the τ-th quantile of Y_j conditional on $X = x$. We can also allow for transformations of these functions satisfying some regularity conditions. The nonparametric estimators we consider are mainly kernel-type estimators but can be allowed to be more general, provided that they satisfy certain Bahadur-type linear expansions.

Inequality restrictions on nonparametric functions arise often as testable implications from economic theory. For example, in first-price auctions, Guerre, Perrigne, and Vuong (2009, GPV hereafter) show that the quantiles of the observed equilibrium bid distributions with different numbers of bidders should satisfy a set of inequality restrictions (Equation (5) of GPV). If the auctions are heterogeneous so that the private values are affected by observed characteristics, we may consider conditionally exogenous participation with a conditional version of the restrictions (see Section 3.2 of GPV). Such restrictions are in the form of multiple inequalities for linear combinations of nonparametric conditional quantile functions. Our test then can be used to test whether the restrictions hold jointly uniformly over quantiles and observed characteristics in a certain range. In this paper, we use this auction example to illustrate the usefulness of our general framework. To the best of our knowledge, there does not exist an alternative test available in the literature for this kind of examples.

In addition to GPV, a large number of auction models are associated with some forms of functional inequalities. See, for example, Haile and Tamer (2003), Haile, Hong, and Shum (2003), Aradillas-López, Gandhi, and Quint (2013a), Aradillas-López, Gandhi, and Quint

(2013b), and Krasnokutskaya, Song, and Tang (2013), among others. Our method can be used to make inference in their setups, while allowing for continuous covariates.

Econometric models of games belong to a related but distinct branch of the literature, compared to the auction models. In this literature, inference on many game theoretic models are recently based on partial identification or functional inequalities. For example, see Tamer (2003), Andrews, Berry, and Jia (2004), Berry and Tamer (2007), Aradillas-López and Tamer (2008), Ciliberto and Tamer (2009), Beresteanu, Molchanov, and Molinari (2011), Galichon and Henry (2011), Chesher and Rosen (2012), and Aradillas-López and Rosen (2013), among others. See de Paula (2013) and references therein for a broad recent development in this literature. Our general method provides researchers in this field a new inference tool when they have continuous covariates.

Inequality restrictions also arise in testing revealed preferences. Blundell, Browning, and Crawford (2008) used revealed preference inequalities to provide the nonparametric bounds on average consumer responses to price changes. In addition, Blundell, Kristensen, and Matzkin (2014) used the same inequalities to bound quantile demand functions. It would be possible to use our framework to test revealed preference inequalities either in average demand functions or in quantile demand functions. See also Hoderlein and Stoye (2013) and Kitamura and Stoye (2013) for related issues of testing revealed preference inequalities.

In addition to the literature mentioned above, many results on partial identification can be written as functional inequalities (see, e.g., Imbens and Manski (2004), Manski (2003), Manski (2007), Manski and Pepper (2000), Tamer (2010), and references therein). In Section 3, we provide a couple of motivating examples of partially identified econometric models (one from Chesher and Rosen (2014) and the other from Khan, Ponomareva, and Tamer (2013)) for which our testing approach can be used to construct confidence regions but to which none of the currently available methods can be applied.

Our framework has several distinctive merits. First, our proposal is easy to implement in general, mainly due to its recourse to the bootstrap method. The bootstrap procedure is based on nonparametric bootstrap applied to kernel-based test statistics. We establish the general asymptotic (uniform) validity of the bootstrap procedure under high level conditions and provide low level conditions for an empirical example based on GPV.

Second, our proposed test is shown to exhibit good power properties both in finite and large samples. Good power properties can be achieved by the use of critical values that adapt to the binding restrictions of functional inequalities. This could be done in various ways; in this paper, we follow the "contact set" approach of Linton, Song, and Whang (2010) and propose bootstrap critical values. As is shown in this paper, the bootstrap critical values yield significant power improvements. Furthermore, we find through our local power analysis that this class of tests exhibit dual convergence rates depending on Pitman directions, and in many cases, the faster of the two rates achieves a parametric rate of $\sqrt{n}$, despite the use of kernel-type test statistics.

Third, we establish the asymptotic validity of the proposed test uniformly over a large class of distributions, without imposing restrictions on the covariance structure among nonparametric estimates of $v_{\tau,j}(\cdot)$, thereby allowing for degenerate cases. Such a uniformity result is crucial for ensuring good finite sample properties for tests whose (pointwise) limiting distribution under the null hypothesis exhibits various forms of discontinuity. The discontinuity in the context of this paper is highly complex, as the null hypothesis involves inequality restrictions on a multiple number of (or even a continuum of) nonparametric functions. We establish the uniform validity of the test in a way that covers these various incidences of discontinuity. Our new uniform asymptotics may be of independent interest in many other contexts.

Much of the recent literature on testing inequality restrictions focuses on conditional moment inequalities.[1] Researches on conditional moment inequalities include Andrews and Shi (2013), Andrews and Shi (2014), Armstrong (2011a), Armstrong (2011b), Armstrong and Chan (2013), Chernozhukov, Lee, and Rosen (2013), Chetverikov (2011), Fan and Park (2014), Khan and Tamer (2009), Kim (2009), Lee, Song, and Whang (2013), Menzel (2009), Ponomareva (2010), among others. In contrast, this paper's approach naturally covers a wide class of inequality restrictions among nonparametric functions that the moment inequality framework does not (or at least is cumbersome to) apply. Such examples include testing multiple inequalities that are defined by differences in conditional quantile functions uniformly over covariates and quantiles.[2] If we restrict our attention to the conditional moment inequalities, then our approach is mostly comparable to the moment selection approach of Andrews and Shi (2013). Our general framework is also related to testing qualitative nonparametric hypotheses such as monotonicity in mean regression. See, for example, Baraud, Huet, and Laurent (2005), Chetverikov (2012), Dümbgen and Spokoiny (2001), and Ghosal, Sen, and van der Vaart (2000) among many others. See also Lee, Linton, and Whang (2009) and Delgado and Escanciano (2012) for testing stochastic monotonicity.

Among aforementioned papers, Chernozhukov, Lee, and Rosen (2013) developed a sup-norm approach in testing inequality restrictions on nonparametric functions using pointwise asymptotics, and in principle, could be extended to test general functional inequalities as

[1]There exists large literature on inference on models with a finite number of unconditional moment inequality restrictions. Some examples include Andrews and Barwick (2012), Andrews and Guggenberger (2009), Andrews and Soares (2010), Beresteanu and Molinari (2008), Bugni (2010), Canay (2010), Chernozhukov, Hong, and Tamer (2007), Galichon and Henry (2009), Romano and Shaikh (2008), Romano and Shaikh (2010), and Rosen (2008), among others.

[2]A working paper version (Andrews and Shi 2009) of Andrews and Shi (2013) covers testing moment inequalities indexed by $\tau \in \mathcal{T}$, but their framework does not appear to be easily extendable to deal with functions of multiple conditional quantiles such as differences in conditional quantiles.

in (1.1).[3] Example 4 of Chernozhukov, Lee, and Rosen (2013) considered the case of one inequality with a conditional quantile function at a particular quantile, but it is far from trivial to extend this example to multiple inequalities of differences in conditional quantile functions uniformly over a range of quantiles. As this paper demonstrates through empirical applications, such testing problems arise frequently in the fields of industrial organization and labor economics (see Sections 3.3 and 3.4).

The uniformity result in this paper is non-standard since our test is based on asymptotically non-tight processes, in contrast to Andrews and Shi (2013) who convert conditional moment inequalities into an infinite number of unconditional moment inequalities. This paper's development of asymptotic theory draws on the method of Poissonization (see, e.g., Horváth (1991) and Giné, Mason, and Zaitsev (2003)). For applications of this method, see Anderson, Linton, and Whang (2012) for inference on a polarization measure, Lee and Whang (2009) for testing for conditional treatment effects, and Lee, Song, and Whang (2013) for testing inequalities for nonparametric regression functions using the numerator of the Nadaraya-Watson estimator (based on pointwise asymptotics). Also, see Mason and Polonik (2009) and Biau, Cadre, Mason, and Pelletier (2009) for support estimation.

The remainder of the paper is as follows. Section 2 gives an informal description of our general framework by introducing test statistics and critical values and by providing intuitions behind our approach. In Section 3, we present four motivating examples that include two examples of partially identified models and two empirical examples to demonstrate the usefulness of our test. The first empirical example is based on GPV and the second one is about testing functional inequalities in the context of wage inequality, inspired by Acemoglu and Autor (2011). In Section 4, we establish the uniform asymptotic validity of our bootstrap test using high-level conditions. We also provide a class of distributions for which the asymptotic size is exact. In Section 5, we give primitive conditions for the uniform asymptotic validity of our inference method for the first empirical example in Section 3. In Section 6, we establish consistency of our test and its local power properties. Section 7 concludes. Appendices consist of two parts. The first part presents results of Monte Carlo experiments and more examples of testing functional inequalities that include an alternative statistic for the first empirical example and testing monotonicity with respect to a covariate

[3]Our test involves a one-sided version of L_p-type functionals of nonparametric estimators ($1 \leq p < \infty$). We regard the sup-norm and L_p norm approaches complementary, each with its own strength and weakness. For example, our test and also the test of Andrews and Shi (2013) have higher power against relatively flat alternatives, whereas the test of Chernozhukov, Lee, and Rosen (2013) has higher power against sharply-peaked alternatives. See the results of Monte Carlo experiments reported in Appendix I. See also Andrews and Shi (2013), Andrews and Shi (2014), and Chernozhukov, Lee, and Rosen (2013) for related discussions and further Monte Carlo evidence.

in conditional expectation, cumulative distribution, and quantile functions. The remaining part provides all the proofs of theorems.

2. General Overview

2.1. Test Statistics. We present a general overview of this paper's framework by introducing test statistics and critical values. To ease the exposition, we confine our attention to the case of $J = 2$ here. The definitions and formal results for general J are given later in Section 4.

Throughout this paper, we assume that $\mathcal{T}$ is a connected compact subset of a Euclidean space. This does not lose much generality because when $\mathcal{T}$ is a finite set, we can redefine our test statistic by taking $\mathcal{T}$ as part of the finite index j indexing the nonparametric functions.

For $j = 1, 2$, let $\hat{v}_{\tau,j}(x)$ be a kernel-based nonparametric estimator of $v_{\tau,j}(x)$ and let its appropriately scaled version be

$$\hat{u}_{\tau,j}(x) \equiv \frac{r_{n,j}\hat{v}_{\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)},$$

where $r_{n,j}$ is an appropriate normalizing sequence that diverges to infinity,[4] and $\hat{\sigma}_{\tau,j}(x)$ is an appropriate (possibly data-dependent) scale normalization.[5] Then the inference is based on the following statistic:

$$\begin{aligned} \hat{\theta} &\equiv \int_{\mathcal{T}}\int_{\mathcal{X}} \max\left\{\hat{u}_{\tau,1}(x), \hat{u}_{\tau,2}(x), 0\right\}^p dx d\tau \\ &\equiv \int_{\mathcal{X}\times\mathcal{T}} \max\left\{\hat{u}_{\tau,1}(x), \hat{u}_{\tau,2}(x), 0\right\}^p dQ(x,\tau), \end{aligned} \tag{2.1}$$

where Q is Lebesgue measure on $\mathcal{X} \times \mathcal{T}$. In this overview section, we focus on the case of using the max function under the integral in (2.1). In addition, we consider the sum $\sum_{j=1}^{2} \max\left\{\hat{u}_{\tau,j}(x), 0\right\}^p$ in one of our empirical examples (see Section 3.3).

2.2. Bootstrap Critical Values. As we shall see later, the asymptotic distribution of the test statistic exhibits complex ways of discontinuities as one perturbs the data generating processes. This suggests that the finite sample properties of the asymptotic critical values may not be stable. Furthermore, the location-scale normalization requires nonparametric estimation and thus a further choice of tuning parameters. This can worsen the finite sample properties of the critical values further. To address these issues, this paper develops a bootstrap procedure.

[4]Permitting the convergence rate $r_{n,j}$ to differ across $j \in \mathbb{N}_J$ can be convenient, when the nonparametric estimators have different convergence rates. For example, this accommodates a situation where one jointly tests the non-negativity and monotonicity of a nonparametric function.

[5]While our framework permits the case where $\hat{\sigma}_{\tau,j}(x)$ is simply chosen to be 1, we allow for a more general case where $\hat{\sigma}_{\tau,j}(x)$ is a consistent estimator for some nonparametric quantity.

As we shall show formally in a more general form in Lemma 1 in Section 4 below, it is satisfied that under H_0, for each sequence $c_n \to \infty$ such that $\sqrt{\log n}/c_n \to 0$ as $n \to \infty$,

$$
\begin{aligned}
(2.2) \qquad \hat{\theta} &= \int_{B_{n,\{1\}}(c_n)} \max\left\{\hat{u}_{\tau,1}(x), 0\right\}^p dQ(x,\tau) \\
&\quad + \int_{B_{n,\{2\}}(c_n)} \max\left\{\hat{u}_{\tau,2}(x), 0\right\}^p dQ(x,\tau) \\
&\quad + \int_{B_{n,\{1,2\}}(c_n)} \max\left\{\hat{u}_{\tau,1}(x), \hat{u}_{\tau,2}(x), 0\right\}^p dQ(x,\tau),
\end{aligned}
$$

with probability approaching one, where, letting $u_{n,\tau,j}(x) \equiv r_{n,j} v_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x)$, i.e., a population version of $\hat{u}_{\tau,j}(x)$,[6] we define

$$
\begin{aligned}
B_{n,\{1\}}(c_n) &\equiv \left\{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |u_{n,\tau,1}(x)| \le c_n \text{ and } u_{n,\tau,2}(x) < -c_n\right\}, \\
B_{n,\{2\}}(c_n) &\equiv \left\{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |u_{n,\tau,2}(x)| \le c_n \text{ and } u_{n,\tau,1}(x) < -c_n\right\} \text{ and} \\
B_{n,\{1,2\}}(c_n) &\equiv \left\{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |u_{n,\tau,1}(x)| \le c_n \text{ and } |u_{n,\tau,2}(x)| \le c_n\right\}.
\end{aligned}
$$

For example, the set $B_{n,\{1\}}(c_n)$ is a set of points (x,τ) such that $|v_{n,\tau,1}(x)/\sigma_{n,\tau,1}(x)|$ is close to zero, and $v_{n,\tau,2}(x)/\sigma_{n,\tau,2}(x)$ is negative and away from zero. We call *contact sets* such sets as $B_{n,\{1\}}(c_n)$, $B_{n,\{2\}}(c_n)$, and $B_{n,\{1,2\}}(c_n)$.

Now, comparing (2.2) with (2.1) reveals that the limiting distribution of $\hat{\theta}$ under the null hypothesis will not depend on points outside the union of the contact sets. Thus the main idea of this paper is to base the bootstrap critical values on the quantity on the right hand side of (2.2) instead of that on the last integral in (2.1). As we will explain shortly in the next subsection, this leads to a test that is uniformly valid and exhibits substantial improvement in power.

To construct bootstrap critical values, we introduce sample versions of the contact sets:

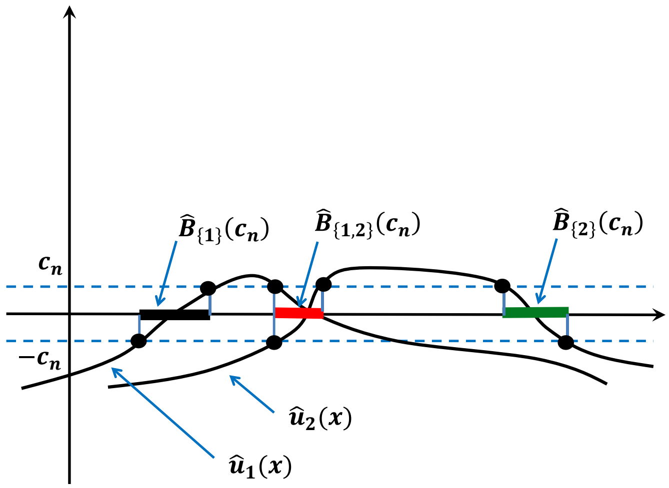
$$
\begin{aligned}
\hat{B}_{\{1\}}(c_n) &\equiv \left\{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |\hat{u}_{\tau,1}(x)| \le c_n \text{ and } \hat{u}_{\tau,2}(x) < -c_n\right\}, \\
\hat{B}_{\{2\}}(c_n) &\equiv \left\{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |\hat{u}_{\tau,2}(x)| \le c_n \text{ and } \hat{u}_{\tau,1}(x) < -c_n\right\} \text{ and} \\
\hat{B}_{\{1,2\}}(c_n) &\equiv \left\{(x,\tau) \in \mathcal{X} \times \mathcal{T} : |\hat{u}_{\tau,1}(x)| \le c_n \text{ and } |\hat{u}_{\tau,2}(x)| \le c_n\right\}.
\end{aligned}
$$

See Figure 1 for illustration of estimation of contact sets when $J = 2$.

Given the contact sets, we construct bootstrap critical values as follows. Let $\hat{v}^*_{\tau,j}(x)$ and $\hat{\sigma}^*_{\tau,j}(x)$, $j = 1, 2$, denote the bootstrap counterparts of $\hat{v}_{\tau,j}(x)$ and $\hat{\sigma}_{\tau,j}(x)$, $j = 1, 2$. Let the bootstrap counterparts be constructed in the same way as the nonparametric estimators

[6] It is convenient for general development to let the population quantities $v_{n,\tau,j}(x)$ and $\sigma_{n,\tau,j}(x)$ depend on n.

Figure 1. Contact Set Estimation



Note: This figure illustrates estimated contact sets when $J = 2$. The black, red, and green line segments on the x-axis represent estimated contact sets.

$\hat{v}_{\tau,j}(x)$ and $\hat{\sigma}_{\tau,j}(x)$, $j = 1, 2$, with the bootstrap sample independently drawn with replacement from the empirical distribution of the original sample. We let

$$\hat{s}^*_{\tau,j}(x) \equiv \frac{r_{n,j}\{\hat{v}^*_{\tau,j}(x) - \hat{v}_{\tau,j}(x)\}}{\hat{\sigma}^*_{\tau,j}(x)}, \; j = 1, 2. \tag{2.3}$$

Note that $\hat{s}^*_{\tau,j}(x)$ is a centered and scale normalized version of the bootstrap quantity $\hat{v}^*_{\tau,j}(x)$. We construct a bootstrap version of the right hand side of (2.2) as

$$\begin{aligned} \hat{\theta}^* \equiv & \int_{\hat{B}_{\{1\}}(\hat{c}_n)} \max\left\{\hat{s}^*_{\tau,1}(x), 0\right\}^p dQ(x,\tau) \\ & + \int_{\hat{B}_{\{2\}}(\hat{c}_n)} \max\left\{\hat{s}^*_{\tau,2}(x), 0\right\}^p dQ(x,\tau) \\ & + \int_{\hat{B}_{\{1,2\}}(\hat{c}_n)} \max\left\{\hat{s}^*_{\tau,1}(x), \hat{s}^*_{\tau,2}(x), 0\right\}^p dQ(x,\tau), \end{aligned} \tag{2.4}$$

where $\hat{c}_n$ is a data dependent version of c_n. We will discuss a way to construct $\hat{c}_n$ shortly. We also define

$$\hat{a}^* \equiv \mathbf{E}^* \hat{\theta}^*,$$

where $\mathbf{E}^*$ denotes the expectation under the bootstrap distribution. Let c^*_α be the $(1-\alpha)$-th quantile from the bootstrap distribution of $\hat{\theta}^*$. Then for a small $\eta > 0$ such as $\eta = 10^{-6}$, we take $c^*_{\alpha,\eta} \equiv \max\{c^*_\alpha, h^{d/2}\eta + \hat{a}^*\}$ as the critical value to form the following test:

$$\text{Reject } H_0 \text{ if and only if } \hat{\theta} > c^*_{\alpha,\eta}. \tag{2.5}$$

Then it is shown later that the test has asymptotically correct size, i.e.,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_0} P\{\hat{\theta} > c^*_{\alpha,\eta}\} \leq \alpha, \tag{2.6}$$

where $\mathcal{P}_0$ is the collection of potential distributions that satisfy the null hypothesis.

2.3. **Obtaining tuning parameters.** To construct $\hat{c}_n$, we suggest the following procedure. First, define

$$S^*_n \equiv \max\left\{\sup_{(j,\tau,x)} \hat{s}^*_{\tau,j}(x), \varepsilon\sqrt{\log n}\right\},$$

where $\varepsilon > 0$ is a small number. Then, set

$$\hat{c}_n = C_{\rm cs}(\log\log n) q_{1-\alpha_n}(S^*_n), \tag{2.7}$$

where $q_{1-\alpha_n}(S^*_n)$ is the $(1-\alpha_n)$-th quantile of the bootstrap distribution of S^*_n with $\alpha_n = 0.1/\log n$, and $C_{\rm cs}$ is a "sensitivity" constant that needs to be chosen by a researcher. Although the rule-of-thumb for c_n in (2.7) is not completely data-driven, it has the advantage that the scale of $\hat{u}_{\tau,j}(x)$ is invariant, due to the term $q_{1-\alpha_n}(S^*_n)$; see Chernozhukov, Lee, and Rosen (2013) for a similar idea.[7] This data-dependent choice of $\hat{c}_n$ is encompassed by the theoretical framework of this paper, while many other choices are also admitted.[8]

To implement our bootstrap test, it is necessary to fix three constants: η, ε, and $C_{\rm cs}$, in addition to the bandwidth used in kernel-based nonparametric estimation. Based on our experiences in Monte Carlo experiments, we suggest the following rule-of-thumb: set η and ε to be small numbers, say $\eta = \epsilon = 10^{-6}$ and check sensitivity with respect to $C_{\rm cs}$ by varying it over a certain range. In particular, we recommend taking $C_{\rm cs} = 0.5$ and performing sensitivity check by increasing the value of $C_{\rm cs}$ up to 1.5.[9]

Regarding the bandwidth selection, we suggest the following rule. First, choose a bandwidth, say $\tilde{h}$, using a readily available bandwidth selection rule that is typically designed

[7]Note that $q_{1-\alpha_n}(S^*_n)$ is the $(1-\alpha_n)$ quantile of the supremum of $\hat{s}^*_{\tau,j}(x)$ over (j,τ,x) for a sufficiently small ε, provided that $\hat{s}^*_{\tau,j}(x)$ is non-degenerate. Note that $(1-\alpha_n)$ converges to 1 as n gets large. Thus, this observation leads to the choice of $\hat{c}_n$ in (2.7) that is proportional to $q_{1-\alpha_n}(S^*_n)$ times a very slowing growing term such as $\log\log n$, to insure that $\hat{c}_n$ diverges to infinity but as slowly as possible, while having the property of scale invariance.

[8]See Assumption A4(ii) below for sufficient conditions for a data dependent choice of $\hat{c}_n$. It is not hard to see that the conditions are satisfied, once the uniform convergence rates of $\hat{v}_{\tau,j}(x)$ and $\hat{\sigma}_{\tau,j}(x)$ and their bootstrap versions hold as required in Assumptions A3, A5, and B2 and B3.

[9]The rationale behind this particular recommendation is that in Monte Carlo experiments reported in Appendix I, our test performed well with $C_{\rm cs} = 0.5$ and we would like to be on the more conservative side when we check the sensitivity to $C_{\rm cs}$.

for the purpose of optimal estimation (e.g. see Fan and Gijbels (1996) for local polynomial estimators). When $d = 1$ and the underlying function is twice continuously differentiable, the bandwidth has the form $\tilde{h} = Cn^{-1/5}$ with some constant C. Second, if necessary, modify $\tilde{h}$ so that it satisfies the regularity conditions imposed in this paper. For example, in case of estimating conditional quantile functions, Assumption AUC-3 in Section 5 is satisfied by the choice of $h = n^{-s}$ with the condition $1/4 < s < 1/3$ if the local linear estimator is used with $d = 1$. Then we can take $h = \tilde{h} \times n^{1/5} \times n^{-s}$ for some s satisfying $1/4 < s < 1/3$.

2.4. **Discontinuity, Uniformity, and Power.** Many tests of inequality restrictions exhibit discontinuity in its limiting distribution under the null hypothesis. When the inequality restrictions involve nonparametric functions, this discontinuity takes a complex form, as emphasized in Section 5 of Andrews and Shi (2013).

To see the discontinuity problem in our context, let $\{(Y_i, X_i)^\top\}_{i=1}^n$ be i.i.d. copies from an observable bivariate random vector, $(Y, X)^\top \in \mathbf{R} \times \mathbf{R}$, where X_i is a continuous random variable with density f. We consider a simple testing example:

$$H_0 : \mathbf{E}[Y|X = x] \leq 0 \text{ for all } x \in \mathcal{X} \quad \text{vs.} \quad H_1 : \mathbf{E}[Y|X = x] > 0 \text{ for some } x \in \mathcal{X}. \tag{2.8}$$

Here, with the subscript τ suppressed, we set $J = 1$, $r_{n,1} = \sqrt{nh}$, $p = d = 1$, and define $[v]_+ \equiv \max\{v, 0\}$. Let

$$\hat{v}_1(x) = \frac{1}{nh}\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right) \text{ and } \hat{\sigma}_1^2(x) = \frac{1}{nh}\sum_{i=1}^n Y_i^2 K^2\left(\frac{X_i - x}{h}\right), \tag{2.9}$$

where K is a nonnegative, univariate kernel function with compact support and h is a bandwidth.

Assume that the density of X is strictly positive on $\mathcal{X}$. Then, in this example, $v_{n,1}(x) \equiv \mathbf{E}\hat{v}_1(x) \leq 0$ for almost every x in $\mathcal{X}$ whenever the null hypothesis is true. Define

$$Z_{n,1}(x) = \sqrt{nh}\left\{\frac{\hat{v}_1(x) - v_{n,1}(x)}{\hat{\sigma}_1(x)}\right\} \text{ and } B_{n,1}(0) = \left\{x \in \mathcal{X} : \left|\sqrt{nh}v_{n,1}(x)\right| = 0\right\}.$$

We analyze the asymptotic properties of $\hat{\theta}$ as follows. We first write

$$\begin{aligned} h^{-1/2}(\hat{\theta} - a_{n,1}) &= h^{-1/2}\left\{\int_{B_{n,1}(0)} [Z_{n,1}(x)]_+ \, dx - a_{n,1}\right\} \\ &\quad + h^{-1/2}\int_{\mathcal{X}\setminus B_{n,1}(0)} \left[Z_{n,1}(x) + \frac{\sqrt{nh}v_{n,1}(x)}{\hat{\sigma}_1(x)}\right]_+ dx, \end{aligned} \tag{2.10}$$

where

$$a_{n,1} = \mathbf{E}\left[\int_{B_{n,1}(0)} [Z_{n,1}(x)]_+ \, dx\right].$$

When $\liminf_{n\to\infty} Q\left(B_{n,1}(0)\right) > 0$ with $Q(B_{n,1}(0))$ denoting Lebesgue measure of $B_{n,1}(0)$, we can show that the leading term on the right hand side in (2.10) becomes asymptotically $N(0, \sigma_0^2)$ for some $\sigma_0^2 > 0$. On the other hand, the second term vanishes in probability as $n \to \infty$ under H_0 because for each $x \in \mathcal{X} \backslash B_{n,1}(0)$,

$$0 > \sqrt{nh} v_{n,1}(x) \to -\infty$$

as $n \to \infty$ under H_0. Thus we conclude that when $\liminf_{n\to\infty} Q\left(B_{n,1}(0)\right) > 0$ under H_0,

$$h^{-1/2}(\hat{\theta} - a_{n,1}) \approx h^{-1/2} \left\{ \int_{B_{n,1}(0)} \left[Z_{n,1}(x)\right]_+ dx - a_{n,1} \right\} \to_d N(0, \sigma_0^2). \tag{2.11}$$

This asymptotic theory is pointwise in P (with P fixed and letting $n \to \infty$), and may not be adequate for finite sample approximation. There are two sources of discontinuity. First, the pointwise asymptotic theory essentially regards the drift component $\sqrt{nh} v_{n,1}(x)$ as $-\infty$, whereas in finite samples, the component can be very negative, but not $-\infty$. Second, even if the nonparametric function $\sqrt{nh} v_{n,1}(x)$ changes continuously, the contact set $B_{n,1}(0)$ may change discontinuously in response.[10] While there is no discontinuity in the finite sample distribution of the test statistic, there may arise discontinuity in its pointwise asymptotic distribution. Furthermore, the complexity of the discontinuity makes it harder to trace its source, when we have $J > 2$. As a result, the asymptotic validity of the test that is established pointwise in P is not a good justification of the test. We need to establish the asymptotic validity that is *uniform* in P over a reasonable class of probabilities.

Under regularity conditions, bootstrap critical values based on the least favorable configuration (LFC) such that

$$\hat{\theta}^*_{\text{LFC}} \equiv \int_{\mathcal{X}} \left[\hat{\mathbf{s}}^*(x)\right]_+ dx, \text{ where } \hat{\mathbf{s}}^*(x) = \sqrt{nh} \left\{ \frac{\hat{v}_1^*(x) - \hat{v}_1(x)}{\hat{\sigma}_1^*(x)} \right\}, \tag{2.12}$$

can be shown to yield tests that are asymptotically valid *uniformly* in P. However, they are often too conservative in practice. Using a critical value based on

$$\hat{\theta}_1^* \equiv \int_{\hat{B}_{\{1\}}(c_n)} \left[\hat{\mathbf{s}}^*(x)\right]_+ dx$$

also yields an asymptotically valid test, and yet $\hat{\theta}^*_{\text{LFC}} > \hat{\theta}_1^*$ in general. Thus the bootstrap tests that use the contact set have better power properties than those that do not. The power

[10]For example, take $\sqrt{nh} v_{n,1}(x) = -x^2/n$ on $\mathcal{X} = [-1, 1]$. Let $v_0(x) \equiv 0$. Then $\sqrt{nh} v_{n,1}(x)$ goes to $v_0(x)$ uniformly in $x \in \mathcal{X}$ as $n \to \infty$. However, for each n, $B_{n,1}(0) = \{x \in \mathcal{X} : \sqrt{nh} v_{n,1}(x) = 0\} = \{0\}$, which does not converge in Hausdorff distance to $B_1(0) \equiv \{x \in \mathcal{X} : v_0(x) = 0\} = \mathcal{X}$.

improvement is substantial in many simulation designs and can be important in real-data applications.[11]

Now, let us see how the choice of $c^*_{\alpha,\eta} \equiv \max\{c^*_\alpha, h^{1/2}\eta + \hat{a}^*\}$ (with $d = 1$ here) leads to bootstrap inference that is valid even when the test statistic becomes degenerate under the null hypothesis. The degeneracy arises when the inequality restrictions hold with large slackness, so that the convergence in (2.11) holds with $\sigma_0^2 = 0$, and hence

$$h^{-1/2}(\hat{\theta} - a_{n,1}) = o_P(1).$$

For the bootstrap counterpart, note that

$$\begin{aligned} h^{-1/2}(c^*_{\alpha,\eta} - a_{n,1}) &= h^{-1/2}\max\{c^*_\alpha - a_{n,1}, h^{1/2}\eta + \hat{a}^* - a_{n,1}\} \\ &\geq \eta + h^{-1/2}(\hat{a}^* - a_{n,1}), \end{aligned}$$

where it can be shown that $h^{-1/2}(\hat{a}^* - a_{n,1}) = o_P(1)$. Therefore, the bootstrap inference is designed to be asymptotically valid even when the test statistic becomes degenerate.

Note that for the sake of validity only, one may replace $h^{1/2}\eta$ by a fixed constant, say $\bar{\eta} > 0$. However, this choice would render the test asymptotically too conservative. The choice of $h^{1/2}\eta$ in this paper makes the test asymptotically exact for a wide class of probabilities, while preserving the uniform validity in both the cases of degeneracy and nondegeneracy.[12] The precise class of probabilities under which the test becomes asymptotically exact is presented in Section 4.

There are two remarkable aspects of the local power behavior of our bootstrap test. First, the test exhibits two different kinds of convergence rates along different directions of Pitman local alternatives. Second, despite the fact that the test uses the approach of local smoothing by kernel as in Härdle and Mammen (1993), the faster of the two convergence rates achieves a parametric rate of $\sqrt{n}$. To see this more closely, let us return to the simple example in (2.8), and consider the following local alternatives:

$$v_n(x) = v_0(x) + \frac{\delta(x)}{b_n}, \tag{2.13}$$

where $v_0(x) \leq 0$ for all $x \in \mathcal{X}$ and $\delta(x) > 0$ for some $x \in \mathcal{X}$, and $b_n \to \infty$ as $n \to \infty$ such that $v_n(x) > 0$ for some $x \in \mathcal{X}$. The function $\delta(\cdot)$ represents a Pitman direction of the local alternatives. Suppose that the test has nontrivial local power against local alternatives of

[11] There may exist an alternative approach to improve the power of our test. Romano, Shaikh, and Wolf (2013) proposed a computationally attractive two-step method for testing a finite number of unconditional moment inequalities. It is an interesting topic to extend their two-step approach to our setup, but it is beyond the scope of this paper.

[12] Our fixed positive constant η plays a role similar to a fixed constant in Andrews and Shi (2013)'s modification of the sample variance-covariance matrix of unconditional moment conditions, transformed by instruments (ε in their notation in equation (3.5) of Andrews and Shi (2013)).

the form in (2.13), but trivial power whenever b_n in (2.13) is replaced by b'_n that diverges faster than b_n. In this case, we say that the test has convergence rate equal to b_n against the Pitman direction δ.

As we show later, there exist two types of convergence rates of our test, depending on the choice of $\delta(x)$. Let $B^0(0) \equiv \{x \in \mathcal{X} : v_0(x) = 0\}$ and $\sigma_1^2(x) \equiv \mathbf{E}[Y_i^2|X_i = x]f(x)\int K^2(u)du$. When $\delta(\cdot)$ is such that

$$\int_{B^0(0)} \frac{\delta(x)}{\sigma_1(x)}dx > 0,$$

the test achieves a parametric rate $b_n = \sqrt{n}$. On the other hand, when $\delta(\cdot)$ is such that

$$\int_{B^0(0)} \frac{\delta(x)}{\sigma_1(x)}dx = 0 \text{ and } \int_{B^0(0)} \frac{\delta^2(x)}{\sigma_1^2(x)}dx > 0,$$

the test achieves a slower rate $b_n = \sqrt{n}h^{1/4}$. See Section 6.2 for heuristics behind the results. In Section 6.3, the general form of local power functions is derived.

3. Motivating Examples

In this section, we first provide two examples of partially identified econometric models for which our testing approach can be used to construct confidence regions. One example is based on generalized instrumental variables models of Chesher and Rosen (2014), and the other is from a panel data model of Khan, Ponomareva, and Tamer (2013). In addition, we give two empirical examples. The first empirical example is on testing auction models following GPV, and the second one is about testing functional inequalities via differences-in-differences in conditional quantiles, inspired by Acemoglu and Autor (2011). All four examples given in this section are not covered easily by existing inference methods, when continuous covariates exist; however, they are all special cases of our general framework.

Appendix II gives more examples of testing problems that can be included in our general framework. In particular, these additional examples include new methods for testing monotonicity with respect to a covariate by constructing one-sided L_p-type functionals in a suitable fashion in three examples: one in mean regression, another in conditional distribution function, and the third in quantile regression.

3.1. **Generalized Instrumental Variables Models.** First, we consider generalized instrumental variables models of Chesher and Rosen (2014). Specially, we illustrate usefulness of our framework using Example 5 of Chesher and Rosen (2014) with the restriction that the structural error U is independent of the instrument Z. In Example 5 of Chesher and Rosen (2014), the outcome variable Y_1 is fully observed, whereas the endogenous explanatory variable Y_2^* is interval censored, that is, $Y_2^* \in [Y_{2l}, Y_{2u}]$.

One of semiparametric specifications imposed in Chesher and Rosen (2014) is to assume the linear index for the structural function without any parametric specification of the distribution of U. In this specification, Chesher and Rosen (2014) show that the full independence between U and Z implies that the identified set for the structural parameter β is given by the set of b's that satisfy

$$G_1(b, \tau_1, \tau_2, z) \leq G_2(b, \tau_1, \tau_2, z) \tag{3.1}$$

for every z and $(\tau_1, \tau_2) \in \mathcal{T} \equiv \{(\tau_1, \tau_2) \in \mathbf{R}^2 : \tau_1 \leq \tau_2\}$, where

$$G_1(b, \tau_1, \tau_2, z) \equiv P(\tau_1 + bY_{2u} \leq Y_1 \leq \tau_2 + bY_{2l} | Z = z),$$
$$G_2(b, \tau_1, \tau_2, z) \equiv P(\tau_1 + bY_{2l} \leq Y_1 \leq \tau_2 + bY_{2u} | Z = z).$$

The identified set in (3.1) is a simplified version of the identified set obtained in Section 4 of Chesher and Rosen (2014), without including exogenous explanatory variables. Then, a confidence region for β can be obtained by inverting pointwise (in b) tests with $v_{\tau,b}(x) \equiv G_1(b, \tau_1, \tau_2, x) - G_2(b, \tau_1, \tau_2, x)$, where $\tau = (\tau_1, \tau_2)$.

3.2. **Panel Data Models with Endogenous Censoring.** Consider a panel data model of Khan, Ponomareva, and Tamer (2013). In their framework, a researcher only observes $\{(Y_{it}, D_{it}, X_{it}) : i = 1, \ldots, n, t = 1, \ldots, T\}$ generated from

$$Y_{it} = \max\{Y_{it}^*, C_{it}\},$$
$$D_{it} = 1\{Y_{it}^* \geq C_{it}\},$$
$$Y_{it}^* = \alpha_i + X_{it}'\beta + U_{it},$$

where α_i is the unobserved fixed effect that can be correlated with $X_i = (X_{i1}, \ldots, X_{iT})'$ and $U_i = (U_{i1}, \ldots, U_{iT})'$. Khan, Ponomareva, and Tamer (2013) consider endogenous censoring and obtain bounds under alternative modeling assumptions. To illustrate their approach, note that

$$Y_{it}^L \leq Y_{it}^* \leq Y_{it},$$

where $Y_{it}^L = D_{it}Y_{it} + (1 - D_{it})(-\infty)$.

When $\alpha_i + U_{it}$ has the same distribution as $\alpha_i + U_{is}$ conditional on X_i for $t \neq s$ (which they call Model 1), they show that the identified set is the set of b's that satisfy

$$P(Y_{it} - X_{it}'b \leq y | X_i = x) \leq P(Y_{it}^L - X_{it}'b \leq y | X_i = x)$$

for every (y, x) and every $t = 1, \ldots, T$. Then, to construct a confidence region for β, we may take the following route: for each $j = 1, \ldots, T$, we define

$$\begin{aligned} v_{\tau,j,b}(x) &= v_{\tau,j}(x;b) \\ &\equiv P(Y_{ij} - X_{ij}'b \leq \tau | X_i = x) - P(Y_{ij}^L - X_{ij}'b \leq \tau | X_i = x), \end{aligned}$$

and carry out our test pointwise in b. Khan, Ponomareva, and Tamer (2013) focus on the case when covariates have discrete distribution with finite support. Our method provides an inference method for the case of continuous covariates. Our general framework also applies to other partially identified panel data models. For example, see Jun, Lee, and Shin (2011), Li and Oka (2013) and Rosen (2012) among others.

3.3. **Empirical Example 1: Testing Functional Inequalities in Auction Models.** In this example, we go back to the auction environment of GPV mentioned earlier. We first state the testing problem formally, give the form of test statistic, and present empirical results.

3.3.1. *Testing Problem.* Suppose that the number I of bidders can take two values, 2 and 3 (that is, $I \in \{2, 3\}$). For each τ such that $0 < \tau < 1$, let $q_k(\tau|x)$ denote the τ-th conditional quantile (given $X = x$) of the observed equilibrium bid distribution when the number of bidders is $I = k$, where $k = 2, 3$. A conditional version of Equation (5) of GPV (with $I_1 = 2$ and $I_2 = 3$ in their notation) provides the following testing restrictions:

$$\begin{aligned} &q_2(\tau|x) - q_3(\tau|x) < 0, \\ &\underline{b} - 2q_2(\tau|x) + q_3(\tau|x) < 0 \end{aligned} \tag{3.2}$$

for any $\tau \in (0, 1]$ and for any $x \in \text{supp}(X)$, where $\text{supp}(X)$ is the (common) support of X, and $\underline{b}$ is the left endpoint of the support of the observed bids.[13] The restrictions in (3.2) are based on conditionally exogenous participation for which the latent private value distribution is independent of the number of bidders conditional on observed characteristics (X), e.g. appraisal values.

A slightly weaker version of (3.2) can be put into our general testing problem in (1.1).[14] That is, we can test the following null hypothesis:

$$\begin{aligned} &v_{\tau,1}(x) \equiv q_2(\tau|x) - q_3(\tau|x) \leq 0, \\ &v_{\tau,2}(x) \equiv \underline{b} - 2q_2(\tau|x) + q_3(\tau|x) \leq 0 \end{aligned} \tag{3.3}$$

for any $(\tau, x) \in \mathcal{T} \times \mathcal{X} \subset (0, 1] \times \text{supp}(X)$.

[13] In GPV, it is assumed that for $I = k$, the support of the observed equilibrium bid distribution is $[\underline{b}, \overline{b}_k] \subset [0, \infty)$ with $\underline{b} < \overline{b}_k$, where $k = 2, 3$. Note that $\underline{b}$ is common across k's, while $\overline{b}_k$'s are not.

[14] If necessary, we may test the strict inequalities (3.1), instead of the weak inequalities (3.2). However, such test would require a test statistic that is different from ours and needs a separate treatment.

The example in (3.3) illustrates that in order to test the implications of auction theory, it is essential to test the null hypothesis uniformly in τ and x. More specifically, testing for a wide range of τ is important because testable implications are expressed in terms of conditional stochastic dominance relations. Furthermore, testing the relations uniformly over x is natural since theoretical predictions given by conditionally exogenous participation should hold for any realization of observed auction heterogeneity. It also shows that it is important to go beyond the $J = 1$ case and to include a general $J > 1$. In fact, if the number of bidders can take more than two values, there could be many more functional inequalities (see Corollary 1 of GPV). Finally, we note that $v_{\tau,1}(x)$ and $v_{\tau,2}(x)$ are not forms of conditional moment inequalities and each involves two different conditional quantile functions indexed by τ. Therefore, tests developed for conditional moment inequalities are not directly applicable to this empirical example. There exist related but distinct papers regarding this empirical example. See, e.g., Marmer, Shneyerov, and Xu (2013) who developed a nonparametric test for selective entry, and Gimenes and Guerre (2013) who proposed augmented quantile regression for first-price auction models.

3.3.2. *Test Statistic.* To implement the test, it is necessary to estimate conditional quantile functions. In estimation of $q_j(\tau|x)$, $j = 2, 3$, we may use a local polynomial quantile regression estimator, say $\widehat{q}_j(\tau|x)$. Now write

$$\begin{aligned} \hat{v}_{\tau,1}(x) &= \hat{q}_2(\tau|x) - \hat{q}_3(\tau|x), \\ \hat{v}_{\tau,2}(x) &= \hat{\underline{b}} - 2\hat{q}_2(\tau|x) + \hat{q}_3(\tau|x), \end{aligned}$$

where $\hat{\underline{b}}$ is a consistent estimator of $\underline{b}$.[15] Then testing (3.3) can be carried out using $\{\hat{v}_{\tau,j}(x) : j = 1, 2\}$ based on our general framework. In this application, our test statistics take the following forms:

$$\begin{aligned} \hat{\theta}_{\text{sum}} &= \int_{\mathcal{X}\times\mathcal{T}} [r_n \hat{v}_{\tau,1}(x)]_+^p \, dQ(x,\tau) + \int_{\mathcal{T}\times\mathcal{X}} [r_n \hat{v}_{\tau,2}(x)]_+^p \, dQ(x,\tau), \text{ or} \\ \hat{\theta}_{\max} &= \int_{\mathcal{X}\times\mathcal{T}} \left(\max\left\{[r_n \hat{v}_{\tau,1}(x)]_+, [r_n \hat{v}_{\tau,2}(x)]_+\right\}\right)^p dQ(x,\tau). \end{aligned} \tag{3.4}$$

Note that in (3.4), we set $\hat{\sigma}_{\tau,j}(x) \equiv 1$.

As a matter of fact, it is possible to develop an alternative test statistic by rewriting (3.3) in terms of distribution functions. Appendix II.1 illustrates the usefulness and flexibility of our framework by reconsidering the implications from GPV using a test statistic based on estimating conditional cumulative distribution functions.

[15]In our application, we set $\hat{\underline{b}}$ to be the observed minimum value.

3.3.3. *Empirical Results.* We now present empirical results using the timber auction data used in Lu and Perrigne (2008).[16] They used the timer auction data to estimate bidders' risk aversion, taking advantage of bidding data from ascending auctions as well as those from first-price sealed-bid auctions. In our empirical example, we use only the latter auctions with 2 and 3 bidders, and we use the appraisal value as the only covariate X_i ($d = 1$). Summary statistics and visual presentation of data are given in Table 1 and Figure 2. It can be seen from Table 1 that average bids become higher as the number of bidders increases from 2 to 3. The top panel of Figure 2 suggests that this more aggressive bidding seems to be true, conditional on appraisal values.

Table 1. Summary Statistics for Empirical Example 1

	2 bidders (Sample size = 107)		3 bidders (Sample size = 108)	
	Mean	Standard Deviation	Mean	Standard Deviation
Appraisal Value	66.0	47.7	53.3	41.4
Highest bid	96.1	55.6	100.8	56.7
Second highest bid	80.9	49.2	83.1	51.5
Third highest bid			69.4	44.6

Notes: Bids and appraisal values are given in dollars per thousand board-feet (MBF). Source: Timber auction data are from the *Journal of Applied Econometrics* website.

Before estimation, the covariate was transformed to lie between 0 and 1 by studentizing it and then applying the standard normal CDF transformation. The bottom panel of Figure 2 shows local linear estimates of conditional quantile functions at $\tau = 0.1, 0.5, 0.9$.[17] In this figure, estimates are only shown between the 10% and 90% sample quantiles of the covariate.

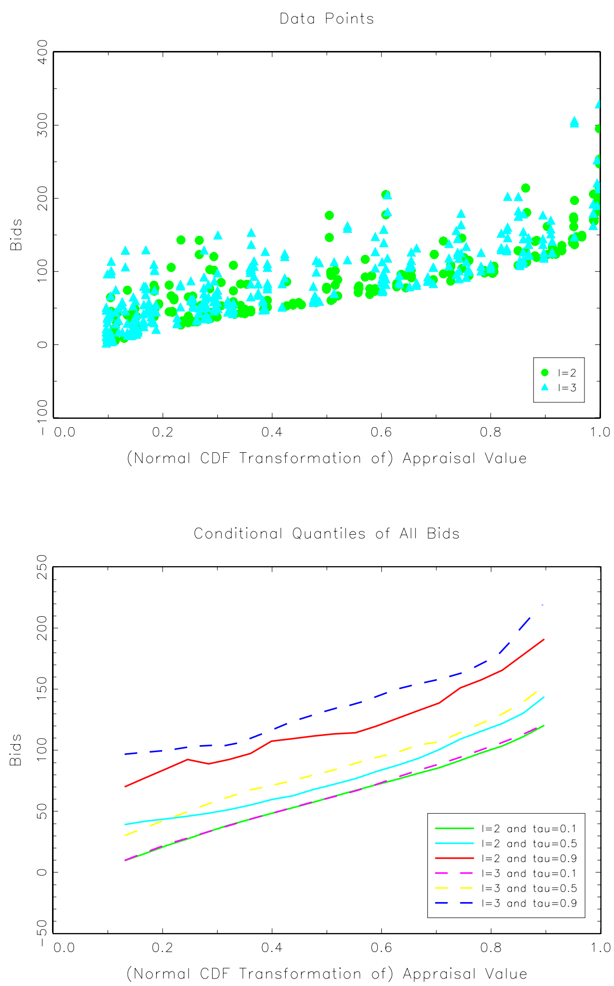
On one hand, the 10% conditional quantiles are almost identical between auctions with two bidders ($I = 2$) and those with three bidders ($I = 3$). On the other hand, the 50% and 90% conditional quantiles are higher with three bidders for most values of appraisal values. There is a crossing of two conditional median curves at the lower end of appraisal values.

To check whether inequalities in (3.3) hold in this empirical example, we plot estimates of $v_{\tau,1}(x)$ and $v_{\tau,2}(x)$ in Figure 3. The top panel of the figure shows that 20 estimated curves of $v_{\tau,1}(x)$, each representing a particular conditional quantile, ranging from the 10th percentile to the 90th percentile. There are strictly positive values of $v_{\tau,1}(x)$ at the lower end of appraisal values. The bottom panel of Figure 3 depicts 20 estimated curves of $v_{\tau,2}(x)$,

[16]The data are available on the *Journal of Applied Econometrics* website.

[17]Specifically, the conditional quantile functions $q_2(\tau|x)$ and $q_3(\tau|x)$ are estimated via the local linear quantile regression estimator with the kernel function $K(u) = 1.5[1-(2u)^2]\times 1\{|u| \leq 0.5\}$ and the bandwidth $h = 0.6$. See Section 5.1 for more details on estimating conditional quantile functions.

FIGURE 2. Data for Empirical Illustration for Empirical Example 1

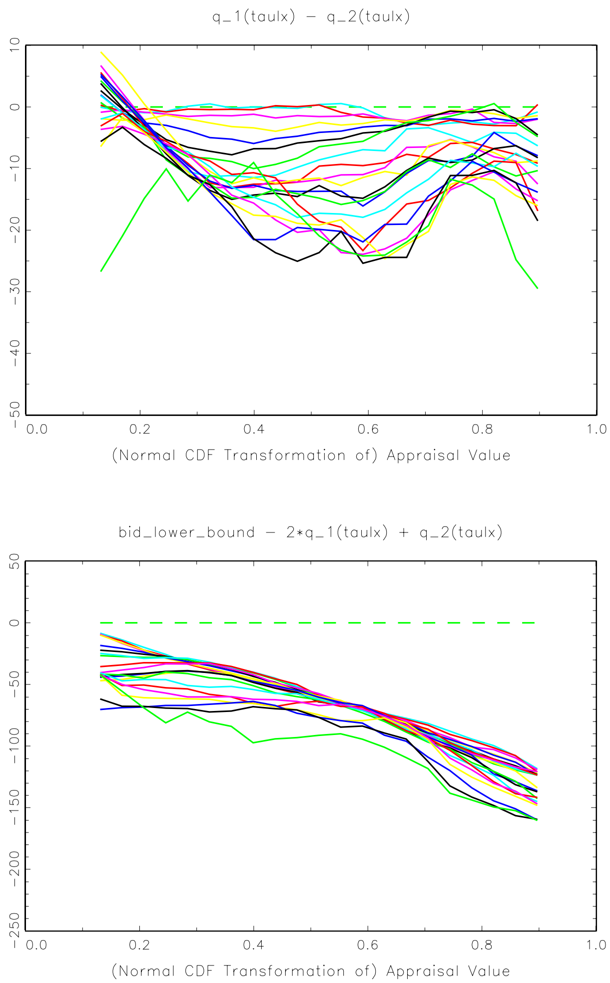


Note: The top panel of the figure shows observations and the bottom panel depicts local linear quantile regression estimates.

showing that they are all strictly negative. The test based on (3.4) can tell formally whether positive values of $v_{\tau,1}(x)$ at the lower end of appraisal values can be viewed as evidence against economic restrictions imposed by (3.3).

We considered both the L_1 and L_2 test statistics described in (3.4). We set $\mathcal{T}$ to be the interval between the 10th and 90th percentiles of the covariate, and also set $\mathcal{X} = [0.1, 0.9]$. The contact set was estimated with $\hat{c}_n = C_{\text{cs}} \log\log(n) q_{1-0.1/\log(n)}(S_n^*)$ with $r_n = \sqrt{nh}$. We checked the sensitivity to the tuning parameters with $C_{\text{cs}} \in \{0.5, 1, 1.5\}$ and $h \in \{0.3, 0.6, 0.9\}$. All cases resulted in bootstrap p-values of 1, thereby suggesting that positive values of $v_{\tau,1}(x)$ at the lower end of appraisal values cannot be interpreted as evidence

FIGURE 3. Estimates of $v_{\tau,1}(x)$ and $v_{\tau,2}(x)$ for Empirical Example 1



Note: The top and bottom panels of the figure show estimates of $v_{\tau,1}(x)$ and $v_{\tau,2}(x)$, respectively, where $\hat{v}_{\tau,1}(x) = \hat{q}_1(\tau|x) - \hat{q}_2(\tau|x)$ and $\hat{v}_{\tau,2}(x) = \underline{b} - 2\hat{q}_1(\tau|x) + \hat{q}_2(\tau|x)$.

against the null hypothesis beyond random sampling errors. Therefore, we have not found any evidence against economic implications imposed by (3.3).

3.4. **Empirical Example 2: Testing Functional Inequalities in the Context of Wage Inequality.** We now give an example based on Acemoglu and Autor (2011).

3.4.1. *Testing Problem.* Figures 9a-9c in Acemoglu and Autor (2011) depict changes in log hourly wages by percentile relative the median. Specifically, they consider the following differences-in-differences in quantiles:

$$\Delta_{t,s}(\tau, x) \equiv [q_t(\tau|x) - q_s(\tau|x)] - [q_t(0.5|x) - q_s(0.5|x)]$$

for time periods t and s and for quantiles τ, where $q_t(\tau|x)$ denotes the τ-quantile of log hourly wages conditional on $X = x$ in year t. Acemoglu and Autor (2011) consider males and females together in Figure 9a, males only in Figure 9b, and females only in Figure 9c. Thus, in their setup, the only covariate X is gender.

Figures 9a-9c in Acemoglu and Autor (2011) suggest that (1) $\Delta_{1988,1974}(\tau, x) \geq 0$ for quantiles above the median, but $\Delta_{1988,1974}(\tau, x) \leq 0$ for quantiles below the median (hence, widening the wage inequality, while the lower quantiles losing most), and that (2) $\Delta_{2008,1988}(\tau, x) \geq 0$ for most of quantiles (hence, 'polarization' of wage growth, while middle quantiles losing most). In this subsection, we consider testing

$$H_0 : \Delta_{t,s}(\tau, x) \geq 0 \ \forall (x, \tau) \in \mathcal{X} \times \mathcal{T}, \tag{3.5}$$

with a continuous covariate, where $(t, s) = (1988, 1974)$ or $(t, s) = (2008, 1988)$.[18] Note that degeneracy of the test statistic could occur if the contact set consists of values of (x, τ) only around $\tau = 0.5$. Therefore, the uniformity of our test could be potentially important in this example.

3.4.2. *Test Statistic.* To implement the test, we again use a local polynomial quantile regression estimator, say $\hat{q}_t(\tau|x)$. Then $\Delta_{t,s}(\tau, x)$ can be estimated by

$$\hat{\Delta}_{t,s}(\tau, x) \equiv [\hat{q}_t(\tau|x) - \hat{q}_s(\tau|x)] - [\hat{q}_t(0.5|x) - \hat{q}_s(0.5|x)].$$

Then testing (3.5) can be carried out using

$$\hat{\theta}_{t,s} \equiv \int_{\mathcal{X} \times \mathcal{T}} [r_n \hat{v}_{\tau,t,s}(x)]_+^p \, dQ(x, \tau), \tag{3.6}$$

where $\hat{v}_{\tau,t,s}(x) = -\hat{\Delta}_{t,s}(\tau, x)$.[19] Here, to reflect different sample sizes between two time periods, we set

$$r_n = \sqrt{\frac{(n_t h_t) \times (n_s h_s)}{(n_t h_t) + (n_s h_s)}},$$

where n_j and h_j are the sample size and the bandwidth used for nonparametric estimation for year $j = t, s$.

---

[18] Note that H_0 in (3.5) includes the case $\Delta_{t,s}(\tau, x) \equiv 0$, which does not correspond to the notion of polarization. In view of this, our null hypothesis in (3.5) can be regarded as a weak form of polarization hypothesis, whereas a more strict version can be written as the inequality in (3.5) holds strictly for some high and low quantiles.

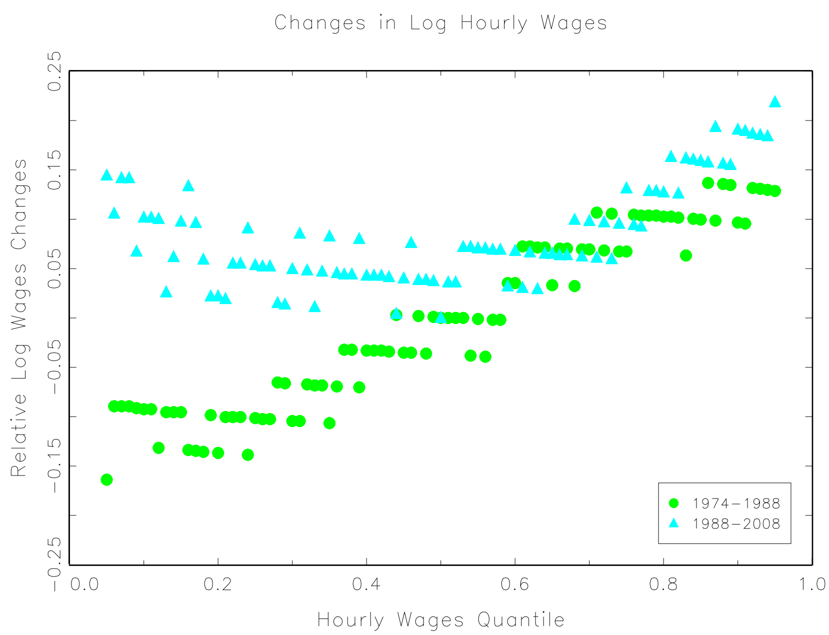
[19] Note that the null hypothesis is written as positivity in (3.5). Hence $\hat{v}_{\tau,t,s}(x)$ is defined accordingly.

Table 2. Summary Statistics for Empirical Example 2

Year	1974	1988	2008
Log Real Hourly Wages	2.780	2.769	2.907
Age in Years	35.918	35.501	39.051
Sample Size	19575	64682	48341

Notes: The sample is restricted to white males, with age between 16 and 64. Entries for log real hourly wages and age show CPS sample weighted means. Source: May/ORG CPS data extract from David Autor's web site.

Figure 4. Changes in Log Hourly Wages by Percentile Relative to the Median

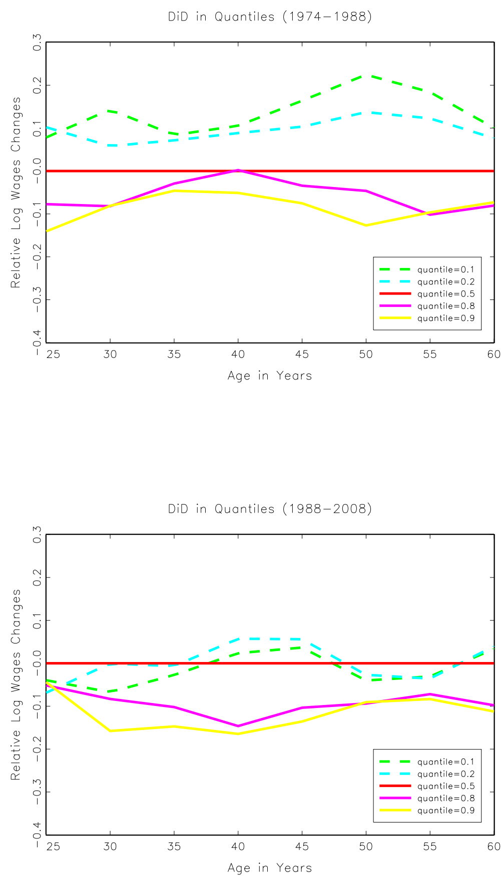


Notes: The figure shows differences-in-differences in quantiles of log hourly wages, measured by $[q_t(\tau) - q_s(\tau)] - [q_t(0.5) - q_s(0.5)]$. Triangles correspond to changes from 1974 to 1988, whereas circles those from 1988 to 2008. All quantiles are computed using CPS sample weight. Source: May/ORG CPS data extract from David Autor's web site.

3.4.3. *Empirical Results.* We used the CPS data extract of Acemoglu and Autor (2011).[20] In our empirical example, we use age in years as the only covariate. Summary statistics and

[20]The data are available on David Autor's web site. We would like to thank him for posting the data set on a public domain. They used three-year averages around the year of interest to produce Figures 9a-9c in Acemoglu and Autor (2011); however, we used just annual data.

FIGURE 5. Estimates of $\hat{v}_{\tau,t,s}(x)$



Note: The top and bottom panels of the figure show local linear estimates of $-\Delta_{1988,1974}(\tau, x)$ and $-\Delta_{2008,1988}(\tau, x)$, respectively, where x is age in years.

visual presentation of data are given in Table 2 and Figure 4. Note that Figure 4 replicates the basic patterns of Figures 9 of Acemoglu and Autor (2011).

We now turn to the conditional version of Figure 4, using age as a conditioning variable. As an illustration, let $\mathcal{X}$ be an interval of ages between 25 and 60 and let $\mathcal{T} = [0.1, 0.9]$. To check whether inequalities in $\hat{\Delta}_{t,s}(\tau, x) \geq 0$ hold for each value of $(x, \tau) \in \mathcal{X} \times \mathcal{T}$, we plot estimates of $\hat{v}_{\tau,t,s}(x) = -\hat{\Delta}_{t,s}(\tau, x)$ in Figure 5. The top panel of the figure shows that 5 estimated curves of $\hat{v}_{\tau,1988,1974}(x)$, each representing a particular conditional quantile, and the bottom panel shows the corresponding graph for period 1988-2008.[21] By construction, the estimated curve is a flat line at zero when $\tau = 0.5$. As consistent with Figure 4, the

[21]As before, underlying conditional quantile functions are estimated via the local linear quantile regression estimator with the kernel function $K(u) = 1.5[1 - (2u)^2] \times 1\{|u| \leq 0.5\}$. One important difference from the

lower quantiles seem to violate the null hypothesis, especially for the period 1974-1988. As before, our test can tell formally whether positive values of $\hat{v}_{\tau,t,s}(x)$ lead to rejection of the null hypothesis of polarization of wage growth.

We considered both the L_1 and L_2 test statistics described in (3.6). As before, the contact set was estimated with $\hat{c}_n = C_{\text{cs}} \log\log(n) q_{1-0.1/\log(n)}(S_n^*)$ with $r_n = \sqrt{nh}$.[22] We checked the sensitivity to the tuning parameters with $C_{\text{cs}} \in \{0.5, 1, 1.5\}$.

For period 1974-1988, we rejected the null hypothesis at the 1% level across all three values of C_{cs}. However, for period 1988-2008, we fail to reject the null hypothesis at the 5% level for any value of C_{cs}. Therefore, the changing patterns of the US wage distribution around 1988, reported in Acemoglu and Autor (2011), seem to hold up conditionally on age as well.

4. Uniform Asymptotics under General Conditions

In this section, we establish uniform asymptotic validity of our bootstrap test using high-level conditions. We also provide a class of distributions for which the asymptotic size is exact. We first define the set of distributions we consider.

Definition 1. Let $\mathcal{P}$ denote the collection of the potential joint distributions of the observed random vectors that satisfy Assumptions A1-A6, and B1-B4 given below. Let $\mathcal{P}_0 \subset \mathcal{P}$ be the sub-collection of potential distributions that satisfy the null hypothesis.

Let $||\cdot||$ denote the Euclidean norm throughout the paper. For any given sequence of subcollections $\mathcal{P}_n \subset \mathcal{P}$, any sequence of real numbers $b_n > 0$, and any sequence of random vectors Z_n, we say that $Z_n/b_n \to_P 0$, $\mathcal{P}_n$-uniformly, or $Z_n = o_P(b_n)$, $\mathcal{P}_n$-uniformly, if for any $a > 0$,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} P\left\{||Z_n|| > ab_n\right\} = 0.$$

Similarly, we say that $Z_n = O_P(b_n)$, $\mathcal{P}_n$-uniformly, if for any $a > 0$, there exists $M > 0$ such that

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} P\left\{||Z_n|| > Mb_n\right\} < a.$$

We also define their bootstrap counterparts. Let P^* denote the probability under the bootstrap distribution. For any given sequence of subcollections $\mathcal{P}_n \subset \mathcal{P}$, any sequence of real numbers $b_n > 0$, and any sequence of random vectors Z_n^*, we say that $Z_n^*/b_n \to_{P^*} 0$, $\mathcal{P}_n$-uniformly, or $Z_n^* = o_{P^*}(b_n)$, $\mathcal{P}_n$-uniformly, if for any $a > 0$,

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} P\left\{P^*\left\{||Z_n^*|| > ab_n\right\} > a\right\} = 0.$$

first empirical example is that we used the CPS sample weight, which were incorporated by multiplying it to the kernel weight for each observation. Finally, the bandwidth was $h = 2.5$ for all years.

[22]To accommodate different sample sizes across years, we set $n = (n_{1974} + n_{1988} + n_{2008})/3$ in computing $\hat{c}_n$.

Similarly, we say that $Z_n^* = O_{P^*}(b_n)$, $\mathcal{P}_n$-uniformly, if for any $a > 0$, there exists $M > 0$ such that

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}_n} P\left\{P^*\left\{||Z_n^*|| > Mb_n\right\} > a\right\} < a.$$

In particular, when we say $Z_n = o_P(b_n)$ or $O_P(b_n)$, $\mathcal{P}$-uniformly, it means that the convergence holds uniformly over $P \in \mathcal{P}$, and when we say $Z_n = o_P(b_n)$ or $O_P(b_n)$, $\mathcal{P}_0$-uniformly, it means that the convergence holds uniformly over all the probabilities in $\mathcal{P}$ that satisfy the null hypothesis.

4.1. **Test Statistics and Critical Values in General Form.** First, let us extend the test statistics and the bootstrap procedure to the general case of $J \geq 1$. Let $\Lambda_p : \mathbf{R}^J \to [0,\infty)$ be a nonnegative, increasing function indexed by $p \geq 1$. While the theory of this paper can be extended to various general forms of map Λ_p, we focus on the following type:

$$\Lambda_p(v_1,\cdots,v_J) = (\max\{[v_1]_+,\cdots,[v_J]_+\})^p \text{ or } \Lambda_p(v_1,\cdots,v_J) = \sum_{j=1}^J [v_j]_+^p, \tag{4.1}$$

where for $a \in \mathbf{R}$, $[a]_+ = \max\{a,0\}$. The test statistic is defined as

$$\hat{\theta} = \int_{\mathcal{X}\times\mathcal{T}} \Lambda_p\left(\hat{u}_{\tau,1}(x),\cdots,\hat{u}_{\tau,J}(x)\right) dQ(x,\tau).$$

To motivate our bootstrap procedure, it is convenient to begin with the following lemma. Let us introduce some notation. Define $\mathcal{N}_J \equiv 2^{\mathbb{N}_J}\backslash\{\varnothing\}$, i.e., the collection of all the nonempty subsets of $\mathbb{N}_J \equiv \{1,2,\cdots,J\}$. For any $A \in \mathcal{N}_J$ and $\mathbf{v} = (v_1,\cdots,v_J)^\top \in \mathbf{R}^J$, we define $\mathbf{v}_A$ to be $\mathbf{v}$ except that for each $j \in \mathbb{N}_J\backslash A$, the j-th entry of $\mathbf{v}_A$ is zero, and let

$$\Lambda_{A,p}(\mathbf{v}) \equiv \Lambda_p(\mathbf{v}_A). \tag{4.2}$$

That is, $\Lambda_{A,p}(\mathbf{v})$ is a "censoring" of $\Lambda_p(\mathbf{v})$ outside the index set A. Now, we define a general version of contact sets: for $A \in \mathcal{N}_J$ and for $c_{n,1}, c_{n,2} > 0$,

(4.3)

$$B_{n,A}(c_{n,1},c_{n,2}) \equiv \left\{(x,\tau) \in \mathcal{X}\times\mathcal{T} : \begin{array}{lll} |r_{n,j}v_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x)| & \leq c_{n,1}, & \text{for all } j \in A \\ r_{n,j}v_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x) & < -c_{n,2}, & \text{for all } j \in \mathbb{N}_J/A \end{array}\right\},$$

where $\sigma_{n,\tau,j}(x)$ is a "population" version of $\hat{\sigma}_{\tau,j}(x)$ (see e.g. Assumption A5 below.) When $c_{n,1} = c_{n,2} = c_n$ for some $c_n > 0$, we write $B_{n,A}(c_n) = B_{n,A}(c_{n,1},c_{n,2})$.

Lemma 1. *Suppose that Assumptions A1-A3 and A4(i) in Section 4.2 hold. Suppose further that $c_{n,1} > 0$ and $c_{n,2} > 0$ are sequences such that*

$$\sqrt{\log n}\{c_{n,1}^{-1} + c_{n,2}^{-1}\} \to 0,$$

as $n \to \infty$. Then as $n \to \infty$,

$$\inf_{P \in \mathcal{P}_0} P\left\{\hat{\theta} = \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,1},c_{n,2})} \Lambda_{A,p}(\hat{u}_{\tau,1}(x), \cdots, \hat{u}_{\tau,J}(x)) dQ(x,\tau)\right\} \to 1,$$

where $\mathcal{P}_0$ is the set of potential distributions of the observed random vector under the null hypothesis.

The lemma above shows that the test statistic $\hat{\theta}$ is uniformly approximated by the integral with domain restricted to the contact sets $B_{n,A}(c_{n,1}, c_{n,2})$ in large samples. Note that the asymptotic result is remarkable, in the sense that the approximation error between $\hat{\theta}$ and the expression on the right-hand side is $o_P(\varepsilon_n)$ for any $\varepsilon_n \to 0$. The result of Lemma 1 suggests that one may consider a bootstrap procedure that mimics the representation of $\hat{\theta}$ in Lemma 1.

We begin by introducing a sample version of the contact sets. For $A \in \mathcal{N}_J$,

$$\hat{B}_A(\hat{c}_n) \equiv \left\{(x,\tau) \in \mathcal{X} \times \mathcal{T} : \begin{array}{ll} |r_{n,j}\hat{v}_{\tau,j}(x)/\hat{\sigma}_{\tau,j}(x)| \leq \hat{c}_n, & \text{for all } j \in A \\ r_{n,j}\hat{v}_{\tau,j}(x)/\hat{\sigma}_{\tau,j}(x) < -\hat{c}_n, & \text{for all } j \in \mathbb{N}_J \backslash A \end{array}\right\}.$$

The explicit condition for $\hat{c}_n$ is found in Assumption A4 below. Given the bootstrap counterparts, $\{[\hat{v}^*_{\tau,j}(x), \hat{\sigma}^*_{\tau,j}(x)] : j \in \mathbb{N}_J\}$, of $\{[\hat{v}_{\tau,j}(x), \hat{\sigma}_{\tau,j}(x)] : j \in \mathbb{N}_J\}$, we define our bootstrap test statistic as follows:

$$\hat{\theta}^* \equiv \sum_{A \in \mathcal{N}_J} \int_{\hat{B}_A(\hat{c}_n)} \Lambda_{A,p}(\hat{s}^*_{\tau,1}(x), \cdots, \hat{s}^*_{\tau,J}(x)) dQ(x,\tau),$$

where for $j \in \mathbb{N}_J$, $\hat{s}^*_{\tau,j}(x) \equiv r_{n,j}(\hat{v}^*_{\tau,j}(x) - \hat{v}_{\tau,j}(x))/\hat{\sigma}^*_{\tau,j}(x)$. We also define

$$\hat{a}^* \equiv \sum_{A \in \mathcal{N}_J} \int_{\hat{B}_A(\hat{c}_n)} \mathbf{E}^* \Lambda_{A,p}(\hat{s}^*_{\tau,1}(x), \cdots, \hat{s}^*_{\tau,J}(x)) dQ(x,\tau).$$

Let c^*_α be the $(1-\alpha)$-th quantile from the bootstrap distribution of $\hat{\theta}^*$ and take

$$c^*_{\alpha,\eta} = \max\{c^*_\alpha, h^{d/2}\eta + \hat{a}^*\}$$

as our critical value, where $\eta > 0$ is a small fixed number.

One of the main technical contributions of this paper is to present precise conditions under which this proposal of bootstrap test works. We present and discuss them in subsequent sections.

To see the intuition for the bootstrap validity, first note that the uniform convergence of $r_{n,j}\{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)\}$ over (x,τ) implies that

$$B_{n,A}(c_{n,L}, c_{n,U}) \subset \hat{B}_A(\hat{c}_n) \subset B_{n,A}(c_{n,U}, c_{n,L}) \tag{4.4}$$

with probability approaching one, whenever $P\{c_{n,L} \leq \hat{c}_n \leq c_{n,U}\} \to 1$. Therefore, if $\sqrt{\log n}/c_{n,L} \to 0$, then, (letting $\hat{s}_{\tau,j} \equiv r_{n,j}(\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x))/\hat{\sigma}_{\tau,j}(x)$), we have

$$\hat{\theta} \leq \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,L}, c_{n,U})} \Lambda_{A,p}\left(\hat{s}_{\tau,1}(x), \cdots, \hat{s}_{\tau,J}(x)\right) dQ(x,\tau), \tag{4.5}$$

with probability approaching one, by Lemma 1 and the null hypothesis. When the last sum has a nondegenerate limit, we can approximate its distribution by the bootstrap distribution

$$\begin{aligned} & \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,L}, c_{n,U})} \Lambda_{A,p}\left(\hat{s}^*_{\tau,1}(x), \cdots, \hat{s}^*_{\tau,J}(x)\right) dQ(x,\tau) \\ \leq & \sum_{A \in \mathcal{N}_J} \int_{\hat{B}_A(\hat{c}_n)} \Lambda_{A,p}\left(\hat{s}^*_{\tau,1}(x), \cdots, \hat{s}^*_{\tau,J}(x)\right) dQ(x,\tau) \equiv \hat{\theta}^*, \end{aligned}$$

where the inequality follows from (4.4).[23] Thus the critical value is read from the bootstrap distribution of $\hat{\theta}^*$. On the other hand, if the last sum in (4.5) has limiting distribution degenerate at zero, we simply take a small positive number η to control the size of the test. This results in our choice of $c^*_{\alpha,\eta} = \max\{c^*_\alpha, h^{d/2}\eta + \hat{a}^*\}$.

4.2. **High-Level Regularity Conditions.** In this section, we provide high-level conditions needed to develop general results. We assume that $\mathcal{S} \equiv \mathcal{X} \times \mathcal{T}$ is a compact subset of a Euclidean space. We begin with the following assumption.

Assumption A1. *(Asymptotic Linear Representation) For each $j \in \mathbb{N}_J \equiv \{1, \cdots, J\}$, there exists a nonstochastic function $v_{n,\tau,j}(\cdot) : \mathbf{R}^d \to \mathbf{R}$ such that (a) $v_{n,\tau,j}(x) \leq 0$ for all $(x,\tau) \in \mathcal{S}$ under the null hypothesis, and (b) as $n \to \infty$,*

(4.6)
$$\sup_{(x,\tau)\in\mathcal{S}} \left| r_{n,j}\left\{\frac{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)}\right\} - \sqrt{nh^d}\{\hat{g}_{\tau,j}(x) - \mathbf{E}\hat{g}_{\tau,j}(x)\}\right| = o_P(\sqrt{h^d}),\ \mathcal{P}\text{-uniformly},$$

where, with $\{(Y_i^\top, X_i^\top)\}_{i=1}^n$ being a random sample such that $Y_i = (Y_{i1}^\top, \ldots, Y_{iJ}^\top)^\top \in \mathbf{R}^{J\bar{L}}$, $Y_{ij} \in \mathbf{R}^{\bar{L}}$, $X_i \in \mathbf{R}^d$, and the distribution of X_i is absolutely continuous with respect to Lebesgue measure,[24] *we define*

$$\hat{g}_{\tau,j}(x) \equiv \frac{1}{nh^d}\sum_{i=1}^n \beta_{n,x,\tau,j}\left(Y_{ij}, \frac{X_i - x}{h}\right),$$

and $\beta_{n,x,\tau,j} : \mathbf{R}^{\bar{L}} \times \mathbf{R}^d \to \mathbf{R}$ is a function which may depend on $n \geq 1$.

[23] In fact, the main challenge here is to prove the bootstrap approximation using the method of Poissonization that is uniform in $P \in \mathcal{P}_0$.

[24] Throughout the paper, we assume that $X_i \in \mathbf{R}^d$ is a continuous random vector. It is straightforward to extend the analysis to the case where X_i has a subvector of discrete random variables.

Assumption A1 requires that there exist a nonparametric function $v_{n,\tau,j}(x)$ around which the asymptotic linear representation holds uniformly in $P \in \mathcal{P}$, and $v_{n,\tau,j}(x) \leq 0$ under the null hypothesis. The required rate of convergence in (4.6) is $o_P(h^{d/2})$ instead of $o_P(1)$. We need this stronger convergence rate primarily because $\hat{\theta} - a_n$ is $O_P(h^{d/2})$ for some nonstochastic sequence a_n.[25]

When $\hat{v}_{\tau,j}(x)$ is a sample mean of i.i.d. random quantities involving nonnegative kernels and $\hat{\sigma}_{n,\tau}(x) = 1$, we may take $v_{n,\tau,j}(x) = \mathbf{E}\hat{v}_{\tau,j}(x)$, and then $o_P(\sqrt{h^d})$ is in fact precisely equal to 0. If the original nonparametric function $v_{\tau,j}(\cdot)$ satisfies some smoothness conditions, we may take $v_{n,\tau,j}(x) = v_{\tau,j}(x)$, and handle the bias part $\mathbf{E}\hat{v}_{\tau,j}(x) - v_{\tau,j}(x)$ using the standard arguments to deduce the error rate $o_P(\sqrt{h^d})$. Assumption A1 admits both set-ups. For instance, consider the simple example in Section 2.4. The asymptotic linear representation in Assumption 1 can be shown to hold with

$$\beta_{n,x,1}\left(Y_i, (X_i - x)/h\right) = Y_i K((X_i - x)/h)/\sigma_{n,1}(x),$$

where $\sigma_{n,1}^2(x) = \mathbf{E}[Y_i^2 K^2((X_i - x)/h)]/h$, if $\hat{\sigma}_{n,1}(x)$ is chosen as in (2.9).

The following assumption for $\beta_{n,x,\tau,j}$ essentially defines the scope of this paper's framework.

Assumption A2. *(Kernel-Type Condition) For some compact $\mathcal{K}_0 \subset \mathbf{R}^d$ that does not depend on $P \in \mathcal{P}$ or n, it is satisfied that $\beta_{n,x,\tau,j}(y,u) = 0$ for all $u \in \mathbf{R}^d \backslash \mathcal{K}_0$ and all $(x,\tau,y) \in \mathcal{X} \times \mathcal{T} \times \mathcal{Y}_j$ and all $j \in \mathbb{N}_J$, where $\mathcal{Y}_j$ denotes the support of Y_{ij}.*

Assumption A2 can be immediately verified when the asymptotic linear representation in (4.6) is established. This condition is satisfied in particular when the asymptotic linear representation involves a multivariate kernel function with bounded support in a multiplicative

[25] To see this more clearly, we assume that $\mathcal{T} = \{\tau\}$, $p = 1$, and $J = 1$, and suppress the subscripts τ and j from the notation, and take $\hat{\sigma}(x) = 1$ for simplicity. We write (in the case where $v_n(x) = 0$)

$$\begin{aligned} h^{-d/2}\hat{\theta} &= h^{-d/2}\int_{\mathcal{X}} \max\{r_n\{\hat{v}(x) - v_n(x)\}, 0\}\, dx \\ &= h^{-d/2}\int_{\mathcal{X}} \max\left\{\sqrt{nh^d}\{\hat{g}(x) - \mathbf{E}\hat{g}(x)\}, 0\right\} dx + h^{-d/2}R_n, \end{aligned}$$

where R_n is an error term that has at least the same convergence rate as the convergence rate of the remainder term in the asymptotic linear representation for $\hat{v}(x)$. Now we let

$$a_n = \mathbf{E}\left[\int_{\mathcal{X}} \max\left\{\sqrt{nh^d}\{\hat{g}(x) - \mathbf{E}\hat{g}(x)\}, 0\right\} dx\right]$$

and write $h^{-d/2}\hat{\theta} - h^{-d/2}a_n$ as

$$h^{-d/2}\left(\int_{\mathcal{X}} \max\left\{\sqrt{nh^d}\{\hat{g}(x) - \mathbf{E}\hat{g}(x)\}, 0\right\} dx - a_n\right) + h^{-d/2}R_n.$$

It can be shown that the leading term is asymptotically normal using the method of Poissonization. Hence $h^{-d/2}\hat{\theta} - h^{-d/2}a_n$ becomes asymptotically normal, if $R_n = o_P(h^{d/2})$. This is where the faster error rate in the asymptotic linear representation in Assumption A1(i) plays a role.

form. In such a case, the set $\mathcal{K}_0$ depends only on the choice of the kernel function, not on any model primitives.

Assumption A3. *(Uniform Convergence Rate for Nonparametric Estimators) For all $j \in \mathbb{N}_J$,*

$$\sup_{(x,\tau)\in\mathcal{S}} r_{n,j}\left|\frac{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)}\right| = O_P\left(\sqrt{\log n}\right), \ \mathcal{P}\text{-uniformly.}$$

Assumption A3 requires that $\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)$ have the uniform convergence rate of $O_P(r_{n,j}^{-1}\sqrt{\log n})$ uniformly over $P \in \mathcal{P}$. Lemma 2 in Section 4.4 provides some sufficient conditions for this convergence.

We now introduce conditions for the bandwidth h and the tuning parameter c_n for the contact sets.

Assumption A4. *(Rate Conditions for Tuning Parameters) (i) As $n \to \infty$, $h \to 0$, $\sqrt{\log n}/r_n \to 0$, and $n^{-1/2}h^{-d-\nu_1} \to 0$ for some arbitrarily small $\nu_1 > 0$, where $r_n \equiv \min_{j\in\mathbb{N}_J} r_{n,j}$.*
(ii) For each $n \geq 1$, there exist nonstochastic sequences $c_{n,L} > 0$ and $c_{n,U} > 0$ such that $c_{n,L} \leq c_{n,U}$, and

$$\inf_{P\in\mathcal{P}} P\left\{c_{n,L} \leq \hat{c}_n \leq c_{n,U}\right\} \to 1, \ \text{and} \ \sqrt{\log n}/c_{n,L} + c_{n,U}/r_n \to 0,$$

as $n \to \infty$.

The requirement that $\sqrt{\log n}/r_n \to 0$ is satisfied easily for most cases where r_n increases at a polynomial order in n. Assumption A4(ii) requires that $\hat{c}_n$ increase faster than $\sqrt{\log n}$ but slower than r_n with probability approaching one.

Assumption A5. *(Regularity Conditions for $\hat{\sigma}_{\tau,j}(x)$) For each $(\tau, j) \in \mathcal{T} \times \mathbb{N}_J$, there exists $\sigma_{n,\tau,j}(\cdot) : \mathcal{X} \to (0,\infty)$ such that $\liminf_{n\to\infty} \inf_{(x,\tau)\in\mathcal{S}} \inf_{P\in\mathcal{P}} \sigma_{n,\tau,j}(x) > 0$, and*

$$\sup_{(x,\tau)\in\mathcal{S}} |\hat{\sigma}_{\tau,j}(x) - \sigma_{n,\tau,j}(x)| = o_P(1), \ \mathcal{P}\text{-uniformly.}$$

Assumption A5 requires that the scale normalization $\hat{\sigma}_{\tau,j}(x)$ should be asymptotically well defined. The condition precludes the case where estimator $\hat{\sigma}_{\tau,j}(x)$ converges to a map that becomes zero at some point (x,τ) in $\mathcal{S}$. Assumption A5 is usually satisfied by an appropriate choice of $\hat{\sigma}_{\tau,j}(x)$. When one chooses $\hat{\sigma}_{\tau,j}(x) = 1$, which is permitted in our framework, Assumption A5 is immediately satisfied with $\sigma_{n,\tau,j}(x) = 1$. Again, if we go back to the simple example considered in Section 2.4, it is straightforward to see that under regularity conditions, with the subscript τ suppressed, $\hat{\sigma}_1^2(x) = \sigma_{n,1}^2(x) + o_P(1)$ and $\sigma_{n,1}^2(x) = \sigma_1^2(x) + o(1)$, where $\sigma_1^2(x) \equiv \mathbf{E}(Y^2|X = x)f(x)\int K^2(u)du$, as $n \to \infty$. The convergence can be strengthened to a uniform convergence when $\sigma_1^2(x)$ is bounded away from zero uniformly over $x \in \mathcal{X}$ and $P \in \mathcal{P}$, so that Assumption A5 holds.

We introduce assumptions about the moment conditions for $\beta_{n,x,\tau,j}(\cdot,\cdot)$ and other regularity conditions. For $\tau \in \mathcal{T}$ and $\varepsilon_1 > 0$, let $\mathcal{S}_\tau(\varepsilon_1) \equiv \{x + a : x \in \mathcal{S}_\tau,\ a \in [-\varepsilon_1, \varepsilon_1]^d\}$, where $\mathcal{S}_\tau \equiv \{x \in \mathcal{X} : (x,\tau) \in \mathcal{S}\}$ for each $\tau \in \mathcal{T}$. Let $\mathcal{U} \equiv \mathcal{K}_0 + \mathcal{K}_0$ such that $\mathcal{U}$ contains $\{0\}$ in its interior and $\mathcal{K}_0$ is the same as Assumption 2.

Assumption A6. *(i) There exist* $M \geq 2(p+2)$, $C > 0$, *and* $\varepsilon_1 > 0$ *such that*

$$\mathbf{E}[|\beta_{n,x,\tau,j}(Y_{ij}, u)|^M | X_i = x] f(x) \leq C,$$

for all $(x,u) \in \mathcal{S}_\tau(\varepsilon_1) \times \mathcal{U}$, $\tau \in \mathcal{T}$, $j \in \mathbb{N}_J$, $n \geq 1$, *and* $P \in \mathcal{P}$, *where* $f(\cdot)$ *is the density of* X_i.[26]
(ii) For each $a \in (0, 1/2)$, *there exists a compact set* $\mathcal{C}_a \subset \mathbf{R}^d$ *such that*

$$0 < \inf_{P \in \mathcal{P}} P\{X_i \in \mathbf{R}^d \backslash \mathcal{C}_a\} \leq \sup_{P \in \mathcal{P}} P\{X_i \in \mathbf{R}^d \backslash \mathcal{C}_a\} < a.$$

Assumption A6(i) requires that conditional moments of $\beta_{n,x,\tau,j}(Y_{ij}, z)$ be bounded. Assumption A6(ii) is a technical condition for the distribution of X_i. The third inequality in Assumption A6(ii) is satisfied if the distribution of X_i is uniformly tight in $\mathcal{P}$, and follows, for example, if $\sup_{P\in\mathcal{P}} \mathbf{E}||X_i|| < \infty$. The first inequality in Assumption A6(ii) requires that there be a common compact set outside which the distribution of X_i still has positive probability mass uniformly over $P \in \mathcal{P}$. The main thrust of Assumption A6(ii) lies in the requirement that such a compact set be independent of $P \in \mathcal{P}$. While it is necessary to make this technical condition explicit as stated here, the condition itself appears very weak.

This paper's asymptotic analysis adopts the approach of Poissonization (see, e.g., Horváth (1991) and Giné, Mason, and Zaitsev (2003)). However, existing methods of Poissonization are not readily applicable to our testing problem, mainly due to the possibility of local or global redundancy among the nonparametric functions. In particular, the conditional covariance matrix of $\beta_{n,x,\tau,j}(Y_{ij}, u)$'s across different (x,τ,j)'s given X_i can be singular in the limit. Since the empirical researcher rarely knows *a priori* the local relations among nonparametric functions, it is important that the validity of the test is not sensitive to the local relations among them, i.e., the validity should be uniform in P.

This paper deals with this challenge in three steps. First, we introduce a Poissonized version of the test statistic and apply a certain form of regularization to facilitate the derivation

[26] The conditional expectation $\mathbf{E}_P\left[|\beta_{n,x,\tau,j}(Y_{ij}, u)|^M | X_i = x\right]$ is of type $\mathbf{E}\left[f(Y,x)|X = x\right]$, which is not well defined according to Kolmogorov's definition of conditional expectations. See, e.g. Proschan and Presnell (1998) for this problem. Here we define the conditional expectation in an elementary way by using conditional densities or conditional probability mass functions of (Y_{ij}, Y_{ik}) given $X_i = x$, depending on whether (Y_{ij}, Y_{ik}) is continuous or discrete.

of its limiting distribution uniformly in $P \in \mathcal{P}$, i.e., regardless of singularity or degeneracy in the original test statistic. Second, we use a Berry-Esseen-type bound to compute the finite sample influence of the regularization bias and let the regularization parameter go to zero carefully, so that the bias disappears in the limit. Third, we translate thus computed limiting distribution into that of the original test statistic, using so-called de-Poissonization lemma. This is how the uniformity issue in this complex situation is covered through the Poissonization method combined with the method of regularization.

4.3. **Asymptotic Validity of Bootstrap Procedure.** Recall that $\mathbf{E}^*$ and P^* denote the expectation and the probability under the bootstrap distribution. We make the following assumptions for $\hat{v}^*_{\tau,j}(x)$.

Assumption B1. *(Bootstrap Asymptotic Linear Representation) For each $j \in \mathbb{N}_J$,*

$$\sup_{(x,\tau)\in\mathcal{S}} \left| r_{n,j} \left\{ \frac{\hat{v}^*_{\tau,j}(x) - \hat{v}_{\tau,j}(x)}{\hat{\sigma}^*_{\tau,j}(x)} \right\} - \sqrt{nh^d}\{\hat{g}^*_{\tau,j}(x) - \mathbf{E}^*\hat{g}^*_{\tau,j}(x)\} \right| = o_{P^*}(\sqrt{h^d}),\ \mathcal{P}\text{-uniformly},$$

where

$$\hat{g}^*_{\tau,j}(x) \equiv \frac{1}{nh^d}\sum_{i=1}^{n} \beta_{n,x,\tau,j}\left(Y^*_{ij}, \frac{X^*_i - x}{h}\right),$$

and $\beta_{n,x,\tau,j}$ is a real valued function introduced in Assumption A1.

Assumption B2. *For all $j \in \mathbb{N}_J$,*

$$\sup_{(x,\tau)\in\mathcal{S}} r_{n,j} \left| \frac{\hat{v}^*_{\tau,j}(x) - \hat{v}_{\tau,j}(x)}{\hat{\sigma}^*_{\tau,j}(x)} \right| = O_{P^*}(\sqrt{\log n}),\ \mathcal{P}\text{-uniformly.}$$

Assumption B3. *For all $j \in \mathbb{N}_J$,*

$$\sup_{(x,\tau)\in\mathcal{S}} \left| \hat{\sigma}^*_{\tau,j}(x) - \hat{\sigma}_{\tau,j}(x) \right| = o_{P^*}(1),\ \mathcal{P}\text{-uniformly.}$$

Assumption B1 is the asymptotic linear representation of the bootstrap estimator $\hat{v}^*_{\tau,j}(x)$. The proof of the asymptotic linear representation can be typically proceeded in a similar way that one obtains the original asymptotic linear representation in Assumption A1. Assumptions B2 and B3 are the bootstrap versions of Assumptions A3 and A5.

Assumption B4. *(Bandwidth Condition) $n^{-1/2}h^{-\left(\frac{3M-4}{2M-4}\right)d-\nu_2} \to 0$ as $n \to \infty$, for some small $\nu_2 > 0$ and for $M > 0$ that appears in Assumption A6(i).*

When $\beta_{n,x,\tau,j}(Y_{ij}, u)$ is bounded uniformly over (n, x, τ, j), the bandwidth condition in Assumption B4 is reduced to $n^{-1/2}h^{-3d/2-\nu_2} \to 0$. If Assumption A4(i) holds with $M = 6$ and $p = 1$ (this choice of (M, p) satisfies Assumption A6(i)), the bandwidth condition in Assumption B4 is reduced to $n^{-1/2}h^{-7d/4-\nu_2} \to 0$.

Note that Assumption B4 is stronger than the bandwidth condition in Assumption A4(i). The main reason is that we need to prove that for some $a_\infty > 0$, we have $a_n = a_\infty + o(h^{d/2})$ and $a_n^* = a_\infty + o_P(h^{d/2})$, $\mathcal{P}$-uniformly, where a_n is an appropriate location normalizer of the test statistic, and a_n^* is a bootstrap counterpart of a_n. To show these, we utilize a Berry-Esseen-type bound for a nonlinear transform of independent sum of random variables. Since the approximation error depends on the moment bounds for the sum, the bandwidth condition in Assumption B4 takes a form that involves $M > 0$ in Assumption A6.

We now present the result of the uniform validity of our bootstrap test.

Theorem 1. *Suppose that Assumptions A1-A6 and B1-B4 hold. Then*

$$\limsup_{n\to\infty} \sup_{P \in \mathcal{P}_0} P\{\hat{\theta} > c^*_{\alpha,\eta}\} \leq \alpha.$$

One might ask whether the bootstrap test $1\{\hat{\theta} > c^*_{\alpha,\eta}\}$ is asymptotically exact, i.e., whether the inequality in Theorem 1 holds as an equality. As we show below, the answer is affirmative in general. The remaining issue is a precise formulation of a subset of $\mathcal{P}_0$ such that the rejection probability of the bootstrap test achieves the level α asymptotically, uniformly over the subset.

To see when the test will have asymptotically exact size, we apply Lemma 1 to find that with probability approaching one,

$$\hat{\theta} = \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,U}, c_{n,L})} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x;\hat{\sigma})\right) dQ(x,\tau),$$

where $\hat{\mathbf{s}}_\tau(x) \equiv [r_{n,j}\{\hat{v}_{n,\tau,j}(x) - v_{n,\tau,j}(x)\}/\hat{\sigma}_{\tau,j}(x)]_{j=1}^J$, and $\mathbf{u}_{n,\tau}(x;\hat{\sigma}) \equiv [r_{n,j} v_{n,\tau,j}(x)/\hat{\sigma}_{\tau,j}(x)]_{j=1}^J$, and $c_{n,U} > 0$ and $c_{n,L} > 0$ are nonstochastic sequences that satisfy Assumption A4(ii).[27] We fix a positive sequence $q_n \to 0$, and write the right hand side as

$$\sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x;\hat{\sigma})\right) dQ(x,\tau) + \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,U}, c_{n,L}) \setminus B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x;\hat{\sigma})\right) dQ(x,\tau). \tag{4.7}$$

Under the null hypothesis, we have $v_{n,\tau,j}(x) \leq 0$, and hence the last sum is bounded by

$$\sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,U}, c_{n,L}) \setminus B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)\right) dQ(x,\tau),$$

[27]Note that we use Lemma 1 with $B_{n,A}(c_{n,U}, c_{n,L})$ here, differently from (4.5). This is because for asymptotic exactness, we need to use different arguments. See the roadmap of Appendix A for detailed explanations.

with probability approaching one. Using the uniform convergence rate in Assumption A3, we find that as long as

$$Q(B_{n,A}(c_{n,U},c_{n,L})\backslash B_{n,A}(q_n))\to 0,$$

fast enough, the second term in (4.7) vanishes in probability. As for the first integral, since for all $x\in B_{n,A}(q_n)$, we have $|r_{n,j}v_{n,\tau,j}(x)/\hat{\sigma}_{\tau,j}(x)|\leq q_n$ for all $j\in A$, we use the Lipschitz continuity of the map $\Lambda_{A,p}$ on a compact set, to approximate the leading sum in (4.7) by

$$\bar{\theta}_{1,n}(q_n)\equiv\sum_{A\in\mathcal{N}_J}\int_{B_{n,A}(q_n)}\Lambda_{A,p}\left(\hat{\mathbf{s}}_{\tau}(x)\right)dQ(x,\tau).$$

Thus we let

$$\tilde{\mathcal{P}}_n(\lambda_n,q_n)\equiv\left\{P\in\mathcal{P}:Q\left(\bigcup_{A\in\mathcal{N}_J}B_n(c_{n,U},c_{n,L})\backslash B_n(q_n)\right)\leq\lambda_n\right\},\tag{4.8}$$

where $B_n(c_{n,U},c_{n,L})\equiv\cup_{A\in\mathcal{N}_J}B_{n,A}(c_{n,U},c_{n,L})$, and find that

$$\hat{\theta}=\bar{\theta}_{1,n}(q_n)+o_P(h^{d/2}),\ \tilde{\mathcal{P}}_n(\lambda_n,q_n)\cap\mathcal{P}_0\text{-uniformly},$$

as long as λ_n and q_n converge to zero fast enough. We will specify the conditions in Theorem 2 below.

Let us deal with $\bar{\theta}_{1,n}(q_n)$. First, it can be shown that there are sequences of nonstochastic numbers $a_n(q_n)\in\mathbf{R}$ and $\sigma_n(q_n)>0$ that depend on q_n such that

$$h^{-d/2}\{\bar{\theta}_{1,n}(q_n)-a_n(q_n)\}/\sigma_n(q_n)\overset{d}{\to}N(0,1),\tag{4.9}$$

if $\text{liminf}_{n\to\infty}\sigma_n(q_n)>0$. We provide the precise formulae for $\sigma_n(q_n)$ and $a_n(q_n)$ in Section 6.3. Since the distribution of $h^{-d/2}\{\bar{\theta}_{1,n}(q_n)-a_n(q_n)\}/\sigma_n(q_n)$ is approximated by the bootstrap distribution of $h^{-d/2}\{\hat{\theta}^*-a_n(q_n)\}/\sigma_n(q_n)$ in large samples, we find that

$$\frac{h^{-d/2}\{c_\alpha^*-a_n(q_n)\}}{\sigma_n(q_n)}=\Phi^{-1}(1-\alpha)+o_P(1).$$

Hence the bootstrap critical value c_α^* will dominate $h^{-d/2}\eta+\hat{a}^*>0$, if for all $n\geq 1$,

$$\begin{aligned}\Phi^{-1}(1-\alpha)&\geq\frac{h^{-d/2}\{h^{d/2}\eta+\hat{a}^*-a_n(q_n)\}}{\sigma_n(q_n)}\\&=\frac{\eta+h^{-d/2}\{\hat{a}^*-a_n(q_n)\}}{\sigma_n(q_n)}.\end{aligned}$$

We can show that $\hat{a}^*-a_n(q_n)=o_P(h^{d/2})$, which follows if λ_n in (4.8) vanishes to zero sufficiently fast. Hence if

$$\sigma_n(q_n)\geq\eta/\Phi^{-1}(1-\alpha),$$

we have c^*_α becomes approximately equal to our bootstrap critical value $c^*_{\alpha,\eta}$. This leads to the following formulation of probabilities.

Definition 2. Define

$$\mathcal{P}_n(\lambda_n, q_n) \equiv \left\{ P \in \tilde{\mathcal{P}}_n(\lambda_n, q_n) : \sigma_n(q_n) \geq \eta/\Phi^{-1}(1-\alpha) \right\},$$

where $\tilde{\mathcal{P}}_n(\lambda_n, q_n)$ is as defined in (4.8).

The following theorem establishes the asymptotic exactness of the size of the bootstrap test over $P \in \mathcal{P}_n(\lambda_n, q_n) \cap \mathcal{P}_0$.

Theorem 2. *Suppose that Assumptions A1-A6 and B1-B4 hold. Let* $\lambda_n \to 0$ *and* $q_n \to 0$ *be positive sequences such that*

$$\begin{aligned} h^{-d/2} (\log n)^{p/2} \lambda_n &\to 0 \text{ and} \\ h^{-d/2} q_n \{(\log n)^{(p-1)/2} + q_n^{p-1}\} &\to 0. \end{aligned} \tag{4.10}$$

Then

$$\limsup_{n\to\infty} \sup_{P \in \mathcal{P}_n(\lambda_n, q_n)\cap\mathcal{P}_0} \left| P\{\hat{\theta} > c^*_{\alpha,\eta}\} - \alpha \right| = 0.$$

Theorem 2 shows that the rejection probability of our bootstrap test achieves exactly the level α uniformly over the set of probabilities in $\mathcal{P}_n(\lambda_n, q_n) \cap \mathcal{P}_0$. If $v_{n,\tau,j}(x) \equiv 0$ for each (x,τ) and for each j (the least favorable case, say P_{LFC}), then it is obvious that the distribution P_{LFC} belongs to $\mathcal{P}_n(\lambda_n, q_n)$ for any positive sequences $\lambda_n \to 0$ and $q_n \to 0$. This would be the only case of asymptotically exact coverage if bootstrap critical values were obtained as in (2.12), without contact set estimation. By estimating the contact sets and obtaining a critical value based on them, Theorem 2 establishes the asymptotically uniform exactness of the bootstrap test for distributions such that they may not satisfy $v_{n,\tau,j}(x) \equiv 0$ everywhere.

4.4. **Sufficient Conditions for Uniform Convergences in Assumptions A3 and B2.** This subsection gives sufficient conditions that yield Assumptions A3 and B2. The result is formalized in the following lemma.

Lemma 2. *(i) Suppose that Assumptions A1-A2 hold and that for each* $j \in \mathbb{N}_J$*, there exist finite constants* $C, \gamma_j > 0$*, and a positive sequence* $\delta_{n,j} > 0$ *such that for all* $n \geq 1$*, and all* $(x_1, \tau_1) \in \mathcal{S}$*,*

$$\mathbf{E}\left[\sup_{(x_2,\tau_2)\in\mathcal{S}:||x_1-x_2||+||\tau_1-\tau_2||\leq\lambda} (b_{n,ij}(x_1,\tau_1) - b_{n,ij}(x_2,\tau_2))^2 \right] \leq C\delta^2_{n,j}\lambda^{\gamma_j}, \text{ for all } \lambda > 0, \tag{4.11}$$

where $b_{n,ij}(x_1,\tau_1) \equiv \beta_{n,x_1,\tau_1,j}(Y_{ij},(X_i-x_1)/h)$ *and* $\limsup_{n\to\infty}\mathbf{E}[\sup_{(x,\tau)\in\mathcal{S}} b^4_{n,ij}(x,\tau)] \le C$ *and* $\delta_{n,j} = n^{s_{1,j}}$ *and* $h = n^{s_2}$ *for some* $s_{1,j}, s_2 \in \mathbf{R}$. *Furthermore, assume that*

$$n^{-1/2}h^{-d-\nu} \to 0,$$

for some small $\nu > 0$. *Then, Assumption A3 holds.*
(ii) Suppose further that Assumptions B1 and B3 hold. Then, Assumption B2 holds.

The condition (4.11) is the local L_2-continuity condition for $\beta_{n,x,\tau,j}(Y_{ij},(X_i-x)/h)$ in (x,τ). The condition corresponds to what Andrews (1994) called "Type IV class". The condition is satisfied by numerous maps that are continuous or discontinuous, as long as regularity conditions for the random vector (Y_i, X_i) are satisfied.[28] Typically, $\delta_{n,j}$ diverges to infinity at a polynomial rate in h^{-1}. The constant γ_j is 2 or can be smaller than 2, depending on the smoothness of the underlying function $b_{n,ij}(x,\tau)$. The value of γ_j does not affect the asymptotic theory of this paper, as long as it is strictly positive.

5. Verifying High-Level Conditions for the First Empirical Example

In this section, we use the auction model of GPV to illustrate how to verify high-level regularity conditions in Section 4.[29]

5.1. **Details on Estimating Conditional Quantile Functions.** We provide further details on the empirical example considered in Section 3.3. Assume that $q_k(\tau|x)$ is $(r+1)$-times continuously differentiable with respect to x, where $r \ge 1$. We use a local polynomial estimator $\widehat{q}_k(\tau|x)$. For $u \equiv (u_1,\ldots,u_d)$, a d-dimensional vector of nonnegative integers, let $[u] = u_1 + \cdots + u_d$. Let A_r be the set of all d-dimensional vectors u such that $[u] \le r$, and let $|A_r|$ denote the number of elements in A_r. For $z = (z_1,\cdot\cdot\cdot,z_d)^\top \in \mathbf{R}^d$ with $u = (u_1,\cdot\cdot\cdot,u_d)^\top \in A_r$, let $z^u = \prod_{m=1}^d z_m^{u_m}$. Now define $c(z) = (z^u)_{u\in A_r}$, for $z \in \mathbf{R}^d$. Note that $c(z)$ is a vector of dimension $|A_r|$.

Let $\{(B_{\ell i}, X_i, L_i) : \ell = 1,\ldots,L_i, i = 1,\ldots,n\}$ denote the observed data, where $\{B_{\ell i} : \ell = 1,\ldots,L_i\}$ denotes the L_i number of observed bids in the i-th auction, X_i a vector of observed characteristics for the i-th auction, and L_i the number of bids for the i-th auction, taking values from $\mathbb{N}_L \equiv \{2,\cdot\cdot\cdot,\bar{L}\}$. In our application, $\bar{L} = 3$.

Assume that the data $\{(B_{\ell i}, X_i, L_i) : \ell = 1,\ldots,L_i, i = 1,\ldots,n\}$ are i.i.d. over i and that $B_{\ell i}$'s are also i.i.d. over ℓ conditional on X_i and L_i. To implement the test, it is necessary

[28]Chen, Linton, and Van Keilegom (2003, Theorem 3) introduced its extension to functions indexed partly by infinite dimensional parameters, and called it local uniform L_2-continuity. For further discussions, see Andrews (1994) and Chen, Linton, and Van Keilegom (2003).

[29]Similarly one may derive primitive conditions for the second empirical example since it is also concerned with estimating conditional quantile functions. Hence we omit the details.

to estimate $\underline{b}$. In our application, we use $\hat{\underline{b}} = \min\{B_{\ell i} : \ell = 1, \ldots, L_i, i = 1, \ldots, n\}$, that is the overall sample minimum.

For each $x = (x_1, \ldots, x_d)$, the r-th order local polynomial quantile regression estimator of $q_k(\tau|x)$ can be obtained by minimizing

$$S_{n,x,\tau,k}(\gamma) \equiv \sum_{i=1}^{n} 1\{L_i = k\} \sum_{\ell=1}^{L_i} l_\tau \left[B_{\ell i} - \gamma^\top c\left(\frac{X_i - z}{h}\right)\right] K\left(\frac{x - X_i}{h}\right)$$

with respect to $\gamma \in \mathbf{R}^{|A_r|}$, where $l_\tau(u) \equiv \{|u| + (2\tau - 1)u\}/2$ for any $u \in \mathbf{R}$, and $K(\cdot)$ is a d-dimensional kernel function and h a bandwidth. More specifically, let $\widehat{q}_k(\tau|x) = \mathbf{e}_1^\top \hat{\gamma}_k(x)$, where $\hat{\gamma}_k(x) \equiv \arg\min_{\gamma \in \mathbf{R}^{|A_r|}} S_{n,x,\tau,k}(\gamma)$ and $\mathbf{e}_1$ is a column vector whose first entry is one, and the rest zero. Note that all bids are combined in each auction since we consider symmetric bidders. For $u = (u_1, \cdots, u_d)^\top \in A_r$, and $r + 1$ times differentiable map f on $\mathbf{R}^d$, we define the following derivative:

$$(D^u f)(x) \equiv \frac{\partial^{[u]}}{\partial x_1^{u_1} \cdots \partial x_d^{u_d}} f(x),$$

where $[u] = u_1 + \cdots + u_d$. Then we define $\gamma_{\tau,k}(x) \equiv (\gamma_{\tau,k,u}(x))_{u \in A_r}$, where

$$\gamma_{\tau,k,u}(x) \equiv \frac{1}{u_1! \cdots u_d!} D^u q_k(\tau|x).$$

5.2. **Primitive Conditions for the Example.** Let us present primitive conditions for the auction example of GPV.

Assumption AUC-1. *(i) There exists an integer $r > 3d/2 - 1$ and a constant $\varepsilon > 0$ such that for all $(\tau, k) \in \mathcal{T} \times \mathbb{N}_L$, $q_k(\tau|\cdot)$ is $r + 1$ times continuously differentiable on $\mathcal{S}_\tau(\varepsilon)$ with derivatives bounded uniformly over $(\tau, P) \in \mathcal{T} \times \mathcal{P}$.*
(ii) The density f of X is continuously differentiable on $\mathbf{R}^d$ with a derivative bounded uniformly over $P \in \mathcal{P}$.

Assumption AUC-2. *For each $k \in \mathbb{N}_L$, (i) $\inf_{x \in \mathcal{S}_\tau(\varepsilon)} f_{\tau,k}(0|x)$ is bounded away from zero uniformly over $(\tau, P) \in \mathcal{T} \times \mathcal{P}$, with $f_{\tau,k}(0|x)$ being the conditional density of $B_{\ell i} - q_k(\tau|X_i)$ given $X_i = x$ and $L_i = k$. (ii) $\sup_{x \in \mathcal{S}_\tau(\varepsilon)} f_{\tau,k}(0|x)$ is bounded uniformly over $(\tau, P) \in \mathcal{T} \times \mathcal{P}$, and (iii) $f_{\tau,k}(\bar{\varepsilon}|x)$ is continuously differentiable in $(\bar{\varepsilon}, x)$ with a derivative bounded uniformly over $x \in \mathcal{S}_\tau(\varepsilon)$, $\tau \in \mathcal{T}$, and $P \in \mathcal{P}$.*
(iv) $P\{L_i = k|X_i = x\}$ is bounded away from zero uniformly over $x \in \mathcal{S}_\tau(\varepsilon), \tau \in \mathcal{T}$ and $P \in \mathcal{P}$, and is continuously differentiable in x with a derivative bounded uniformly over $x \in \mathcal{S}_\tau(\varepsilon), \tau \in \mathcal{T}$ and $P \in \mathcal{P}$.

Assumption AUC-3. *(i) K is compact-supported, nonnegative, bounded, and Lipschitz continuous on the interior of its support, $\int K(u)du = 1$, and $\int K(u) ||u||^2 du > 0$. (ii) As*

$n \to \infty$,

$$n^{-1/4}h^{-3(d+\nu)/4} + \sqrt{n}h^{r+1} \to 0,$$

for some small $\nu > 0$.

Assumption AUC-4. $\hat{\underline{b}} = \underline{b} + o_P\left(n^{-1/2}\right)$, $\mathcal{P}$*-uniformly.*

Assumption AUC-1(i) is a standard assumption used in the local polynomial approach where one approximates $q_k(\cdot|x)$ by a linear combination of its derivatives through Taylor expansion, except only that the approximation behaves well uniformly over $P \in \mathcal{P}$. Assumption AUC-2 is made to prevent the degeneracy of the asymptotic linear representation of $\hat{\gamma}_{\tau,k}(x) - \gamma_{\tau,k}(x)$ that is uniform over $x \in \mathcal{S}_\tau(\varepsilon), \tau \in \mathcal{T}$ and over $P \in \mathcal{P}$. Assumptions AUC-3 (i) and (ii) are conditions for the kernel and the bandwidth. For example, the choice of $h = n^{-s}$ with the condition $1/(2(r+1)) < s < 1/(3(d+\nu))$ satisfies the bandwidth condition. The small $\nu > 0$ there is introduced to satisfy Assumption B4. Assumption AUC-4 holds in general because the extreme order statistic is super-consistent with the n^{-1} rate of convergence. Recall that $\mathbf{e}_1$ is a unit vector whose first element is one and all other elements are zeros.

Theorem AUC. *If Assumptions AUC-1, AUC-2, AUC-3, and AUC-4 hold, then Assumptions A1-A3, A5-A6, and B1-B4 hold with the following definitions:* $J = 2$, $r_{n,j} \equiv \sqrt{nh^d}$,

$$\begin{aligned}
v_{n,\tau,1}(x) &\equiv \mathbf{e}_1^\top\{\gamma_{\tau,2}(x) - \gamma_{\tau,3}(x)\}, \\
v_{n,\tau,2}(x) &\equiv \underline{b} - \mathbf{e}_1^\top\{2\gamma_{\tau,2}(x) - \gamma_{\tau,3}(x)\}, \\
\beta_{n,x,\tau,1}(Y_i, z) &\equiv \alpha_{n,x,\tau,2}(Y_i, z) - \alpha_{n,x,\tau,3}(Y_i, z), \text{ and} \\
\beta_{n,x,\tau,2}(Y_i, z) &\equiv -2\alpha_{n,x,\tau,2}(Y_i, z) + \alpha_{n,x,\tau,3}(Y_i, z),
\end{aligned}$$

where $\tilde{l}_\tau(u) \equiv \tau - 1\{u \le 0\}$, $Y_i = \{(B_{\ell i}, L_i) : \ell = 1, \ldots, L_i\}$,

$$\begin{aligned}
\alpha_{n,x,\tau,k}(Y_i, z) &\equiv -1\{L_i = k\} \sum_{l=1}^{k} \tilde{l}_\tau\left(B_{\ell i} - \gamma_{\tau,k}^\top(x) \cdot H \cdot c(z)\right) \mathbf{e}_1^\top M_{n,\tau,k}^{-1}(x) c(z) K(z), \\
M_{n,\tau,k}(x) &\equiv k \int P\{L_i = k | X_i = x + th\} f_{\tau,k}(0|x+th) f(x+th) K(t) c(t) c^\top(t) dt,
\end{aligned}$$

and $H =$*diag*$((h^{|u|})_{u \in A_r})$ *is the* $|A_r| \times |A_r|$ *diagonal matrix.*

It is worth commenting on the linear expansion derived in Theorem AUC. The term $\alpha_{n,x,\tau,k}(Y_i, z)$ is not mean zero conditionally on X_i since the bias terms are included inside $\tilde{l}_\tau(\cdot)$. Also, note that $M_{n,\tau,k}(x)$ depends on n and contains the smoothing bias as well. However, the results obtained in Theorem AUC are sufficient enough to verify high-level conditions of the paper.

The main part of the proof is to establish a uniform error rate for the asymptotic linear representation for $\sqrt{nh^d}\{\hat{\gamma}_{\tau,k}(x) - \gamma_{\tau,k}(x)\}$ in a spirit similar to Guerre and Sabbah (2012).[30] Our proof uses some arguments of Guerre and Sabbah (2012), who employ a maximal inequality of Massart (2007, Theorem 6.8).[31]

The theoretical novelty in our derivation of the linear expansion in Theorem AUC is that we have obtained an approximation that is uniform in (x, τ) as well as in P. To the best of our knowledge, there is no established result on linear expansions of local polynomial quantile regression estimators that hold uniformly over three aspects (x, τ, P) simultaneously. Therefore, our results may be of independent interest and can be useful in other contexts.

6. Power Properties

In this section, we go back to the general setup in Section 4 and consider the power properties of the bootstrap test. In Section 6.1, we establish the consistency of our test. Section 6.2 provides heuristic arguments behind local power properties of our tests, and Section 6.3 presents the local power function in a general form.[32]

6.1. **Consistency.**

First, to show consistency of our test, we make the following assumption.

Assumption C1. *For each $j \in \mathbb{N}_J$ and $(x, \tau) \in \mathcal{S}$, $v_{n,\tau,j}(x) = v_{\tau,j}(x) + o(1)$, and*

$$\limsup_{n \to \infty} \sup_{(x,\tau) \in \mathcal{S}} |v_{n,\tau,j}(x)| < \infty. \tag{6.1}$$

The pointwise convergence $v_{n,\tau,j}(x) = v_{\tau,j}(x) + o(1)$ holds typically by an appropriate choice of $v_{n,\tau,j}(x)$. In many examples, condition (6.1) is often implied by Assumptions A1-A6. If we revisit the simple example considered in Section 2.4, it is straightforward to see that under Assumptions A1-A6, with the subscript τ suppressed, $v_{n,1}(x) = v_1(x) + o(1)$, where $v_{n,1}(x) \equiv \mathbf{E}\hat{v}_{n,1}(x)$ and $v_1(x) \equiv \mathbf{E}(Y|X = x)f(x)$, and (6.1) holds easily.

We now establish the consistency of our proposed test as follows.

---

[30]See Lemma QR2 in Appendix B. Our asymptotic approximation is based on plugging the asymptotic linear expansion directly. There is a recent proposal by Mammen, Van Keilegom, and Yu (2013), who developed nonparametric tests for parametric specifications of regression quantiles and showed that calculating moments of linear expansions of nonparametric quantile regression estimators might work better in a sense that their approach requires less stringent conditions for the dimension of covariates and the choice of the bandwidth. It is an interesting future research topic whether their ideas can be applied to our setup.

[31]The main significant difference is that the convergence rate obtained by Guerre and Sabbah (2012) is uniform over h in some interval, while our result is uniform over $P \in \mathcal{P}$.

[32]The local power results in this section are more general than those of Lee, Song, and Whang (2013). In particular, the results accommodate a wider class of local alternatives that may not converge to the least favorable case.

Theorem 3. *Suppose that Assumptions A1-A6, B1-B4, and C1 hold and that we are under a fixed alternative hypothesis such that*

$$\int \Lambda_p \left(v_{\tau,1}(x), \cdots, v_{\tau,J}(x) \right) dQ(x,\tau) > 0.$$

Then as $n \to \infty$,

$$P\{\hat{\theta} > c^*_{\alpha,\eta}\} \to 1.$$

6.2. **Local Power Analysis: Definitions and Heuristics.** In this section, we investigate the local power properties of our test. For local power analysis, we formally define the space of Pitman directions. Let $\mathcal{D}$ be the collection of $\mathbf{R}^J$-valued bounded functions on $\mathcal{X} \times \mathcal{T}$ such that for each $\delta = (\delta_1, \cdots, \delta_J) \in \mathcal{D}$, $Q\{(x,\tau) \in \mathcal{S} : \delta_j(x,\tau) \neq 0\} > 0$ for some $j = 1, \ldots, J$. That is, at least one of the components of any $\delta \in \mathcal{D}$ is a non-zero function a.e. For each $\delta = (\delta_1, \cdots, \delta_J) \in \mathcal{D}$, we write $\delta_{\tau,j}(x) = \delta_j(x,\tau)$, $j = 1, \cdots, J$.

For a given vector of sequences $b_n = (b_{n,1}, \cdots, b_{n,J})$, such that $b_{n,j} \to \infty$, and $\delta \in \mathcal{D}$, we consider the following type of local alternatives:

$$H_\delta : v_{\tau,j}(x) = v^0_{\tau,j}(x) + \frac{\delta_{\tau,j}(x)}{b_{n,j}}, \text{ for all } j \in \mathbb{N}_J, \tag{6.2}$$

where $v^0_{\tau,j}(x) \leq 0$ for all $(x,\tau,j) \in \mathcal{X} \times \mathcal{T} \times \mathbb{N}_J$, $\delta_{\tau,j}(x) > 0$ for some $(x,\tau,j) \in \mathcal{X} \times \mathcal{T} \times \mathbb{N}_J$ such that $v_{\tau,j}(x) > 0$ for some $(x,\tau,j) \in \mathcal{X} \times \mathcal{T} \times \mathbb{N}_J$. Note that in (6.2), $v_{\tau,j}(x)$ is a sequence of Pitman local alternatives that consist of three components: $v^0_{\tau,j}(x)$, b_n, and $\delta_{\tau,j}(x)$.

The first component $v^0_{\tau,j}(x)$ determines where the sequence of local alternatives converges to. For example, if $v^0_{\tau,j}(x) \equiv 0$ for all (x,τ,j), then we have a sequence of local alternatives that converges to the least favorable case. We allow for negative values for $v^0_{\tau,j}(x)$, so that we include the local alternatives that do not converge to the least favorable case as well.

From here on, we assume the local alternative hypotheses of the form in (6.2). We fix $v^0_{\tau,j}(x)$ and identify each local alternative with a pair (b_n, δ) for each Pitman direction $\delta \in \mathcal{D}$. The following definitions are useful to explain our local power results.

Definition 3. (i) Given a Pitman direction $\delta \in \mathcal{D}$, we say that an α-level test, $1\{T > c_\alpha\}$, *has nontrivial local power against* (b_n, δ), if under the local alternatives (b_n, δ),

$$\text{liminf}_{n\to\infty} \, P\left\{T > c_\alpha\right\} > \alpha,$$

and say that the test has *trivial local power against* (b_n, δ), if under the local alternatives (b_n, δ),

$$\text{limsup}_{n\to\infty} \, P\left\{T > c_\alpha\right\} \leq \alpha.$$

(ii) Given a collection $\mathcal{D}$, we say that a test *has convergence rate* b_n *against* $\mathcal{D}$, if the test has nontrivial local power against (b_n, δ) for some $\delta \in \mathcal{D}$, and has trivial local power against (b_n', δ) for all $\delta \in \mathcal{D}$ and all b_n' such that $b_{n,j}'/b_{n,j} \to \infty$ as $n \to \infty$, for all $j = 1, \ldots, J$.

One of the remarkable aspects of the local power properties is that our test has *two types of convergence rates*. More specifically, there exists a partition $(\mathcal{D}_1, \mathcal{D}_2)$ of $\mathcal{D}$, where our test has a rate b_n against $\mathcal{D}_1$ and another rate b_n' against $\mathcal{D}_2$. Furthermore, in many nonparametric inequality testing environments, the faster of the two rates b_n and b_n' achieves the parametric rate of $\sqrt{n}$.

To see this closely, let us assume the set-up of testing inequality restrictions on a mean regression function in Section 2.4, and consider the following local alternatives:

$$v_{n,1}(x) = v_0(x) + \frac{\delta(x)}{b_n}, \tag{6.3}$$

where $v_0(x) \leq 0$ for all $x \in \mathcal{X}$, and $\delta \in \mathcal{D}$.

First, we set $b_n = \sqrt{n}$. Then under this local alternative hypothesis (b_n, δ), we can verify that with probability approaching one,

$$h^{-1/2}(\hat{\theta} - a_{n,0}) = h^{-1/2} \left\{ \int_{B_n^0(c_n)} \left[Z_{n,1}(x) + \frac{\sqrt{nh} v_0(x)}{\hat{\sigma}_1(x)} + \frac{h^{1/2}\delta(x)}{\hat{\sigma}_1(x)} \right]_+ dx - a_{n,0} \right\}, \tag{6.4}$$

where $Z_{n,1}(x) = \sqrt{nh}\left\{\hat{v}_1(x) - v_{n,1}(x)\right\}/\hat{\sigma}_1(x)$, $B_n^0(c_n) = \left\{ x \in \mathcal{X} : \left|\sqrt{nh} v_0(x)\right| \leq c_n \right\}$, $c_n \to \infty$, $\sqrt{\log n}/c_n \to 0$, and

$$a_{n,0} = \mathbf{E}\left[\int_{B_n^0(c_n)} \left[Z_{n,1}(x)\right]_+ dx \right].$$

Under regularity conditions, the right-hand side of (6.4) is approximated by

$$h^{-1/2} \left\{ \int_{B^0(0)} \left[Z_{n,1}(x) + \frac{h^{1/2}\delta(x)}{\sigma_1(x)} \right]_+ dx - a_{n,\delta} \right\} + h^{-1/2}\left\{a_{n,\delta} - a_{n,0}\right\}, \tag{6.5}$$

where $B^0(0) = \{x \in \mathcal{X} : v_0(x) = 0\}$ and

$$a_{n,\delta} = \mathbf{E}\left[\int_{B^0(0)} \left[Z_{n,1}(x) + \frac{h^{1/2}\delta(x)}{\sigma_1(x)} \right]_+ dx \right].$$

The leading term in (6.5) converges in distribution to $\mathbb{Z}_1 \sim N(0, \sigma_0^2)$ precisely as in (2.11). Furthermore, we can show that

$$\begin{aligned} a_{n,\delta} &= \int_{B^0(0)} \mathbf{E}\left[\mathbb{Z}_1 + \frac{h^{1/2}\delta(x)}{\sigma_1(x)} \right]_+ dx + o(h^{1/2}) \text{ and} \\ a_{n,0} &= \int_{B^0(0)} \mathbf{E}\left[\mathbb{Z}_1\right]_+ dx + o(h^{1/2}). \end{aligned}$$

Therefore, as for the last term in (6.5), we find that

$$\begin{aligned} h^{-1/2}\left\{a_{n,\delta}-a_{n,0}\right\} &= \int_{B^0(0)} h^{-1/2}\left(\mathbf{E}\left[\mathbb{Z}_1+\frac{h^{1/2}\delta(x)}{\sigma_1(x)}\right]_+ - \mathbf{E}\left[\mathbb{Z}_1\right]_+\right)dx \\ &= 2\phi(0)\int_{B^0(0)}\frac{\delta(x)}{\sigma_1(x)}dx+o(1), \end{aligned}$$

where the last equality follows from expanding $h^{-1/2}\left\{\mathbf{E}\left[\mathbb{Z}_1+h^{1/2}\delta(x)/\sigma_1(x)\right]_+ - \mathbf{E}\left[\mathbb{Z}_1\right]_+\right\}$. We conclude that under the local alternatives, we have

$$h^{-1/2}(\hat{\theta}-a_{n,0}) \quad \to_d \quad \mathbb{Z}_1+2\phi(0)\int_{B^0(0)}\frac{\delta(x)}{\sigma_1(x)}dx.$$

The magnitude of the last term in the limit determines the local power of the test. Thus under Pitman local alternatives such that

$$\int_{B^0(0)}\frac{\delta(x)}{\sigma_1(x)}dx>0, \tag{6.6}$$

the test has nontrivial power against $\sqrt{n}$-converging Pitman local alternatives. Note that the integral in (6.6) is defined on the population contact set $B^0(0)$. Thus, the test has nontrivial power, unless the contact set has Lebesgue measure zero or $\delta(\cdot)$ is "too often negative" on the contact set.

When the integral in (6.6) is zero, we consider the local alternatives (b_n,δ) with a slower convergence rate $b_n=n^{1/2}h^{1/4}$. Following similar arguments as before, we now have

$$h^{-1/2}(\hat{\theta}-a_{n,0}) \to_d \mathbb{Z}_1+\lim_{n\to\infty}h^{-1/2}\left\{\bar{a}_{n,\delta}-a_{n,0}\right\},$$

where

$$\bar{a}_{n,\delta}=\int_{B^0(0)}\mathbf{E}\left[Z_{n,1}(x)+\frac{h^{1/4}\delta(x)}{\sigma_1(x)}\right]_+dx,$$

which can be shown again to be equal to

$$\int_{B^0(0)}\mathbf{E}\left[\mathbb{Z}_1+\frac{h^{1/4}\delta(x)}{\sigma_1(x)}\right]_+dx+o(h^{1/2}).$$

However, observe that

$$\begin{aligned} & h^{-1/2}\int_{B^0(0)}\left\{\mathbf{E}\left[\mathbb{Z}_1+\frac{h^{1/4}\delta(x)}{\sigma_1(x)}\right]_+ - \mathbf{E}\left[\mathbb{Z}_1\right]_+\right\}dx \\ = {} & h^{-1/4}2\phi(0)\int_{B^0(0)}\frac{\delta(x)}{\sigma_1(x)}dx+\frac{1}{2}\int_{B^0(0)}\frac{\delta^2(x)}{\sigma_1^2(x)}dx+o(1) \\ = {} & \frac{1}{2}\int_{B^0(0)}\frac{\delta^2(x)}{\sigma_1^2(x)}dx+o(1), \end{aligned}$$

because $\int_{B^0(0)}\{\delta(x)/\sigma_1(x)\}dx = 0$. We find that under the local alternative hypothesis in (6.3) with $b_n = n^{1/2}h^{1/4}$,

$$h^{-1/2}(\hat{\theta} - a_{n,0}) \quad \to_d \quad \mathbb{Z}_1 + \frac{1}{2}\int_{B^0(0)} \frac{\delta^2(x)}{\sigma_1^2(x)}dx.$$

Therefore, even when $\int_{B^0(0)}\{\delta(x)/\sigma_1(x)\}dx = 0$, the test still has nontrivial power against $n^{1/2}h^{1/4}$-converging Pitman local alternatives, if the Pitman directions are such that

$$\int_{B^0(0)} \{\delta^2(x)/\sigma_1^2(x)\}dx > 0.$$

Now let us consider the partition $(\mathcal{D}_1, \mathcal{D}_2)$ of $\mathcal{D}$, where

$$\begin{aligned}\mathcal{D}_1 &= \left\{\delta \in \mathcal{D} : \int_{B^0(0)} \delta(x)/\sigma_1(x)dx \neq 0\right\} \text{ and} \\ \mathcal{D}_2 &= \left\{\delta \in \mathcal{D} : \int_{B^0(0)} \delta(x)/\sigma_1(x)dx = 0 \text{ and } \int_{B^0(0)} \{\delta^2(x)/\sigma_1^2(x)\}dx > 0\right\}.\end{aligned}$$

When $\inf_{x\in\mathcal{X}}\sigma_1^2(x) > c > 0$ for some $c > 0$ (recall Assumption A5) and $Q(B^0(0)) > 0$, we have $\int_{B^0(0)}\{\delta^2(x)/\sigma_1^2(x)\}dx > 0$ and the set $\{\mathcal{D}_1, \mathcal{D}_2\}$ becomes a partition of $\mathcal{D}$. Thus the bootstrap test has a convergence rate of $\sqrt{n}$ against $\mathcal{D}_1$ and $n^{1/2}h^{1/4}$-rate against $\mathcal{D}_2$. In the next section, Corollary 1 provides a general result of this phenomenon of dual convergence rates of our bootstrap test.

6.3. **Local Power Analysis: Results.** We now provide general local power functions explicitly. We first present explicit forms of location and scale normalizers, $a_n(q_n)$ and $\sigma_n(q_n)$ in (4.9). Let for $j, k \in \mathbb{N}_J$, and $\tau_1, \tau_2 \in \mathcal{T}$,

$$\rho_{n,\tau_1,\tau_2,j,k}(x,u) \equiv \frac{1}{h^d}\mathbf{E}\left[\beta_{n,x,\tau_1,j}\left(Y_{ij}, \frac{X_i - x}{h}\right)\beta_{n,x,\tau_2,k}\left(Y_{ik}, \frac{X_i - x}{h} + u\right)\right]. \tag{6.7}$$

This function approximates the asymptotic covariance between $\sqrt{n}(\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x))/\hat{\sigma}_{\tau,j}(x)$ and $\sqrt{n}(\hat{v}_{\tau,j}(x+uh) - v_{n,\tau,j}(x+uh))/\hat{\sigma}_{\tau,j}(x)$. We define $\Sigma_{n,\tau_1,\tau_2}(x,u)$ to be the J-dimensional square matrix with (j,k)-th entry given by $\rho_{n,\tau_1,\tau_2,j,k}(x,u)$.

Define for $\mathbf{v} \in \mathbf{R}^J$,

$$\bar{\Lambda}_{x,\tau}(\mathbf{v}) \equiv \sum_{A\in\mathcal{N}_J} \Lambda_{A,p}(\mathbf{v})1\left\{(x,\tau) \in B_{n,A}(q_n)\right\}.$$

Then we define

$$a_n(q_n) \equiv \int_{\mathcal{X}}\int_{\mathcal{T}} \mathbf{E}\left[\bar{\Lambda}_{x,\tau_1}(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0))\right]d\tau dx,$$

and

$$\sigma_n^2(q_n) \equiv \int_{\mathcal{U}}\int_{\mathcal{X}}\int_{\mathcal{T}}\int_{\mathcal{T}} C_{n,\tau_1,\tau_2}(x,u)d\tau_1 d\tau_2 dx du, \tag{6.8}$$

where

$$C_{n,\tau_1,\tau_2}(x,u) \equiv Cov\left(\bar{\Lambda}_{x,\tau_1}(\mathbb{W}^{(1)}_{n,\tau_1,\tau_2}(x,u)), \bar{\Lambda}_{x,\tau_2}(\mathbb{W}^{(2)}_{n,\tau_1,\tau_2}(x,u))\right),$$

and $[\mathbb{W}^{(1)}_{n,\tau_1,\tau_2}(x,u)^\top, \mathbb{W}^{(2)}_{n,\tau_1,\tau_2}(x,u)^\top]^\top$ is a mean zero $\mathbf{R}^{2J}$-valued Gaussian random vector whose covariance matrix is given by

$$\left[\begin{array}{cc} \Sigma_{n,\tau_1,\tau_1}(x,0) & \Sigma_{n,\tau_1,\tau_2}(x,u) \\ \Sigma_{n,\tau_1,\tau_2}(x,u)^\top & \Sigma_{n,\tau_2,\tau_2}(x+uh,0) \end{array}\right]. \tag{6.9}$$

The multiple integral in (6.8) is nonnegative.

The limit of the quantity $\sigma_n^2(q_n)$ as $n \to \infty$, if it is positive, is nothing but the asymptotic variance of the test statistic $\hat{\theta}$ (after location-scale normalization). Not surprisingly the asymptotic variance does not depend on points (x,τ) of $\mathcal{X} \times \mathcal{T}$ such that $v_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x)$ is away below zero, as is expressed through its dependence on the contact sets $B_{n,A}(q_n)$ and the "truncated map" $\bar{\Lambda}_{x,\tau}(\cdot)$ involving A's restricted to $\mathcal{N}_J$.

We first make the following assumptions.

Assumption C2. *(i) For each $(\tau,j) \in \mathcal{T} \times \mathbb{N}_J$, there exists a map $v^0_{n,\tau,j} : \mathbf{R}^d \to \mathbf{R}$ such that for each $x \in \mathcal{S}_\tau(\varepsilon_1)$, $v^0_{n,\tau,j}(x) \le 0$, and*

$$v_{n,\tau,j}(x) = v^0_{n,\tau,j}(x) + \frac{\delta_{\tau,j}(x)}{b_{n,j}}\left(1 + o(1)\right), \tag{6.10}$$

where $o(1)$ is uniform in $x \in \mathcal{S}_\tau$ and in $\tau \in \mathcal{T}$, as $n \to \infty$ and $b_{n,j} \to \infty$ is the positive sequence in (6.2).
(ii) $\sup_{(x,\tau)\in\mathcal{S}} |\sigma_{n,\tau,j}(x) - \sigma_{\tau,j}(x)| = o(1)$, as $n \to \infty$, for some function $\sigma_{\tau,j}(x)$ such that $\inf_{(x,\tau)\in\mathcal{S}} \sigma_{\tau,j}(x) > 0$.

Assumption C2 can also be shown to hold in many examples. When appropriate smoothness conditions for $v_{\tau,j}(x)$ hold and a suitable (possibly higher-order) kernel function is used, we can take $v_{n,\tau,j}(x)$ in Assumption A1 to be identical to $v_{\tau,j}(x)$, and hence Assumption C2 is implied by (6.2). For the simple example in Section 2.4, if we take $v_{n,j}(x) = \mathbf{E}\hat{v}_j(x)$, it follows that $v_{n,j}(x) = v^0_{n,j}(x) + b^{-1}_{n,j}\int \delta_j(x+zh)K(z)dz$, with $v^0_{n,j}(x) = \int v^0_j(x+zh)K(z)dz$. Hence when $\delta_j(x)$ is uniformly continuous in x, we obtain Assumption C2.

The local asymptotic power function is based on the asymptotic normal approximation of the distribution of $\hat{\theta}$ (after scale and location normalization) under the local alternatives. For this purpose, we define the sequence of probability sets that admit the normal approximation under local alternatives. For $c_1, c_2 > 0$, let $B^0_n(c_1,c_2)$ and $B^0_{n,A}(c_1,c_2)$ denote $B_n(c_1,c_2)$ and $B_{n,A}(c_1,c_2)$ except that $v_{n,\tau,j}(x)$'s are replaced by $v^0_{n,\tau,j}(x)$'s in Assumption C2. As before, we write $B^0_n(c) \equiv B^0_n(c,c)$.

Definition 4. For any positive sequence $\lambda_n \to 0$, define

$$\mathcal{P}_n^0(\lambda_n) \equiv \left\{ P \in \tilde{\mathcal{P}}_n^0(\lambda_n) : \sigma_n^2(0) \geq \eta/\Phi^{-1}(1-\alpha) \right\},$$

where $\tilde{\mathcal{P}}_n^0(\lambda_n)$ is equal to $\tilde{\mathcal{P}}_n(\lambda_n, q_n)$ except that $B_{n,A}(c_{n,U}, c_{n,L})$ and $B_{n,A}(q_n)$ are replaced by $B_{n,A}^0(c_{n,U}, c_{n,L})$ and $B_{n,A}^0(q_n)$ for all $A \in \mathcal{N}_J$, and q_n is set to be zero.

To give a general form of the local power function, let us define $\psi_{n,A,\tau}(\cdot; x) : \mathbf{R}^J \to [0, \infty)$, $(x, \tau) \in \mathcal{X} \times \mathcal{T}$ and $A \subset \mathbb{N}_J$, as

$$\psi_{n,A,\tau}(\mathbf{y}; x) = \frac{1}{\sigma_n(0)} \mathbf{E}\left[\Lambda_{A,p}\left(\mathbb{W}_{n,\tau,\tau}^{(1)}(x,0) + \mathbf{y}\right)\right] \cdot 1\left\{(x,\tau) \in B_{n,A}^0(0)\right\}.$$

The local power properties of the bootstrap test are mainly determined by the slope and the curvature of this function. So, we define

$$\psi_{n,A,\tau}^{(1)}(\mathbf{y}; x) \equiv \frac{\partial}{\partial \mathbf{y}} \psi_{n,A,\tau}(\mathbf{y}; x) \text{ and } \psi_{n,A,\tau}^{(2)}(\mathbf{y}; x) \equiv \frac{\partial^2}{\partial \mathbf{y} \partial \mathbf{y}^\top} \psi_{n,A,\tau}(\mathbf{y}; x), \tag{6.11}$$

if the first derivatives and the second derivatives in the definition exist respectively.

Assumption C3. *(i) There exists* $\varepsilon_1 > 0$ *such that for all* $(\tau, A) \in \mathcal{T} \times \mathcal{N}_J$ *and all* x *in the interior of* $\mathcal{S}_\tau(\varepsilon_1)$, $\psi_{n,A,\tau}^{(1)}(\mathbf{0}; x)$ *exists for all* $n \geq 1$ *and*

$$\psi_{A,\tau}^{(1)}(\mathbf{0}; x) \equiv \lim_{n\to\infty} \psi_{n,A,\tau}^{(1)}(\mathbf{0}; x)$$

exists, and $\limsup_{n\to\infty} \sup_{(x,\tau)\in\mathcal{S}} |\psi_{n,A,\tau}^{(1)}(0; x)| < C$ *for some* $C > 0$.
(ii) There exists $\varepsilon_1 > 0$ *such that for all* $(\tau, A) \in \mathcal{T} \times \mathcal{N}_J$ *and all* x *in the interior of* $\mathcal{S}_\tau(\varepsilon_1)$, $\psi_{n,A,\tau}^{(2)}(\mathbf{0}; x)$ *exists for all* $n \geq 1$ *and*

$$\psi_{A,\tau}^{(2)}(\mathbf{0}; x) \equiv \lim_{n\to\infty} \psi_{n,A,\tau}^{(2)}(\mathbf{0}; x)$$

exists, and $\limsup_{n\to\infty} \sup_{(x,\tau)\in\mathcal{S}} |\psi_{n,A,\tau}^{(2)}(\mathbf{0}; x)| < C$ *for some* $C > 0$.

To appreciate Assumption C3, consider the case where $J = 2$, $A = \{1, 2\}$, and $\mathbb{W}_{n,\tau,\tau}^{(1)}(x, 0)$ has a distribution denoted by G_n. Choose $y_1 \geq y_2$ without loss of generality. We take $\Lambda_p(v_1, v_2) = \max\{v_1, v_2, 0\}^p$. Then we can write $\mathbf{E}[\Lambda_{A,p}(\mathbb{W}_{n,\tau,\tau}^{(1)}(x, 0) + \mathbf{y})]$ as

$$\begin{aligned}
&\int_{\mathbf{R}^2} (w_1 + y_1)^p 1\left\{w_1 \in [w_2 + y_2 - y_1, \infty) \text{ and } w_2 \in [-y_2, \infty)\right\} dG_n(w_1, w_2) \\
&+ \int_{\mathbf{R}^2} (w_2 + y_2)^p \, 1\left\{w_1 \in (-\infty, w_2 + y_2 - y_1) \text{ and } w_2 \in [-y_2, \infty)\right\} dG_n(w_1, w_2) \\
&+ \int_{\mathbf{R}^2} (w_1 + y_1)^p 1\left\{w_1 \in [-y_1, \infty) \text{ and } w_2 \in (-\infty, -y_2)\right\} dG_n(w_1, w_2).
\end{aligned}$$

Certainly the three quantities are all differentiable in (y_1, y_2).

The following theorem offers the local power function of the bootstrap test in a general form.

Theorem 4. *Suppose that Assumptions A1-A6, B1-B4, C1-C2, and C3(i) hold and that*

$$h^{-d/2} (\log n)^{p/2} \lambda_n \to 0, \tag{6.12}$$

as $n \to \infty$. Then for each sequence $P_n \in \mathcal{P}_n^0(\lambda_n)$, $n \geq 1$, which satisfies the local alternative hypothesis (b_n, δ) for some $\delta \in \mathcal{D}$ with $b_n = (r_{n,j} h^{-d/2})_{j=1}^J$,

$$\lim_{n\to\infty} P_n\{\hat{\theta} > c^*_{\alpha,\eta}\} = 1 - \Phi\left(z_{1-\alpha} - \mu_1(\delta)\right),$$

where Φ denotes the standard normal cdf,

$$\mu_1(\delta) \equiv \sum_{A\in\mathcal{N}_J} \int \psi_{A,\tau}^{(1)}(\mathbf{0}; x)^\top \delta_{\tau,\sigma}(x) dQ(x,\tau),$$

and

$$\delta_{\tau,\sigma}(x) \equiv \left(\frac{\delta_{\tau,1}(x)}{\sigma_{\tau,1}(x)}, \cdots, \frac{\delta_{\tau,J}(x)}{\sigma_{\tau,J}(x)}\right). \tag{6.13}$$

Theorem 4 shows that if we take b_n such that $b_{n,j} = r_{n,j} h^{-d/2}$ for each $j = 1, \ldots, J$, the local asymptotic power of the test against (b_n, δ) is determined by the shift $\mu_1(\delta)$. Thus, the bootstrap test has nontrivial local power against (b_n, δ) if and only if

$$\mu_1(\delta) > 0.$$

The test is asymptotically biased against (b_n, δ) such that $\mu_1(\delta) < 0$.

Suppose that

$$\mu_1(\delta) = 0, \tag{6.14}$$

for all $A \in \mathcal{N}_J$, i.e., when $\delta_{\tau,\sigma}$ has positive and negative parts which precisely cancels out in the integration. Then, we show that the bootstrap test has nontrivial asymptotic power against local alternatives that converges at a rate slower than $n^{-1/2}$ to the null hypothesis.

Theorem 5. *Suppose that the conditions of Theorem 4 and Assumption C3(ii) hold. Then for each sequence $P_n \in \mathcal{P}_n^0(\lambda_n)$, $n \geq 1$, which satisfies the local alternative hypothesis (b_n, δ) for some $\delta \in \mathcal{D}$ such that $\mu_1(\delta) = 0$ and $b_n = (r_{n,j} h^{-d/4})_{j=1}^J$,*

$$\lim_{n\to\infty} P_n\{\hat{\theta} > c^*_{\alpha,\eta}\} = 1 - \Phi\left(z_{1-\alpha} - \mu_2(\delta)\right),$$

where

$$\mu_2(\delta) \equiv \frac{1}{2} \sum_{A\in\mathcal{N}_J} \int \delta_{\tau,\sigma}^\top(x) \psi_{A,\tau}^{(2)}(\mathbf{0}; x) \delta_{\tau,\sigma}(x) dQ(x,\tau).$$

The local power function depends on the limit of the curvature of the function $\psi_{n,A,\tau}(\mathbf{y};x)$ at $\mathbf{y}=\mathbf{0}$, for all $A \in \mathcal{N}_J$. When the function is strictly concave at $\mathbf{0}$ in the limit, $\psi_{A,\tau}^{(2)}(\mathbf{0};x)$ is positive definite on $\mathcal{X} \times \mathcal{T}$, and in this case, the bootstrap test has nontrivial power whenever $\delta_{\tau,\sigma}(x)$ is nonzero on a set whose intersection with $B_n^0(0)$ has Lebesgue measure greater than $c > 0$ for all $n \geq 1$, for some $c > 0$.

From Theorems 4 and 5, it is seen that the phenomenon of dual convergence rates generally hold for our tests. To formally state the result, define

$$\mathcal{D}_1 \equiv \{\delta \in \mathcal{D} : \mu_1(\delta) \neq 0\} \text{ and}$$
$$\mathcal{D}_2 \equiv \{\delta \in \mathcal{D} : \mu_1(\delta) = 0 \text{ and } \mu_2(\delta) > 0\}.$$

When $\liminf_{n\to\infty} Q(B_n^0(0)) > 0$, the set $\{\mathcal{D}_1, \mathcal{D}_2\}$ becomes a partition of the space of Pitman directions $\mathcal{D}$.

Corollary 1. *Suppose that the conditions of Theorem 5 hold. Then the bootstrap test has convergence rate $b_n = (r_{n,j}h^{-d/2})_{j=1}^J$ against $\mathcal{D}_1$, and convergence rate $b_n = (r_{n,j}h^{-d/4})_{j=1}^J$ against $\mathcal{D}_2$.*

When $r_{n,j}$'s diverge to infinity at the usual nonparametric rate $r_{n,j} = n^{1/2}h^{d/2}$ as in many kernel-based estimators, the test has a parametric rate of convergence $b_n = \sqrt{n}$ and nontrivial local power against $\mathcal{D}_1$. However, the test has a convergence rate slower than the parametric rate against $\mathcal{D}_2$.

When $r_{n,j}$'s diverge slower than the rate $n^{1/2}h^{d/2}$ as in the case of kernel-based derivative estimators, the test has a convergence rate slower than the parametric rate. In Appendix II.2, we present several nonparametric tests for monotonicity where $d = 1$, $J = 1$, and $r_{n,1} = n^{1/2}h^{3/2}$. In this case, the monotonicity tests have convergence rate with $b_n = n^{1/2}h$ against $\mathcal{D}_1$, and convergence rate with $b_n = n^{1/2}h^{5/4}$ against $\mathcal{D}_2$.

7. Conclusions

In this paper, we have proposed a general method for testing inequality restrictions on nonparametric functions and have illustrated its usefulness by looking at two particular empirical applications. We regard our examples as just some illustrative applications and believe that our framework can be useful in a number of other settings.

Our bootstrap test is based on a one-sided version of L_p functionals of kernel-type estimators ($1 \leq p < \infty$). We have provided regularity conditions under which the bootstrap test is asymptotically valid uniformly over a large class of distributions and have also provided a class of distributions for which the asymptotic size is exact. We have shown the consistency of our test and have obtained a general form of the local power function.

There are different notions of efficiency for nonparametric tests and hence there is no compelling sense of an asymptotically optimal test for the hypothesis considered in this paper. See Nitikin (1995) and Bickel, Ritov, and Stoker (2006) for a general discussion. It would be interesting to consider a multiscale version of our test based on a range of bandwidths to see if it achieves adaptive rate-optimality against a sequence of smooth alternatives along the lines of Armstrong and Chan (2013) and Chetverikov (2011).

Appendices

Appendix I reports the results of Monte Carlo experiments, and Appendix II presents more examples of testing problems that can be included in our general framework. Appendix A gives the proofs of Theorems 1-5, Appendix B provides the proof of Theorem AUC, and Appendices C and D offer auxiliary results for the proofs of Theorems 1-5.

I. Monte Carlo Experiments

This part of the appendix reports the finite-sample performance of our proposed test for the Monte Carlo design considered in Andrews and Shi (2013, Section 10.3, hereafter AS). The null hypothesis has the form

$$H_0 : \mathbf{E}(Y - \theta | X = x) \leq 0 \text{ for each } x \in \mathcal{X}$$

with a fixed θ. AS generated a random sample of (Y, X) from the following model:

$$Y = f(X) + U,$$

where $X \sim \text{Unif}[-2, 2]$, U follows truncated normal such that $U_i = \min\{\max\{-3, \sigma\tilde{U}_i\}, 3\}$ with $\tilde{U}_i \sim N(0, 1)$ and $\sigma = 1$, and $f(\cdot)$ is a function with an alternative shape. AS considered two functions:

$$\begin{aligned} f_{AS1}(x) &:= L\phi(x^{10}), \\ f_{AS2}(x) &:= L \cdot \max\{\phi((x - 1.5)^{10}), \phi((x + 1.5)^{10})\}, \end{aligned}$$

These two functions have steep slopes, f_{AS1} being a roughly plateau-shaped function and f_{AS2} a roughly double-plateau-shaped function, respectively. AS considered the following Monte Carlo designs:

$$\begin{aligned} &\text{DGP1: } f(x) = f_{AS1}(x) \text{ and } L = 1; \quad \text{DGP2: } f(x) = f_{AS1}(x) \text{ and } L = 5; \\ &\text{DGP3: } f(x) = f_{AS2}(x) \text{ and } L = 1; \quad \text{DGP4: } f(x) = f_{AS2}(x) \text{ and } L = 5. \end{aligned}$$

AS compared their tests with Chernozhukov, Lee, and Rosen (2013, hereafter CLR) and Lee, Song, and Whang (2013). The latter test uses conservative standard normal critical values based on the least favorable configuration.

TABLE 3. Results for Monte Carlo Experiments: Coverage Probability

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		AS		CLR		LSW1		LSW2	
	n	CvM	KS	series	local linear		$C_{\mathrm{cs}} = 0.4$	$C_{\mathrm{cs}} = 0.5$	$C_{\mathrm{cs}} = 0.6$
DGP1	100	.986	.986	.707	.804	1.00	.980	.990	.999
	250	.975	.973	.805	.893	1.00	.951	.960	.971
	500	.975	.970	.872	.925	1.00	.968	.976	.977
	1000	.971	.966	.909	.935	1.00	.962	.971	.973
DGP2	100	1.00	1.00	.394	.713	1.00	.996	.999	1.00
	250	1.00	1.00	.683	.856	1.00	.953	.963	.975
	500	1.00	1.00	.833	.908	1.00	.963	.972	.976
	1000	1.00	1.00	.900	.927	1.00	.965	.968	.968
DGP3	100	.970	.969	.620	.721	1.00	.987	.991	.993
	250	.969	.964	.762	.854	1.00	.952	.965	.973
	500	.963	.957	.854	.900	1.00	.966	.971	.976
	1000	.969	.963	.901	.927	1.00	.949	.957	.962
DGP3	100	.998	.999	.321	.655	1.00	.998	.999	1.00
	250	.997	.998	.612	.826	1.00	.952	.965	.976
	500	.994	.994	.808	.890	1.00	.964	.971	.973
	1000	.994	.991	.893	.918	1.00	.943	.950	.958

Notes: Figures in columns (1)-(5) are from Table V of Andrews and Shi (2013), whereas those in columns (6)-(8) are based 1000 Monte Carlo replications in each experiment, with the number of bootstrap replications being 200. LSW1 refers to the test of Lee, Song, and Whang (2013), which uses conservative standard normal critical values based on the least favorable configuration. LSW2 refers to this paper that uses bootstrap critical values based on the estimated contact set. The tuning parameter is chosen by the rule $\hat{c}_n = C_{\mathrm{cs}} \log\log(n) q_{1-0.1/\log(n)}(S_n^*)$, where $C_{\mathrm{cs}} \in \{0.4, 0.5, 0.6\}$.

In this paper, we used the same statistic for Lee, Song, and Whang (2013) as reported in AS. Specifically, we used the L_1 version of the test with the inverse standard error weight function. In implementing the test, we used $K(u) = (3/2)(1-(2u)^2)I(|u| \leq 1/2)$ and $h = 2 \times \hat{s}_X \times n^{-1/5}$, where $I(A)$ is the usual indicator function that has value one if A is true and zero otherwise and $\hat{s}_X$ is the sample standard deviation of X. Thus, the only difference between the new test (which we call LSW2) and Lee, Song, and Whang (2013) (which we call LSW1) is the use of critical values: LSW1 uses the standard normal critical values based on the least favorable configuration, whereas LSW2 uses bootstrap critical values based on the estimated contact set. For contact set estimation, we set the rule $\hat{c}_n = C_{\mathrm{cs}} \log\log(n) q_{1-0.1/\log(n)}(S_n^*)$, where $C_{\mathrm{cs}} \in \{0.4, 0.5, 0.6\}$.

Table 4. Results for Monte Carlo Experiments: False Coverage Probability

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		AS		CLR		LSW1		LSW2	
	n	CvM	KS	series	local linear		$C_{\text{cs}} = 0.4$	$C_{\text{cs}} = 0.5$	$C_{\text{cs}} = 0.6$
DGP1	100	.84	.89	.88	.83	.98	.81	.90	.95
	250	.57	.67	.82	.69	.92	.44	.49	.54
	500	.25	.37	.72	.50	.70	.17	.18	.20
	1000	.03	.07	.57	.26	.25	.02	.02	.02
DGP2	100	1.0	1.0	.91	.89	.99	.94	.98	1.0
	250	1.0	1.0	.85	.73	.96	.48	.54	.62
	500	.97	.99	.77	.56	.82	.19	.21	.23
	1000	.70	.89	.61	.33	.40	.03	.03	.03
DGP3	100	.70	.79	.89	.84	.90	.69	.79	.86
	250	.30	.46	.83	.66	.65	.27	.32	.35
	500	.06	.15	.70	.47	.26	.06	.06	.08
	1000	.00	.01	.55	.23	.02	.00	.00	.00
DGP4	100	.95	.99	.91	.88	.95	.89	.95	.97
	250	.66	.83	.86	.70	.75	.30	.35	.42
	500	.23	.42	.74	.51	.36	.07	.08	.09
	1000	.01	.04	.59	.29	.04	.00	.00	.00

Notes: See notes in Table 3. Figures in columns (1)-(5) are "CP-corrected", where those in columns (6)-(8) are not "CP-corrected".

The experiments considered sample sizes of $n = 100, 250, 500, 1000$ and the nominal level of $\alpha = 0.05$. We performed 1000 Monte Carlo replications in each experiment. The number of bootstrap replications was 200.

The null hypothesis is tested on $\mathcal{X} = [-1.8, 1.8]$. To compare simulation results from AS, the coverage probability (CP) is computed at nominal level 95% when $\theta = \max_{x \in \mathcal{X}} f(x)$ and the false coverage probability (FCP) is computed at nominal level 95% when $\theta = \max_{x \in \mathcal{X}} f(x) - 0.02$.

Tables 3 and 4 report the results of Monte Carlo experiments. In each table, figures in columns (1)-(5) are from Table V of Andrews and Shi (2013), whereas those in columns (6)-(8) are from our Monte Carlo experiments. Table 3 shows that coverage probabilities of LSW2 are much closer to the nominal level than those of LSW1. When $c = 0.4$ and $n = 100$ or 250, we see some under-coverage for LSW2, but it disappears as n gets larger. Table 4 reports the false coverage probabilities (FCPs). Figures in columns (1)-(5) are "CP-corrected" by AS, where those in columns (6)-(8) are not "CP-corrected". However, CP-correction would not change the results for either $n \geq 500$ or $c \geq 0.5$ since in each of these cases, we have over-coverage. We can see that in terms of FCPs, LSW2 performs much better than LSW1 in all DGPs. Furthermore, the performance of LSW2 is equivalent to that of AS for DGP1, DGP3, and DGP4, and is superior to AS for DGP2. Overall, our simulation results show that our new test is a substantial improved version of LSW1 and is now very much comparable to AS. The relative poor performance of CLR in tables 3 and 4 are mainly due to the experimental design. If the underlying function is sharply peaked, as those in the reported simulations of Chernozhukov, Lee, and Rosen (2013), CLR performs better than AS. In unreported simulations, we confirmed that CLR performs better than LSW2 as well. This is very reasonable since CLR is based on the sup-norm statistic, whereas ours is based on the one-sided L_p norm. Therefore, we may conclude that AS, CLR, and LSW2 complement each other.

II. Further Examples of Testing Functional Inequalities

II.1. **Testing Functional Inequalities in the Auction Model via Estimating Conditional Cumulative Distribution Functions.** This appendix illustrates the usefulness and flexibility of our framework by reconsidering implications from GPV in terms of conditional stochastic dominance. Specifically, relative to the test statistic in the main text (based on estimating conditional quantiles functions), we consider a related but distinct testing statistic based on estimating conditional cumulative distribution functions.

We may rewrite (3.3) as

$$\begin{aligned} G_3(b|x) - G_2(b|x) &\leq 0 \text{ for any } b \in [\underline{b}, \overline{b}_2] \text{ and for any } x \in \mathcal{X} \\ G_2\left[(b+\underline{b})/2|x\right] - G_3(b|x) &\leq 0 \text{ for any } b \in [\underline{b}, \overline{b}_3] \text{ and for any } x \in \mathcal{X}. \end{aligned} \tag{II.1}$$

where $G_k(\cdot|x)$ is the CDF of the observed bid (conditional on $X = x$) when the number of bidders is $I = k$ $(k = 2, 3)$. Recall that in GPV, the support of the observed bid is $[\underline{b}, \overline{b}_k]$. Note that strictly speaking, the restrictions in (II.1) are not identical to those in (3.3) since τ in (3.3) is limited to a compact strict subset of $(0, 1)$.

To implement the test, it is necessary to know $\overline{b}_k$ $(k = 2, 3)$, in addition to the value of $\underline{b}$. As before, in our application, we set $\underline{b}$ the overall minimum value, and $\overline{b}_k$ the maximum value when the number of bids is k for $k = 2, 3$.

Define

$$\begin{aligned} v_1(b, x) &:= G_3(b|x) - G_2(b|x), \\ v_2(b, x) &:= G_2\left[(b+\underline{b})/2|x\right] - G_3(b|x). \end{aligned}$$

To construct the test statistic, it is necessary to estimate $v_j(b, x)$, where $j = 1, 2$.

For each $k = 2, 3$, define $p_k(x) := \Pr(L_i = k | X_i = x) f(x)$, where $f(\cdot)$ is the marginal density of X_i. To describe our estimator of $v_j(b, x)$ in a simple form, define

$$\hat{B}_i(b) := L_i^{-1} \sum_{\ell=1}^{L_i} 1(B_{\ell i} \leq b),$$
$$\hat{p}_k(x) := n^{-1} \sum_{i=1}^{n} 1(L_i = k) K_h(x - X_i),$$

where $K_h(\cdot) = K(\cdot/h)/h^d$, K is a d-dimensional kernel function, h is a bandwidth, and d is the dimension of X. Then $v_1(b, x)$ and $v_2(b, x)$ are estimated by

$$\hat{v}_1(b, x) := n^{-1} \sum_{i=1}^{n} \hat{B}_i(b) \left[\frac{1(L_i = 3)}{\hat{p}_3(x)} - \frac{1(L_i = 2)}{\hat{p}_2(x)} \right] K_h(x - X_i),$$
$$\hat{v}_2(b, x) := n^{-1} \sum_{i=1}^{n} \left[\frac{\hat{B}_i \left[(b + \underline{b})/2 \right] 1(L_i = 2)}{\hat{p}_2(x)} - \frac{\hat{B}_i(b) 1(L_i = 3)}{\hat{p}_3(x)} \right] K_h(x - X_i).$$

Note that again all bids are combined in each auction (see the definition of $\hat{B}_i(b)$) since we consider symmetric bidders.

The sum statistic would be convenient for testing (II.1) since $\overline{b}_2$ can be different from $\overline{b}_3$. Then the test statistic has the form

$$\hat{\theta} = \int_{[\underline{b}, \overline{b}_2] \times \mathcal{X}} [r_n \hat{v}_1(b, x)]_+^p \, dQ(b, x) + \int_{[\underline{b}, \overline{b}_3] \times \mathcal{X}} [r_n \hat{v}_2(b, x)]_+^p \, dQ(b, x), \tag{II.2}$$

where Q is Lebesgue measure. Note that we did not normalize $\hat{v}_j(b, x)$ by its pointwise standard error here. One advantage of doing this is that we can test the null hypothesis on the full support $[\underline{b}, \overline{b}_k]$, $(k = 2, 3)$ without an elaborate use of the trimming function or a decaying weight function at the boundary.

II.2. **Nonparametric Tests of Monotonicity: an L_p Approach.** In this appendix, we present new methods for testing monotonicity by constructing one-sided L_p-type functionals in a suitable fashion. Suppose that we observe n independent and identically distributed random vectors $\{(Y_i, X_i) : i = 1, \ldots, n\}$ from the joint distribution of random variables Y and X, where Y is the dependent variable and X is a univariate explanatory variable. We consider testing monotonicity in three examples: one in mean regression, another in conditional distribution function, and the third in quantile regression. In what follows, we focus on the case that $J = 1$; however, it is straightforward to extend to the $J > 1$ case with a multivariate vector of Y_i.

II.2.1. *Testing monotonicity of mean regression.* Let $\mathcal{X} \subset \mathbf{R}$ be the region of our interest in the domain of the regression function $\mathbf{E}[Y|X = x]$. Consider testing the hypothesis

$$H_0 : \mathbf{E}[Y|X = x] \text{ is increasing on } \mathcal{X}.$$

Let K be a one-dimensional kernel function and h be a bandwidth. Define the following U-process: for $x \in \mathcal{X}$,

$$\hat{v}(x) \equiv \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} (Y_j - Y_i)\text{sgn}(X_i - X_j) K_h(x - X_i) K_h(x - X_j),$$

where $K_h(z) \equiv K(z/h)/h$, and $\text{sgn}(x) \equiv 1\{x > 0\} - 1\{x < 0\}$, $x \in \mathbf{R}$. If $\mathbf{E}[Y|X = \cdot]$ is continuously differentiable, as $n \to \infty$, we have that

$$\mathbf{E}\hat{v}(x) \to -\frac{\partial \mathbf{E}[Y|X = x]}{\partial x} f^2(x) \int\int |u_1 - u_2| K(u_1) K(u_2) du_1 du_2,$$

where $f(\cdot)$ is the density function of X. That is, the limit of $\mathbf{E}\hat{v}(x)$ is less than or equal to zero if and only if $\partial \mathbf{E}[Y|X = x]/\partial x \geq 0$. This suggests we develop a test based on

$$\hat{\theta} = \int_{\mathcal{X}} (\max\{r_n \hat{v}(x), 0\})^p \, dx \tag{II.3}$$

with a suitable choice of r_n.

Define $v_n(x) \equiv h^{-1}\mathbf{E}\left[(Y_j - Y_i)\text{sgn}(X_i - X_j) K_h(x - X_i) K_h(x - X_j)\right]$. It can be shown that the U-process $\hat{v}(x)$ has the following asymptotic representation:

$$\begin{aligned} &\sqrt{nh^3}\,\{\hat{v}(x) - v_n(x)\} \\ &= \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \left\{ \beta_{n,x}\left(Y_i, \frac{x - X_i}{h}\right) - \mathbf{E}\left[\beta_{n,x}\left(Y_j, \frac{x - X_j}{h}\right)\right] \right\} + R_n, \end{aligned}$$

where R_n is a remainder term that is of smaller order than the leading term and

$$\beta_{n,x}(Y_i, z) \equiv 2\{\mathbf{E}[Y|X = x] - Y_i\} K(z) \int \text{sgn}(u - z) K(u)\, du.$$

Therefore, we have $r_n = \sqrt{nh^3}$.

In a contemporaneous paper, Chetverikov (2012) proposed an adaptive test using the sup-norm statistic of a studentized version of a U-process, including $\hat{v}(x)$ as a special case. The test based on (II.3) is an alternative to the sup-norm test of Chetverikov (2012). The test of Chetverikov (2012) is closely related to the tests proposed in Ghosal, Sen, and van der Vaart (2000). They developed monotonicity tests for the function $m(\cdot)$ in the transformation model $\phi(Y) = m(X) + \varepsilon$, where $\phi(\cdot)$ is a monotone function and X and ε are independent. In their setup, independence between X and ε is indispensable, but $\phi(\cdot)$ can be unknown as long as it is strictly monotone. They constructed sup-norm and time spent test statistics ($S_{1,n}$ and

$S_{2,n}$ of Ghosal, Sen, and van der Vaart (2000, page 1060)) using the following U-process:

$$U_n(x) \equiv \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} \mathrm{sgn}(Y_j - Y_i)\mathrm{sgn}(X_i - X_j) K_h(x - X_i) K_h(x - X_j). \tag{II.4}$$

Equation (2.5) of Ghosal, Sen, and van der Vaart (2000) shows that the limit of $h^{-1}\mathbf{E}U_n(x)$ is less than or equal to zero if and only if $\partial m(x)/\partial x \geq 0$. As before, we may develop a test based on (II.4) with a redefined $\hat{v}(x) = h^{-1}U_n(x)$ and the same r_n.

In addition to sup-norm and time spent test statistics, Ghosal, Sen, and van der Vaart (2000, page 1070) suggested test statistics that are similar to our one-sided L_p statistics; however, they did not provide asymptotic theory, remarking that there are no limit theorems for one-sided L_p functionals of a stationary Gaussian process. However, we can obtain the limiting distribution of our suggested test statistic in (II.3) by a direct approximation of $\hat{\theta}$ via Poissonization techniques, without going through strong approximation results such as Rio (1994) and Chernozhukov, Lee, and Rosen (2013).

II.2.2. *Testing stochastic monotonicity.* Let $F_{Y|X}(\cdot|x)$ denote the distribution of Y conditional on $X = x$, where (Y, X) is a pair of random variables whose joint distribution is absolutely continuous with respect to Lebesgue measure. We assume that the function $F_{Y|X}(y|x)$ is continuously differentiable with respect to x for each y. Consider testing the hypothesis $H_0 : \partial F_{Y|X}(y|x)/\partial x \leq 0$ for all $(y, x) \in \mathcal{Y} \times \mathcal{X}$, where $\mathcal{Y} \subset \mathbf{R}$ and $\mathcal{X} \subset \mathbf{R}$ are domains of interest.

In this subsection, consider the following U-process: for $(y, x) \in \mathcal{Y} \times \mathcal{X}$,

$$\hat{v}(y, x) \equiv \frac{1}{n(n-1)h} \sum_{i=1}^{n} \sum_{j=1, j\neq i}^{n} [1(Y_i \leq y) - 1(Y_j \leq y)]\mathrm{sgn}(X_i - X_j) K_h(x - X_i) K_h(x - X_j). \tag{II.5}$$

Lee, Linton, and Whang (2009) proposed a nonparametric test of stochastic monotonicity using the sup-norm statistic based on $\hat{v}(y, x)$. Note that as mentioned in Lee, Linton, and Whang (2009), under the regularity conditions imposed in this paper, as $n \to \infty$, we have that

$$\mathbf{E}\hat{v}(y, x) \to \frac{\partial F_{Y|X}(y|x)}{\partial x} f^2(x) \int\int |u_1 - u_2| K(u_1) K(u_2) du_1 du_2.$$

That is, the limit of $\mathbf{E}\hat{v}(y, x)$ is less than or equal to zero if and only if H_0 holds. Again this suggests we develop a test based on

$$\hat{\theta} = \int_{\mathcal{X}\times\mathcal{T}} \left(\max\left\{r_n \hat{v}(\tau, x), 0\right\}\right)^p dQ(x, \tau) \tag{II.6}$$

with $r_n = \sqrt{nh^3}$. The one-sided L_p functional-based test complements the sup-norm test of Lee, Linton, and Whang (2009). Delgado and Escanciano (2012) proposed an alternative approach based on the sup-norm of the difference between the empirical copula function and its least concave majorant.

II.2.3. *Testing Monotonicity of Quantile Regression.* Let $q(\tau|x)$ denote the τ-th quantile of Y conditional on $X = x$, where $\tau \in (0,1)$. In this subsection, we consider testing monotonicity of quantile regression. The null hypothesis and the alternative hypothesis are as follows:

$$\text{(II.7)} \qquad \begin{aligned} &H_0 : q(\tau|x_1) \leq q(\tau|x_2) \text{ for all } (\tau,(x_1,x_2)) \in \mathcal{T} \times \mathcal{X} \text{ against} \\ &H_1 : q(\tau|x_1) > q(\tau|x_2) \text{ for some } (\tau,(x_1,x_2)) \in \mathcal{T} \times \mathcal{X}, \end{aligned}$$

where $\mathcal{X} \subset \{(x_1,x_2) \in \mathbf{R}^2 : x_1 \leq x_2\}$ and $\mathcal{T} \subset (0,1)$. The null hypothesis states that the quantile functions are increasing on $\mathcal{X}$ for all $\tau \in \mathcal{T}$, and the alternative hypothesis is the negation of the hypothesis. If $\mathcal{T}$ consists of a singleton set, then testing (II.7) amounts to testing monotonicity of quantile regression at a fixed quantile.

Suppose that $q(\tau|x)$ is continuously differentiable on on $\mathcal{X}$ for each $\tau \in \mathcal{T}$. Then one natural approach is to test the sign restriction of the derivative of $q(\tau|x)$. In other words, we again develop a test based on (II.6) with $\hat{v}(\tau,x)$ now being the local polynomial estimator of $\partial q(\tau|x)/\partial x$ and $r_n = \sqrt{nh^3}$.

Our general framework covers various other forms of monotonicity tests for quantile regression. For example, one might be interested in monotonicity of an interquartile regression function. More specifically, let $\tau_1 < \tau_2$ be chosen from $(0,1)$ and write $\Delta q_{\tau_1,\tau_2}(x) \equiv q(\tau_2|x_1) - q(\tau_1|x_1)$. Then the null hypothesis and the alternative hypothesis of monotonicity of the interquartile regression function are as follows:

$$\text{(II.8)} \qquad \begin{aligned} H_{0,\Delta} &: \Delta q_{\tau_1,\tau_2,j}(x_1) \leq \Delta q_{\tau_1,\tau_2,j}(x_2) \text{ for all } (x_1,x_2) \in \mathcal{X} \text{ against} \\ H_{1,\Delta} &: \Delta q_{\tau_1,\tau_2,j}(x_1) > \Delta q_{\tau_1,\tau_2,j}(x_2) \text{ for some } (x_1,x_2) \in \mathcal{X}. \end{aligned}$$

The null hypothesis states that the interquartile regression function $q_{\tau_2,j}(x) - q_{\tau_1,j}(x)$ is increasing on $\mathcal{X}$ for all $j \in \mathbb{N}_J$. This type of monotonicity can be used to investigate whether the income inequality (in terms of interquartile comparison) become severe as certain demographic variable X increases. Once again, we can consider a test based on (II.3) with $\hat{v}(x)$ now being the local polynomial estimator of $[\partial q(\tau_2|x)/\partial x - \partial q(\tau_2|x)/\partial x]$ and $r_n = \sqrt{nh^3}$.

Appendix A. Proofs of Theorems 1-5

The roadmap of Appendix A is as follows. Appendix A begins with the proofs of Lemma 1 (the representation of $\hat{\theta}$) and Lemma 2 (the uniform convergence of $\hat{v}_{\tau,j}(x)$). Then we establish auxiliary results, Lemmas A1-A4, to prepare for the proofs of Theorems 1-3. The brief descriptions of these auxiliary results are given below.

Lemma A1 establishes asymptotic representation of the location normalizers for the test statistic both in the population and in the bootstrap distribution. The crucial implication is that the difference between the population version and the bootstrap version is of order $o_P(h^{d/2})$, $\mathcal{P}$-uniformly. The result is in fact an immediate consequence of Lemma D12 in Appendix D.

Lemma A2 establishes uniform asymptotic normality of the representation of $\hat{\theta}$ and its bootstrap version. The asymptotic normality results use the method of Poissonization as in Giné, Mason, and Zaitsev (2003) and Lee, Song, and Whang (2013). However, in contrast to the preceding researches, the results established here are much more general, and hold uniformly over a wide class of probabilities. The lemma relies on Lemmas C7-C9 in Appendix C and their bootstrap versions in Lemmas D7-D9 in Appendix D. These results are employed to obtain the uniform asymptotic normality of the representation of $\hat{\theta}$ in Lemma A2.

Lemma A3 establishes that the estimated contact sets $\hat{B}_A(\hat{c}_n)$ are covered by its enlarged population version, and covers its shrunk population version with probability approaching one uniformly over $P \in \mathcal{P}$. In fact, this is an immediate consequence of the uniform convergence results for $\hat{v}_{\tau,j}(x)$ and $\hat{\sigma}_{\tau,j}(x)$ in Assumptions 3 and 5. Lemma A3 is used later, when we replace the estimated contact sets by their appropriate population versions, eliminating the nuisance to deal with the estimation errors in $\hat{B}_A(\hat{c}_n)$.

Lemma A4 presents the approximation result of the critical values for the original and bootstrap test statistics in Lemma A2, by critical values from the standard normal distribution uniformly over $P \in \mathcal{P}$. Although we do not propose using the normal critical values, the result is used as an intermediate step for justifying the use of the bootstrap method in this paper. Obviously, Lemma A4 follows as a consequence of Lemma A2.

Equipped with Lemmas A1-A4, we proceed to prove Theorem 1. For this, we first use the representation result of Lemma 1 for $\hat{\theta}$. In doing so, we use $B_A(c_{n,L}, c_{n,U})$ as a population version of $\hat{B}_A(\hat{c}_n)$. This is because

$$B_A(c_{n,L}, c_{n,U}) \subset \hat{B}_A(\hat{c}_n)$$

with probability approaching one by Lemma A3, and thus, makes the bootstrap test statistic $\hat{\theta}^*$ dominate the one that involves $B_A(c_{n,L}, c_{n,U})$ in place of $\hat{B}_A(\hat{c}_n)$. The distribution of the latter bootstrap version with $B_A(c_{n,L}, c_{n,U})$ is asymptotically equivalent to the representation of $\hat{\theta}$ with $B_A(c_{n,L}, c_{n,U})$ after location-scale normalization, as long as the limiting distribution is nondegenerate. When the limiting distribution is degenerate, we use the second component $h^{d/2}\eta + \hat{a}^*$ in the definition of $c^*_{\alpha,\eta}$ to ensure the asymptotic validity of the bootstrap procedure.

For both cases of degenerate and nondegenerate limiting distributions, Lemma A1 which enables one to replace $\hat{a}^*$ by an appropriate population version is crucial.

The proof of Theorem 2 that shows the asymptotic exactness of the bootstrap test modifies the proof of Theorem 1 substantially. Instead of using the representation result of Lemma 1 for $\hat{\theta}$ with $B_{n,A}(c_{n,L}, c_{n,U})$, we now use the same version but with $B_{n,A}(c_{n,U}, c_{n,L})$. This is because for asymptotic exactness, we need to approximate the original and bootstrap quantities by versions using $B_{n,A}(q_n)$ for small q_n, and to do this, we need to control the remainder term in the bootstrap statistic with the integral domain $\hat{B}_A(\hat{c}_n)\backslash B_{n,A}(q_n)$. By our choice of $B_{n,A}(c_{n,U}, c_{n,L})$ and by the fact that we have

$$\hat{B}_A(\hat{c}_n) \subset B_{n,A}(c_{n,U}, c_{n,L}),$$

with probability approaching one by Lemma A3, we can bound the remainder term with a version with the integral domain $B_{n,A}(c_{n,U}, c_{n,L})\backslash B_{n,A}(q_n)$. Thus this remainder term vanishes by the condition for λ_n and q_n in the definition of $\mathcal{P}_n(\lambda_n, q_n)$.

The rest of the proofs are devoted to proving the power properties of the bootstrap procedure. Theorem 3 establishes consistency of the bootstrap test. Theorems 4 and 5 establish local power functions under Pitman local drifts. The proofs of Theorems 4-5 are similar to the proof of Theorem 2, as we need to establish the asymptotically exact form of the rejection probability for the bootstrap test statistic. Nevertheless, we need to employ some delicate arguments to deal with the Pitman local alternatives, and need to expand the rejection probability to obtain the final results. For this, we first establish Lemmas A5-A7. Essentially, Lemma A5 is a version of the representation result of Lemma 1 under local alternatives. Lemma A6 and Lemma A7 parallel Lemma A1 and Lemma 2 under local alternatives.

Let us begin by proving Lemma 1. First, recall the following definitions

$$\text{(A.1)} \quad \hat{\mathbf{s}}_\tau(x) \equiv \left[\frac{r_{n,j}\{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)\}}{\hat{\sigma}_{\tau,j}(x)}\right]_{j\in\mathbb{N}_J} \text{ and } \hat{\mathbf{s}}_\tau^*(x) \equiv \left[\frac{r_{n,j}\{\hat{v}_{\tau,j}^*(x) - \hat{v}_{,\tau,j}(x)\}}{\hat{\sigma}_{\tau,j}^*(x)}\right]_{j\in\mathbb{N}_J}.$$

Also, define

$$\text{(A.2)} \qquad \hat{\mathbf{u}}_\tau(x) \equiv \left[\frac{r_{n,j}\hat{v}_{\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)}\right]_{j\in\mathbb{N}_J} \text{ and } \mathbf{u}_\tau(x;\hat{\sigma}) \equiv \left[\frac{r_{n,j}v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)}\right]_{j\in\mathbb{N}_J}.$$

Proof of Lemma 1. It suffices to show the following two statements:

Step 1: As $n \to \infty$,

$$\inf_{P\in\mathcal{P}_0} P\left\{\int_{\mathcal{S}\backslash B_n(c_{n,1},c_{n,2})} \Lambda_p\left(\hat{\mathbf{u}}_\tau(x)\right) dQ(x,\tau) = 0\right\} \to 1,$$

where we recall $B_n(c_{n,1}, c_{n,2}) \equiv \cup_{A\in\mathcal{N}_J} B_{n,A}(c_{n,1}, c_{n,2})$.

Step 2: For each $A \in \mathcal{N}_J$, as $n \to \infty$,

$$\inf_{P \in \mathcal{P}_0} P \left\{ \int_{B_{n,A}(c_{n,1}, c_{n,2})} \{ \Lambda_p (\hat{\mathbf{u}}_\tau(x)) - \Lambda_{A,p} (\hat{\mathbf{u}}_\tau(x)) \} \, dQ(x,\tau) = 0 \right\} \to 1.$$

First, we prove Step 1. We write the integral in the probability as

$$\int_{\mathcal{S} \backslash B_n(c_{n,1}, c_{n,2})} \Lambda_p \left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_\tau(x; \hat{\sigma}) \right) dQ(x,\tau). \tag{A.3}$$

Let

$$A_n(x,\tau) \equiv \left\{ j \in \mathbb{N}_J : \frac{r_{n,j} v_{n,\tau,j}(x)}{\sigma_{n,\tau,j}(x)} \geq -(c_{n,1} \wedge c_{n,2}) \right\}.$$

We first show that when $(x,\tau) \in \mathcal{S} \backslash B_n(c_{n,1}, c_{n,2})$, we have $A_n(x,\tau) = \varnothing$ under the null hypothesis. Suppose that $(x,\tau) \in \mathcal{S} \backslash B_n(c_{n,1}, c_{n,2})$ but to the contrary, $A_n(x,\tau)$ is nonempty. By the definition of $A_n(x,\tau)$, we have $(x,\tau) \in B_{n,A_n(x,\tau)}(c_{n,1}, c_{n,2})$. However, since

$$\mathcal{S} \backslash B_n(c_{n,1}, c_{n,2}) = \mathcal{S} \cap \left(\cap_{A \in \mathcal{N}_J} B_{n,A}^c(c_{n,1}, c_{n,2}) \right) \subset B_{n,A_n(x,\tau)}^c(c_{n,1}, c_{n,2}),$$

this contradicts the fact that $(x,\tau) \in \mathcal{S} \backslash B_n(c_{n,1}, c_{n,2})$. Hence whenever $(x,\tau) \in \mathcal{S} \backslash B_n(c_{n,1}, c_{n,2})$, we have $A_n(x,\tau) = \varnothing$.

Note that

$$\frac{v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} = \frac{v_{n,\tau,j}(x)}{\sigma_{n,\tau,j}(x)} \left\{ 1 + \frac{\sigma_{n,\tau,j}(x) - \hat{\sigma}_{\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} \right\} = \frac{v_{n,\tau,j}(x)}{\sigma_{n,\tau,j}(x)} \left\{ 1 + o_P(1) \right\},$$

where $o_P(1)$ is uniform over $(x,\tau) \in \mathcal{S}$ and over $P \in \mathcal{P}$ by Assumption A5. Fix a small $\varepsilon > 0$. We have for all $j \in \mathbb{N}_J$,

$$\begin{aligned} & \inf_{P \in \mathcal{P}_0} P \left\{ \frac{r_{n,j} v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} < -\frac{c_{n,1} \wedge c_{n,2}}{1+\varepsilon} \text{ for all } (x,\tau) \in \mathcal{S} \backslash B_n(c_{n,1}, c_{n,2}) \right\} \\ \geq \; & \inf_{P \in \mathcal{P}_0} P \left\{ \frac{r_{n,j} v_{n,\tau,j}(x)}{\sigma_{n,\tau,j}(x)} < -\frac{c_{n,1} \wedge c_{n,2}}{(1+\varepsilon)\{1 + o_P(1)\}} \text{ for all } (x,\tau) \in \mathcal{S} \backslash B_n(c_{n,1}, c_{n,2}) \right\} \to 1, \end{aligned}$$

as $n \to \infty$, where the last convergence follows because $A_n(x,\tau) = \varnothing$ for all $(x,\tau) \in \mathcal{S} \backslash B_n(c_{n,1}, c_{n,2})$. Therefore, with probability approaching one, the term in (A.3) is bounded by

$$\int_{\mathcal{S} \backslash B_n(c_{n,1}, c_{n,2})} \Lambda_p \left(\hat{\mathbf{s}}_\tau(x) - \left(\frac{c_{n,1} \wedge c_{n,2}}{1+\varepsilon} \right) \mathbf{1}_J \right) dQ(x,\tau), \tag{A.4}$$

where $\mathbf{1}_J$ is a J-dimensional vector of ones. Using the definition of $\Lambda_p(\mathbf{v})$, bound the above integral by

$$J^{p/2} \left(\sum_{j=1}^{J} \left[r_{n,j} \sup_{(x,\tau) \in \mathcal{S}} \left| \frac{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} \right| - \frac{c_{n,1} \wedge c_{n,2}}{1+\varepsilon} \right]_+^2 \right)^{p/2}. \tag{A.5}$$

Note that by Assumption A3,

$$r_{n,j} \sup_{(x,\tau)\in\mathcal{S}} \left| \frac{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} \right| = O_P\left(\sqrt{\log n}\right).$$

Fix any arbitrarily large $M > 0$ and denote by E_n the event that

$$r_{n,j} \sup_{(x,\tau)\in\mathcal{S}} \left| \frac{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} \right| \leq M\sqrt{\log n}.$$

The term (A.5), when restricted to this event E_n, is bounded by

$$J^{p/2}\left(\sum_{j=1}^{J}\left[M\sqrt{\log n} - \frac{c_{n,1}\wedge c_{n,2}}{1+\varepsilon}\right]_+^2\right)^{p/2}$$

which becomes zero from some large n on, given that $(c_{n,1}\wedge c_{n,2})/\sqrt{\log n} \to \infty$. Since $\sup_{P\in\mathcal{P}_0} PE_n^c \to 0$ as $n\to\infty$ and then $M\to\infty$ by Assumption A3, we obtain the desired result of Step 1.

As for Step 2, we have for any small $\varepsilon > 0$, and for all $j \in \mathbb{N}_J\backslash A$,

$$\text{(A.6)} \qquad \begin{aligned} & P\left\{\frac{r_{n,j}v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} < -\frac{c_{n,1}\wedge c_{n,2}}{1+\varepsilon} \text{ for all } (x,\tau)\in B_{n,A}(c_{n,1},c_{n,2})\right\} \\ \geq \;& P\left\{\frac{r_{n,j}v_{n,\tau,j}(x)}{\sigma_{n,\tau,j}(x)} < -\frac{c_{n,1}\wedge c_{n,2}}{(1+\varepsilon)\left\{1+o_P(1)\right\}} \text{ for all } (x,\tau)\in B_{n,A}(c_{n,1},c_{n,2})\right\} \to 1, \end{aligned}$$

similarly as before. Let $\bar{\mathbf{s}}_{\tau,A}(x)$ be a J-dimensional vector whose j-th entry is $r_{n,j}\hat{v}_{n,\tau,j}(x)/\hat{\sigma}_{\tau,j}(x)$ if $j\in A$, and $r_{n,j}\{\hat{v}_{n,\tau,j}(x) - v_{n,\tau,j}(x)\}/\hat{\sigma}_{\tau,j}(x)$ if $j\in\mathbb{N}_J\backslash A$. Since by Assumption A5, we have

$$\inf_{P\in\mathcal{P}_0} P\left\{\mathbf{u}_\tau(x;\hat{\sigma}) \leq 0 \text{ for all } (x,\tau)\in\mathcal{S}\right\} \to 1,$$

as $n\to\infty$, using either definition of $\Lambda_p(\mathbf{v})$ in (4.1),

$$\text{(A.7)} \qquad \begin{aligned} & \int_{B_{n,A}(c_{n,1},c_{n,2})} \Lambda_{A,p}\left(\hat{\mathbf{u}}_\tau(x)\right) dQ(x,\tau) \\ \leq \;& \int_{B_{n,A}(c_{n,1},c_{n,2})} \Lambda_p\left(\hat{\mathbf{u}}_\tau(x)\right) dQ(x,\tau) \\ \leq \;& \int_{B_{n,A}(c_{n,1},c_{n,2})} \Lambda_p\left(\bar{\mathbf{s}}_{\tau,A}(x) - \frac{c_{n,1}\wedge c_{n,2}}{1+\varepsilon}\mathbf{1}_{-A}\right) dQ(x,\tau), \end{aligned}$$

where $\mathbf{1}_{-A}$ is the J-dimensional vector whose j-th entry is zero if $j\in A$ and one if $j\in\mathbb{N}_J\backslash A$, and the last inequality holds with probability approaching one by (A.6). Note that by Assumption A3 and by the assumption that $\sqrt{\log n}\{c_{n,1}^{-1} + c_{n,2}^{-1}\} \to \infty$, we deduce that for any $j \in \mathbb{N}_J$,

$$\inf_{P \in \mathcal{P}_0} P\left\{ r_{n,j} \sup_{(x,\tau)\in\mathcal{S}} \left| \frac{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} \right| \leq \frac{c_{n,1} \wedge c_{n,2}}{1+\varepsilon} \right\} \to 1,$$

as $n \to \infty$. Hence, as $n \to \infty$,

$$\inf_{P \in \mathcal{P}_0} P \left\{ \begin{array}{c} \int_{B_{n,A}(c_{n,1},c_{n,2})} \Lambda_p \left(\bar{\mathbf{s}}_{\tau,A}(x) - ((c_{n,1} \wedge c_{n,2})/(1+\varepsilon)) \mathbf{1}_{-A} \right) dQ(x,\tau) \\ = \int_{B_{n,A}(c_{n,1},c_{n,2})} \Lambda_{A,p} \left(\bar{\mathbf{s}}_{\tau,A}(x) \right) dQ(x,\tau) \end{array} \right\} \to 1.$$

Since

$$\int_{B_{n,A}(c_{n,1},c_{n,2})} \Lambda_{A,p}\left(\bar{\mathbf{s}}_{\tau,A}(x)\right) dQ(x,\tau) = \int_{B_{n,A}(c_{n,1},c_{n,2})} \Lambda_{A,p}\left(\hat{\mathbf{u}}_{\tau}(x)\right) dQ(x,\tau),$$

we obtain the desired result from (A.7). ∎

Now let us turn to the proof of Lemma 2 in Section 4.4.

Proof of Lemma 2. (i) Recall the definition $b_{n,ij}(x,\tau) \equiv \beta_{n,x,\tau,j}\left(Y_{ij}, (X_i - x)/h\right)$. Take $M_{n,j} \equiv \sqrt{nh^d}/\sqrt{\log n}$, and let

$$b^a_{n,ij}(x,\tau) \equiv b_{n,ij}(x,\tau) 1_{n,ij} \text{ and } b^b_{n,ij}(x,\tau) \equiv b_{n,ij}(x,\tau)\left(1 - 1_{n,ij}\right),$$

where $1_{n,ij} \equiv 1\{\sup_{(x,\tau)\in\mathcal{S}} |b_{n,ij}(x,\tau)| \leq M_{n,j}/2\}$. First, note that by Assumption A1,

$$r_{n,j}\sqrt{h^d} \sup_{(x,\tau)\in\mathcal{S}} \left| \frac{\hat{v}_{\tau,j}(x) - v_{n,\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} \right| \tag{A.8}$$

$$\leq \sup_{(x,\tau)\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(b^a_{n,ij}(x,\tau) - \mathbf{E}\left[b^a_{n,ij}(x,\tau) \right] \right) \right|$$

$$+ \sup_{(x,\tau)\in\mathcal{S}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(b^b_{n,ij}(x,\tau) - \mathbf{E}\left[b^b_{n,ij}(x,\tau) \right] \right) \right| + o_P(1), \ \mathcal{P}\text{-uniformly.} \tag{A.9}$$

We now prove part (i) by proving the following two steps.

Step 1:

$$\sup_{(x,\tau)\in\mathcal{S}} \left| \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n \left(b^b_{n,ij}(x,\tau) - \mathbf{E}\left[b^b_{n,ij}(x,\tau) \right] \right) \right| = o_P(\sqrt{\log n}), \ \mathcal{P}\text{-uniformly.}$$

Step 2:

$$\sup_{(x,\tau)\in\mathcal{S}} \left| \frac{1}{\sqrt{nh^d}} \sum_{i=1}^n \left(b^a_{n,ij}(x,\tau) - \mathbf{E}\left[b^a_{n,ij}(x,\tau) \right] \right) \right| = O_P(\sqrt{\log n}), \ \mathcal{P}\text{-uniformly.}$$

Step 1 is carried out by some elementary moment calculations, whereas Step 2 is proved using a maximal inequality of Massart (2007, Theorem 6.8).

Proof of Step 1: It is not hard to see that

$$\begin{aligned}
&\mathbf{E}\left[\sup_{(x,\tau)\in\mathcal{S}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(b_{n,ij}^{b}(x,\tau)-\mathbf{E}\left[b_{n,ij}^{b}(x,\tau)\right]\right)\right|\right]\\
\leq\;& 2\sqrt{n}\mathbf{E}\left[\sup_{(x,\tau)\in\mathcal{S}}\left|b_{n,ij}(x,\tau)\right|(1-1_{n,ij})\right]\\
\leq\;& C\sqrt{n}\left(\frac{M_{n,j}}{2}\right)^{-3}\mathbf{E}\left[\sup_{(x,\tau)\in\mathcal{S}}\left|b_{n,ij}(x,\tau)\right|^{4}\right]\leq C_1\sqrt{n}\left(\frac{M_{n,j}}{2}\right)^{-3},
\end{aligned}$$

for some $C_1 > 0$, $C > 0$. The last bound follows by the uniform fourth moment bound for $b_{n,ij}(x,\tau)$ assumed in Lemma 2. Note that

$$\sqrt{n}\left(M_{n,j}\right)^{-3} = n^{-1}h^{-3d/2}\left(\log n\right)^{3/2} = o\left(\sqrt{\log n}h^{d/2}\right),$$

by the condition that $n^{-1/2}h^{-d-\nu} \to 0$ for some small $\nu > 0$.

Proof of Step 2: For each $j \in \mathbb{N}_J$, let $\mathcal{F}_{n,j} \equiv \{\beta^{a}_{n,x,\tau,j}(\cdot,(\cdot - x)/h)/M_{n,j} : (x,\tau)\in\mathcal{S}\}$, where $\beta^{a}_{n,x,\tau,j}(Y_{ij},(X_i - x)/h) \equiv b^{a}_{n,ij}(x,\tau)$. Note that the indicator function $1_{n,ij}$ in the definition of $\beta^{a}_{n,x,\tau,j}$ does not depend on (x,τ) of $\beta^{a}_{n,x,\tau,j}$. Using (4.11) in Lemma 2 and following (part of) the arguments in the proof of Theorem 3 of Chen, Linton, and Van Keilegom (2003), we find that there exist $C_1 > 0$ and $C_{2,j} > 0$ such that for all $\varepsilon > 0$,

$$N_{[]}\left(\varepsilon,\mathcal{F}_{n,j},L_2(P)\right) \leq N\left(\left(\frac{\varepsilon M_{n,j}}{\delta_{n,j}}\right)^{2/\gamma_j},\mathcal{X}\times\mathcal{T},||\cdot||\right) \leq C_1\left(\frac{\varepsilon M_{n,j}}{\delta_{n,j}}\wedge 1\right)^{-C_{2,j}},$$

where $N_{[]}\left(\varepsilon,\mathcal{F}_{n,j},L_2(P)\right)$ denotes the ε-bracketing number of the class $\mathcal{F}_{n,j}$ with respect to the $L_2(P)$-norm and $N\left(\varepsilon,\mathcal{X}\times\mathcal{T},||\cdot||\right)$ denotes the ε-covering number of the space $\mathcal{X}\times\mathcal{T}$ with respect to the Euclidean norm $||\cdot||$. The last inequality follows by the assumption that $\mathcal{X}$ and $\mathcal{T}$ are compact subsets of a Euclidean space. The class $\mathcal{F}_{n,j}$ is uniformly bounded by 1/2.

Let $\{[\beta^{a}_{n,x_k,\tau_k,j}(\cdot,(\cdot - x_k)/h)/M_{n,j} - \Delta_k(\cdot,\cdot)/M_{n,j}, \beta^{a}_{n,x_k,\tau_k,j}(\cdot,(\cdot - x_k)/h)/M_{n,j} + \Delta_k(\cdot,\cdot)/M_{n,j}] : k = 1,\cdots,N_{n,j}\}$ constitutes ε-brackets, where $\Delta_k(Y_{ij},X_i) \equiv \sup|\beta^{a}_{n,x,\tau,j}(Y_{ij},(X_i - x)/h) - \beta^{a}_{n,x_k,\tau_k,j}(Y_{ij},(X_i - x_k)/h)|$ and the supremum is over $(x,\tau)\in\mathcal{S}$ such that

$$\sqrt{||x - x_k||^2 + ||\tau - \tau_k||^2} \leq C_1(\varepsilon M_{n,j}/\delta_{n,j})^{2/\gamma_j}.$$

By the previous covering number bound, we can take $N_{n,j} \leq C_1\left((\varepsilon M_{n,j}/\delta_{n,j})\wedge 1\right)^{-C_{2,j}}$, and

$$\mathbf{E}\Delta_k^2(Y_{ij},X_i)M_{n,j}^{-2} < \varepsilon^2.$$

Note that for any $k \geq 2$,

$$\mathbf{E}\left[|b^a_{n,ij}(x,\tau)/M_{n,j}|^k\right] \leq \mathbf{E}\left[b^2_{n,ij}(x,\tau)\right]/M^2_{n,j} \leq CM^{-2}_{n,j}h^d = C(\log n)/n,$$

by the fact that $|b^a_{n,ij}(x,\tau)/M_{n,j}| \leq 1/2$. Furthermore,

$$\mathbf{E}\left[|\Delta_k(Y_{ij},X_i)/M_{n,j}|^k\right] \leq \mathbf{E}\left[\Delta^2_k(Y_{ij},X_i)/M^2_{n,j}\right] \leq \varepsilon^2,$$

where the first inequality follows because $|\Delta_k(Y_{ij},X_i)/M_{n,j}| \leq 1$. Therefore, by Theorem 6.8 of Massart (2007), we have (from sufficiently large n on)

$$\text{(A.10)} \quad \mathbf{E}\left[\sup_{(x,\tau)\in\mathcal{S}}\left|\frac{1}{M_{n,j}\sqrt{n}}\sum_{i=1}^n\left(b^a_{n,ij}(x,\tau)-\mathbf{E}\left[b^a_{n,ij}(x,\tau)\right]\right)\right|\right]$$

$$\leq C_1\int_0^{\frac{C_2h^{d/2}}{M_{n,j}}}\left\{\left(-C_3\log\left(\frac{\varepsilon M_{n,j}}{\delta_{n,j}}\wedge 1\right)\right)\wedge n\right\}^{1/2}d\varepsilon - \frac{C_4}{\sqrt{n}}\log\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right),$$

where C_1, C_2, C_3, and C_4 are positive constants. (The inequality above follows because $\sqrt{\log n}/\sqrt{n} \to 0$ as $n \to \infty$.) The leading integral has a domain restricted to $[0, \delta_{n,j}/M_{n,j}]$, so that it is equal to

$$C_1\int_0^{\frac{C_2h^{d/2}}{M_{n,j}}\wedge\frac{\delta_{n,j}}{M_{n,j}}}\left\{\left(-C_3\log\left(\frac{\varepsilon M_{n,j}}{\delta_{n,j}}\right)\right)\wedge n\right\}^{1/2}d\varepsilon$$

$$= \frac{C_1\delta_{n,j}}{M_{n,j}}\int_0^{\frac{C_2h^{d/2}}{\delta_{n,j}}\wedge 1}\sqrt{(-C_3\log\varepsilon)\wedge n}\,d\varepsilon$$

$$= O\left(\frac{\delta_{n,j}}{M_{n,j}}\left(\frac{h^{d/2}}{\delta_{n,j}}\wedge 1\right)\sqrt{-\log\left(\frac{h^{d/2}}{\delta_{n,j}}\wedge 1\right)}\right).$$

After multiplying by $M_{n,j}/h^{d/2}$, the last term is of order

$$O\left(\left(1\wedge\frac{\delta_{n,j}}{h^{d/2}}\right)\sqrt{-\log\left(\frac{h^{d/2}}{\delta_{n,j}}\wedge 1\right)}\right) = O\left(\sqrt{-\log\left(\frac{h^{d/2}}{\delta_{n,j}}\wedge 1\right)}\right) = O(\sqrt{\log n}),$$

because $\delta_{n,j} = n^{s_{1,j}}$ and $h = n^{s_2}$ for some $s_{1,j}, s_2 \in \mathbf{R}$.
Also, note that after multiplying by $M_{n,j}/h^{d/2} = \sqrt{n}/\sqrt{\log n}$, the last term in (A.10) (with minus sign) becomes

$$-\frac{C_4}{\sqrt{\log n}}\log\left(\frac{\sqrt{\log n}}{\sqrt{n}}\right) \leq \frac{C_4\sqrt{\log n}}{2} - \frac{C_4\log\sqrt{\log n}}{\sqrt{\log n}} = O\left(\sqrt{\log n}\right),$$

where the inequality follows because $\sqrt{\log n} \geq 1$ for all $n \geq e \equiv \exp(1)$. Collecting the results for both the terms on the right hand side of (A.10), we obtain the desired result of Step 2.

(ii) Define $b^*_{n,ij}(x,\tau) \equiv \beta_{n,x,\tau,j}(Y^*_{ij},(X^*_i - x)/h)$. By Assumptions B1 and B3, it suffices to show that

$$\sup_{(x,\tau)\in\mathcal{S}}\left|\frac{1}{\sqrt{nh^d}}\sum_{i=1}^n\left(b^*_{n,ij}(x,\tau)-\mathbf{E}^*\left[b^*_{n,ij}(x,\tau)\right]\right)\right| = O_{P^*}(\sqrt{\log n}),\ \mathcal{P}\text{-uniformly.}$$

Using Le Cam's Poissonization lemma in Giné and Zinn (1990) (Proposition 2.2 on p.855) and following the arguments in the proof of Theorem 2.2 of Giné (1997), we deduce that

$$\begin{aligned}&\mathbf{E}\left[\mathbf{E}^*\left(\sup_{(x,\tau)\in\mathcal{S}}\left|\frac{1}{\sqrt{nh^d}}\sum_{i=1}^n\left(b^*_{n,ij}(x,\tau)-\mathbf{E}^*\left[b^*_{n,ij}(x,\tau)\right]\right)\right|\right)\right]\\ \leq\ &\frac{e}{e-1}\mathbf{E}\left[\sup_{(x,\tau)\in\mathcal{S}}\left|\frac{1}{\sqrt{nh^d}}\sum_{i=1}^n(N_i-1)\left\{b_{n,ij}(x,\tau)-\frac{1}{n}\sum_{k=1}^n b_{n,kj}(x,\tau)\right\}\right|\right],\end{aligned}$$

where N_i's are i.i.d. Poisson random variables with mean 1 and independent of $\{(X_i,Y_i)\}_{i=1}^n$. The last expectation is bounded by

$$\begin{aligned}&\mathbf{E}\left[\sup_{(x,\tau)\in\mathcal{S}}\left|\frac{1}{\sqrt{nh^d}}\sum_{i=1}^n\left\{(N_i-1)\,b_{n,ij}(x,\tau)-\mathbf{E}\left[(N_i-1)\,b_{n,ij}(x,\tau)\right]\right\}\right|\right]\\ &+\mathbf{E}\left[\sup_{(x,\tau)\in\mathcal{S}}\left|\frac{1}{n}\sum_{i=1}^n(N_i-1)\right|\left|\frac{1}{\sqrt{nh^d}}\sum_{k=1}^n\left(b_{n,kj}(x,\tau)-\mathbf{E}\left[b_{n,kj}(x,\tau)\right]\right)\right|\right].\end{aligned}$$

Using the same arguments as in the proof of (i), we find that the first expectation is $O\left(\sqrt{\log n}\right)$ uniformly in $P\in\mathcal{P}$. Using independence, we write the second expectation as

$$\mathbf{E}\left[\left|\frac{1}{n}\sum_{i=1}^n(N_i-1)\right|\right]\cdot\mathbf{E}\left[\sup_{(x,\tau)\in\mathcal{S}}\left|\frac{1}{\sqrt{nh^d}}\sum_{k=1}^n\left(b_{n,kj}(x,\tau)-\mathbf{E}\left[b_{n,kj}(x,\tau)\right]\right)\right|\right]$$

which, as shown in the proof of part (i), is $O(\sqrt{\log n})$, uniformly in $P\in\mathcal{P}$. ∎

For further proofs, we introduce new notation. Define for any positive sequences $c_{n,1}$ and $c_{n,2}$, and any $\mathbf{v}\in\mathbf{R}^J$,

$$\bar{\Lambda}_{x,\tau}(\mathbf{v}) \equiv \sum_{A\in\mathcal{N}_J}\Lambda_{A,p}(\mathbf{v})1\{(x,\tau)\in B_{n,A}(c_{n,1},c_{n,2})\}. \tag{A.11}$$

We let

$$\begin{aligned}a_n^R(c_{n,1},c_{n,2}) &\equiv \int_{\mathcal{X}\times\mathcal{T}}\mathbf{E}\left[\bar{\Lambda}_{x,\tau}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x))\right]dQ(x,\tau),\ \text{and}\\ a_n^{R*}(c_{n,1},c_{n,2}) &\equiv \int_{\mathcal{X}\times\mathcal{T}}\mathbf{E}^*\left[\bar{\Lambda}_{x,\tau}(\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x))\right]dQ(x,\tau),\end{aligned} \tag{A.12}$$

where $\mathbf{z}_{N,\tau}(x)$ and $\mathbf{z}^*_{N,\tau}(x)$ are random vectors whose j-th entry is respectively given by

$$z_{N,\tau,j}(x) \equiv \frac{1}{nh^d}\sum_{i=1}^{N}\left(\beta_{n,x,\tau,j}\left(Y_{ij},\frac{X_i-x}{h}\right)-\mathbf{E}\left[\beta_{n,x,\tau,j}\left(Y_{ij},\frac{X_i-x}{h}\right)\right]\right) \text{ and}$$
$$z^*_{N,\tau,j}(x) \equiv \frac{1}{nh^d}\sum_{i=1}^{N}\left(\beta_{n,x,\tau,j}\left(Y^*_{ij},\frac{X^*_i-x}{h}\right)-\mathbf{E}^*\left[\beta_{n,x,\tau,j}\left(Y^*_{ij},\frac{X^*_i-x}{h}\right)\right]\right),$$

and N is a Poisson random variable with mean n and independent of $\{Y_i, X_i\}_{i=1}^{\infty}$. We also define

$$a_n(c_{n,1},c_{n,2}) \equiv \int \mathbf{E}\left[\bar{\Lambda}_{x,\tau}(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0))\right]dQ(x,\tau).$$

(See Section 6.3 for the definition of $\mathbb{W}^{(1)}_{n,\tau,\tau}(x,u)$.)

Lemma A1. *Suppose that Assumptions A6(i) and B4 hold and let $c_{n,1}$ and $c_{n,2}$ be any nonnegative sequences. Then*

$$\left|a_n^R(c_{n,1},c_{n,2})-a_n(c_{n,1},c_{n,2})\right| = o(h^{d/2}),\ \textit{uniformly in } P\in\mathcal{P},\ \textit{and}$$
$$\left|a_n^{R*}(c_{n,1},c_{n,2})-a_n(c_{n,1},c_{n,2})\right| = o_P(h^{d/2}),\ \mathcal{P}\textit{-uniformly.}$$

Proof of Lemma A1. The proof is essentially the same as the proof of Lemma D12 in Appendix D. ∎

For any given nonnegative sequences $c_{n,1}, c_{n,2}$, we define

$$\sigma_n^2(c_{n,1},c_{n,2}) \equiv \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{\mathcal{X}} \bar{C}_{\tau_1,\tau_2}(x)dxd\tau_1 d\tau_2, \tag{A.13}$$

where

$$\bar{C}_{\tau_1,\tau_2}(x) \equiv \int_{\mathcal{U}} Cov\left(\bar{\Lambda}_{n,x,\tau_1}(\mathbb{W}^{(1)}_{n,\tau_1,\tau_2}(x,u)),\bar{\Lambda}_{n,x,\tau_2}(\mathbb{W}^{(2)}_{n,\tau_1,\tau_2}(x,u))\right)du.$$

Let

$$\bar{\theta}_n(c_{n,1},c_{n,2}) \equiv \int \bar{\Lambda}_{x,\tau}\left(\hat{\mathbf{s}}_\tau(x)\right)dQ(x,\tau), \tag{A.14}$$

and

$$\bar{\theta}^*_n(c_{n,1},c_{n,2}) \equiv \int \bar{\Lambda}_{x,\tau}\left(\hat{\mathbf{s}}^*_\tau(x)\right)dQ(x,\tau). \tag{A.15}$$

From here on, for any sequence of random quantities Z_n and a random vector Z, we write

$$Z_n \xrightarrow{d} N(0,1),\ \mathcal{P}_0\text{-uniformly},$$

if for each $t>0$,

$$\sup_{P\in\mathcal{P}_0}\left|P\left\{Z_n\le t\right\}-\Phi(t)\right| = o(1).$$

And for any sequence of bootstrap quantities Z_n^* and a random vector Z, we write

$$Z_n^* \overset{d^*}{\to} N(0,1),\ \mathcal{P}_0\text{-uniformly},$$

if for each $t > 0$,

$$|P^* \{Z_n^* \le t\} - \Phi(t)| = o_{P^*}(1),\ \mathcal{P}_0\text{-uniformly}.$$

Lemma A2. *(i) Suppose that Assumptions A1-A3, A4(i), and A5-A6 are satisfied. Then for any sequences* $c_{n,1}, c_{n,2} > 0$ *such that* $\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_0} \sigma_n^2(c_{n,1}, c_{n,2}) > 0$ *and* $\sqrt{\log n}/c_{n,2} \to 0$*, as* $n \to \infty$*,*

$$h^{-d/2}\left(\frac{\bar{\theta}_n(c_{n,1}, c_{n,2}) - a_n^R(c_{n,1}, c_{n,2})}{\sigma_n(c_{n,1}, c_{n,2})}\right) \overset{d}{\to} N(0,1),\ \mathcal{P}_0\text{-uniformly.}$$

(ii) Suppose that Assumptions A1-A3, A4(i), A5-A6, B1 and B4 are satisfied. Then for any sequences $c_{n,1}, c_{n,2} \ge 0$ *such that* $\liminf_{n\to\infty} \inf_{P\in\mathcal{P}_0} \sigma_n^2(c_{n,1}, c_{n,2}) > 0$ *and* $\sqrt{\log n}/c_{n,2} \to 0$*, as* $n \to \infty$*,*

$$h^{-d/2}\left(\frac{\bar{\theta}_n^*(c_{n,1}, c_{n,2}) - a_n^{R*}(c_{n,1}, c_{n,2})}{\sigma_n(c_{n,1}, c_{n,2})}\right) \overset{d^*}{\to} N(0,1),\ \mathcal{P}_0\text{-uniformly.}$$

Proof of Lemma A2. (i) By Lemma 1, we have (with probability approaching one)

$$\bar{\theta}_n(c_{n,1}, c_{n,2}) = \sum_{A\in\mathcal{N}_J} \int_{B_{n,A}(c_{n,1},c_{n,2})} \Lambda_p(\hat{\mathbf{s}}_\tau(x))dQ(x,\tau) = \sum_{A\in\mathcal{N}_J} \int_{B_{n,A}(c_{n,1},c_{n,2})} \Lambda_{A,p}(\hat{\mathbf{s}}_\tau(x))dQ(x,\tau).$$

Note that $a_n^R(c_{n,1}, c_{n,2}) = \sum_{A\in\mathcal{N}_J} a_{n,A}^R(c_{n,1}, c_{n,2})$, where

$$a_{n,A}^R(c_{n,1}, c_{n,2}) \equiv \int_{B_{n,A}(c_{n,1},c_{n,2})} \mathbf{E}\left[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x))\right] dQ(x,\tau).$$

Using Assumption A1, we find that $h^{-d/2}\{\bar{\theta}_n(c_{n,1}, c_{n,2}) - a_n^R(c_{n,1}, c_{n,2})\}$ is equal to

$$h^{-d/2} \sum_{A\in\mathcal{N}_J} \{\zeta_{n,A}(B_{n,A}(c_{n,1}, c_{n,2})) - \mathbf{E}\zeta_{N,A}(B_{n,A}(c_{n,1}, c_{n,2}))\} + o_P(1),$$

where for any Borel set $B \subset \mathcal{S}$,

$$\begin{aligned} \zeta_{n,A}(B) &\equiv \int_B \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{n,\tau}(x))dQ(x,\tau), \\ \zeta_{N,A}(B) &\equiv \int_B \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x))dQ(x,\tau), \end{aligned}$$

and

$$\mathbf{z}_{n,\tau}(x) \equiv \frac{1}{nh^d}\sum_{i=1}^n \beta_{n,x,\tau}(Y_i, (X_i - x)/h) - \frac{1}{h^d}\mathbf{E}\left[\beta_{n,x,\tau}(Y_i, (X_i - x)/h)\right],$$

with

$$\beta_{n,x,\tau}(Y_i, (X_i - x)/h) = (\beta_{n,x,\tau,1}(Y_{i1}, (X_i - x)/h), \cdots, \beta_{n,x,\tau,J}(Y_{iJ}, (X_i - x)/h))^\top.$$

We take $0<\bar{\varepsilon}_l\to 0$ as $l\to\infty$ and take $\mathcal{C}_l\subset\mathbf{R}^d$ such that

$$0<P\left\{X_i\in\mathbf{R}^d\backslash\mathcal{C}_l\right\}\le\bar{\varepsilon}_l,$$

and $Q((\mathcal{X}\backslash\mathcal{C}_l)\times\mathcal{T})\to 0$ as $l\to\infty$. Such a sequence $\{\bar{\varepsilon}_l\}_{l=1}^{\infty}$ exists by Assumption A6(ii) by the condition that $\mathcal{S}$ is compact. We write

$$\begin{aligned}
&\frac{h^{-d/2}\sum_{A\in\mathcal{N}_J}\{\zeta_{n,A}(B_{n,A}(c_{n,1},c_{n,2}))-\mathbf{E}\zeta_{N,A}(B_{n,A}(c_{n,1},c_{n,2}))\}}{\sigma_n^2(c_{n,1},c_{n,2})} \qquad \text{(A.16)}\\
&= \frac{h^{-d/2}\sum_{A\in\mathcal{N}_J}\{\zeta_{n,A}(B_{n,A}(c_{n,1},c_{n,2})\cap(\mathcal{C}_l\times\mathcal{T}))-\mathbf{E}\zeta_{N,A}(B_{n,A}(c_{n,1},c_{n,2})\cap(\mathcal{C}_l\times\mathcal{T}))\}}{\sigma_n^2(c_{n,1},c_{n,2})}\\
&\quad+\frac{h^{-d/2}\sum_{A\in\mathcal{N}_J}\{\zeta_{n,A}(B_{n,A}(c_{n,1},c_{n,2})\backslash(\mathcal{C}_l\times\mathcal{T}))-\mathbf{E}\zeta_{N,A}(B_{n,A}(c_{n,1},c_{n,2})\backslash(\mathcal{C}_l\times\mathcal{T}))\}}{\sigma_n^2(c_{n,1},c_{n,2})}\\
&= A_{1n}+A_{2n},\ \text{say.}
\end{aligned}$$

As for A_{2n}, we apply Lemma C7 in Appendix C, and the condition that $Q((\mathcal{X}\backslash\mathcal{C}_l)\times\mathcal{T})\to 0$, as $l\to\infty$, and

$$\text{liminf}_{n\to\infty}\text{inf}_{P\in\mathcal{P}_0}\sigma_n(c_{1n},c_{2n})>0,$$

to deduce that $A_{2n}=o_P(1)$, as $n\to\infty$ and then $l\to\infty$. As for A_{1n}, first observe that as $n\to\infty$ and then $l\to\infty$,

$$\left|\sigma_n^2(c_{n,1},c_{n,2})-\bar{\sigma}_{n,l}^2(c_{n,1},c_{n,2})\right|\to 0, \qquad \text{(A.17)}$$

where $\bar{\sigma}_{n,l}^2(c_{n,1},c_{n,2})$ is equal to $\sigma_n^2(c_{n,1},c_{n,2})$ except that $B_{n,A}(c_{n,1},c_{n,2})$'s are replaced by $B_{n,A}(c_{n,1},c_{n,2})\cap(\mathcal{C}_l\times\mathcal{T})$. The convergence follows by Assumption 6(i). Also by Lemma C9(i) and the convergence in (A.17) and the fact that

$$\liminf_{n\to\infty}\inf_{P\in\mathcal{P}_0}\sigma_n^2(c_{n,1},c_{n,2})>0,$$

we have

$$A_{1n}\xrightarrow{d}N(0,1),\ \mathcal{P}_0\text{-uniformly},$$

as $n\to\infty$ and as $l\to\infty$. Hence we obtain (i).

(ii) The proof can be done in the same way as in the proof of (i), using Lemmas D7 and D9(i) in Appendix D instead of Lemmas C7 and C9(i) in Appendix C. ∎

Lemma A3. *Suppose that Assumptions A1-A5 hold. Then for any sequences $c_{n,L},c_{n,U}>0$ satisfying Assumption A4(ii), and for each $A\in\mathcal{N}_J$,*

$$\inf_{P\in\mathcal{P}}P\left\{B_{n,A}(c_{n,L},c_{n,U})\subset\hat{B}_A(\hat{c}_n)\subset B_{n,A}(c_{n,U},c_{n,L})\right\}\to 1,\ \textit{as } n\to\infty.$$

Proof of Lemma A3. By using Assumptions A3-A5, and following the proof of Theorem 2, Claim 1 in Linton, Song, and Whang (2010), we can complete the proof. Details are omitted. ∎

Define for $c_{n,1}, c_{n,2} > 0$,

$$T_n(c_{n,1}, c_{n,2}) \equiv h^{-d/2}\left(\frac{\bar{\theta}_n(c_{n,1}, c_{n,2}) - a_n(c_{n,1}, c_{n,2})}{\sigma_n(c_{n,1}, c_{n,2})}\right) \text{ and}$$
$$T_n^*(c_{n,1}, c_{n,2}) \equiv h^{-d/2}\left(\frac{\bar{\theta}_n^*(c_{n,1}, c_{n,2}) - a_n(c_{n,1}, c_{n,2})}{\sigma_n(c_{n,1}, c_{n,2})}\right).$$

We introduce critical values for the finite sample distribution of $\hat{\theta}$ as follows:

$$\gamma_n^\alpha(c_{n,1}, c_{n,2}) \equiv \inf\left\{c \in \mathbf{R} : P\left\{T_n(c_{n,1}, c_{n,2}) \leq c\right\} > 1 - \alpha\right\}.$$

Similarly, let us introduce bootstrap critical values:

$$\gamma_n^{\alpha*}(c_{n,1}, c_{n,2}) \equiv \inf\left\{c \in \mathbf{R} : P^*\left\{T_n^*(c_{n,1}, c_{n,2}) \leq c\right\} > 1 - \alpha\right\}. \tag{A.18}$$

Finally, we introduce asymptotic critical values: $\gamma_\infty^\alpha \equiv \Phi^{-1}(1-\alpha)$, where Φ denotes the standard normal CDF.

Lemma A4. *Suppose that Assumptions A1-A3, A4(i), and A5-A6 hold. Then the following holds.*
(i) For any $c_{n,1}, c_{n,2} \to \infty$ such that

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \sigma_n^2(c_{n,1}, c_{n,2}) > 0,$$

it is satisfied that

$$\sup_{P\in\mathcal{P}} \left|\gamma_n^\alpha(c_{n,1}, c_{n,2}) - \gamma_\infty^\alpha\right| \to 0, \text{ as } n \to \infty.$$

(ii) Suppose further that Assumptions B1 and B4 hold. Then for any $c_{n,1}, c_{n,2} \to \infty$ such that

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \sigma_n^2(c_{n,1}, c_{n,2}) > 0,$$

it is satisfied that

$$\sup_{P\in\mathcal{P}} \left|\gamma_n^{\alpha*}(c_{n,1}, c_{n,2}) - \gamma_\infty^\alpha\right| \to 0, \text{ as } n \to \infty.$$

Proof of Lemma A4. (i) The statement immediately follows from the first statement of Lemma A2(i) and Lemma A1.
(ii) We show only the second statement. Fix $a > 0$. Let us introduce two events:

$$E_{n,1} \equiv \left\{\gamma_n^{\alpha*}(c_{n,1}, c_{n,2}) - \gamma_\infty^\alpha < -a\right\} \text{ and } E_{n,2} \equiv \left\{\gamma_n^{\alpha*}(c_{n,1}, c_{n,2}) - \gamma_\infty^\alpha > a\right\}.$$

On the event $E_{n,1}$, we have

$$\begin{aligned}\alpha &= P^*\left\{h^{-d/2}\left(\frac{\bar{\theta}_n^*(c_{n,1},c_{n,2})-a_n(c_{n,1},c_{n,2})}{\sigma_n(c_{n,1},c_{n,2})}\right)>\gamma_n^{\alpha*}(c_{n,1},c_{n,2})\right\}\\ &\geq P^*\left\{h^{-d/2}\left(\frac{\bar{\theta}_n^*(c_{n,1},c_{n,2})-a_n(c_{n,1},c_{n,2})}{\sigma_n(c_{n,1},c_{n,2})}\right)>\gamma_\infty^{\alpha}-a\right\}.\end{aligned}$$

By Lemma A2(ii) and Lemma A1, the last probability is equal to

$$1-\Phi\left(\gamma_\infty^\alpha-a\right)+o_P(1)>\alpha+o_P(1),$$

where $o_P(1)$ is uniform over $P\in\mathcal{P}$ and the last strict inequality follows by the definition of γ_∞^α and $a>0$. Hence $\sup_{P\in\mathcal{P}}PE_{n,1}\to 0$ as $n\to\infty$. Similarly, on the event $E_{n,2}$, we have

$$\begin{aligned}\alpha &= P^*\left\{h^{-d/2}\left(\frac{\bar{\theta}_n^*(c_{n,1},c_{n,2})-a_n(c_{n,1},c_{n,2})}{\sigma_n(c_{n,1},c_{n,2})}\right)>\gamma_n^{\alpha*}(c_{n,1},c_{n,2})\right\}\\ &\leq P^*\left\{h^{-d/2}\left(\frac{\bar{\theta}_n^*(c_{n,1},c_{n,2})-a_n(c_{n,1},c_{n,2})}{\sigma_n(c_{n,1},c_{n,2})}\right)>\gamma_\infty^{\alpha}+a\right\}.\end{aligned}$$

By the first statement of Lemma A2(ii) and Lemma A1, the last bootstrap probability is bounded by

$$1-\Phi\left(\gamma_\infty^\alpha+a\right)+o_P(1)<\alpha+o_P(1),$$

so that we have $\sup_{P\in\mathcal{P}}PE_{n,2}\to 0$ as $n\to\infty$. We conclude that

$$\sup_{P\in\mathcal{P}}P\left\{\left|\gamma_n^{\alpha*}(c_{n,1},c_{n,2})-\gamma_\infty^\alpha\right|>a\right\}=\sup_{P\in\mathcal{P}}\left(PE_{n,1}+PE_{n,2}\right)\to 0,$$

as $n\to\infty$, obtaining the desired result. ∎

Proof of Theorem 1. By Lemma 1, we have

$$\inf_{P\in\mathcal{P}_0}P\left\{\hat{\theta}=\sum_{A\in\mathcal{N}_J}\int_{B_{n,A}(c_{n,L},c_{n,U})}\Lambda_{A,p}\left(\hat{\mathbf{u}}_\tau(x)\right)dQ(x,\tau)\right\}\to 1,$$

as $n\to\infty$. Since under the null hypothesis, we have $v_{n,\tau,j}(\cdot)/\hat{\sigma}_{\tau,j}(\cdot)\leq 0$ for all $j\in\mathbb{N}_J$, with probability approaching one by Assumption A5, we have

$$\begin{aligned}&\sum_{A\in\mathcal{N}_J}\int_{B_{n,A}(c_{n,L},c_{n,U})}\Lambda_{A,p}\left(\hat{\mathbf{u}}_\tau(x)\right)dQ(x,\tau)\\ \leq\ &\sum_{A\in\mathcal{N}_J}\int_{B_{n,A}(c_{n,L},c_{n,U})}\Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)\right)dQ(x,\tau)\equiv\bar{\theta}_n(c_{n,L},c_{n,U}).\end{aligned}$$

Thus, we have as $n\to\infty$,

$$\inf_{P\in\mathcal{P}_0}P\left\{\hat{\theta}\leq\bar{\theta}_n(c_{n,L},c_{n,U})\right\}\to 1. \tag{A.19}$$

Let the $(1-\alpha)$-th percentile of the bootstrap distribution of

$$\bar{\theta}_n^*(c_{n,L}, c_{n,U}) = \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,L}, c_{n,U})} \Lambda_{A,p}(\hat{\mathbf{s}}_\tau^*(x)) dQ(x,\tau)$$

be denoted by $\bar{c}_{n,L}^{\alpha*}$. By Lemma A3 and Assumption A4(ii), with probability approaching one,

$$\sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,L}, c_{n,U})} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau^*(x)\right) dQ(x,\tau) \leq \sum_{A \in \mathcal{N}_J} \int_{\hat{B}_A(\hat{c}_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau^*(x)\right) dQ(x,\tau). \tag{A.20}$$

This implies that as $n \to \infty$,

$$\inf_{P \in \mathcal{P}} P\left\{c_\alpha^* \geq \bar{c}_{n,L}^{\alpha*}\right\} \to 1. \tag{A.21}$$

There exists a sequence of probabilities $\{P_n\}_{n \geq 1} \subset \mathcal{P}_0$ such that

$$\begin{aligned} \limsup_{n\to\infty} \sup_{P \in \mathcal{P}_0} P\left\{\hat{\theta} > c_{\alpha,\eta}^*\right\} &= \limsup_{n\to\infty} P_n\left\{\hat{\theta} > c_{\alpha,\eta}^*\right\} \\ &= \lim_{n\to\infty} P_{w_n}\left\{\hat{\theta}_{w_n} > c_{w_n,\alpha,\eta}^*\right\}, \end{aligned} \tag{A.22}$$

where $\{w_n\} \subset \{n\}$ is a certain subsequence, and $\hat{\theta}_{w_n}$ and $c_{w_n,\alpha,\eta}^*$ are the same as $\hat{\theta}$ and $c_{\alpha,\eta}^*$ except that the sample size n is now replaced by w_n.

By Assumption A6(i), $\{\sigma_n(c_{n,L}, c_{n,U})\}_{n \geq 1}$ is a bounded sequence. Therefore, there exists a subsequence $\{u_n\}_{n \geq 1} \subset \{w_n\}_{n \geq 1}$, such that $\sigma_{u_n}(c_{u_n,L}, c_{u_n,U})$ converges. We consider two cases:

Case 1: $\lim_{n\to\infty} \sigma_{u_n}(c_{u_n,L}, c_{u_n,U}) > 0$, and

Case 2: $\lim_{n\to\infty} \sigma_{u_n}(c_{u_n,L}, c_{u_n,U}) = 0$.

In both case, we will show below that

$$\limsup_{n\to\infty} P_{u_n}\{\hat{\theta}_{u_n} > c_{u_n,\alpha,\eta}^*\} \leq \alpha. \tag{A.23}$$

Since along $\{w_n\}$, $P_{w_n}\{\hat{\theta}_{w_n} > c_{w_n,\alpha,\eta}^*\}$ converges, it does so along any subsequence of $\{w_n\}$. Therefore, the above limsup is equal to the last limit in (A.22). This completes the proof.

Proof of (A.23) in Case 1: We write $P_{u_n}\{\hat{\theta}_{u_n} > c_{u_n,\alpha,\eta}^*\}$ as

$$\begin{aligned} & P_{u_n}\left(h^{-d/2}\left(\frac{\hat{\theta}_{u_n} - a_{u_n}(c_{u_n,L}, c_{u_n,U})}{\sigma_{u_n}(c_{u_n,L}, c_{u_n,U})}\right) > h^{-d/2}\left(\frac{c_{u_n,\alpha,\eta}^* - a_{u_n}(c_{u_n,L}, c_{u_n,U})}{\sigma_{u_n}(c_{u_n,L}, c_{u_n,U})}\right)\right) \\ \leq\ & P_{u_n}\left(h^{-d/2}\left(\frac{\hat{\theta}_{u_n} - a_{u_n}(c_{u_n,L}, c_{u_n,U})}{\sigma_{u_n}(c_{u_n,L}, c_{u_n,U})}\right) > h^{-d/2}\left(\frac{\bar{c}_{u_n,L}^{\alpha*} - a_{u_n}(c_{u_n,L}, c_{u_n,U})}{\sigma_{u_n}(c_{u_n,L}, c_{u_n,U})}\right)\right) + o(1), \end{aligned}$$

where the inequality follows by the fact that $c^*_{\alpha,\eta} \geq c^*_\alpha \geq \bar{c}^{\alpha*}_{n,L}$ with probability approaching one by (A.21). Using (A.19), we bound the last probability by

(A.24)
$$P_{u_n}\left\{h^{-d/2}\left(\frac{\bar{\theta}_{u_n}(c_{u_n,L},c_{u_n,U}) - a_{u_n}(c_{u_n,L},c_{u_n,U})}{\sigma_{u_n}(c_{u_n,L},c_{u_n,U})}\right) > h^{-d/2}\left(\frac{\bar{c}^{\alpha*}_{u_n,L} - a_{u_n}(c_{u_n,L},c_{u_n,U})}{\sigma_{u_n}(c_{u_n,L},c_{u_n,U})}\right)\right\}+o(1).$$

Therefore, since $\lim_{n\to\infty}\sigma_{u_n}(c_{u_n,L},c_{u_n,U}) > 0$, by Lemmas A2 and A4, we rewrite the last probability in (A.24) as

$$\begin{aligned}
& P_{u_n}\left\{h^{-d/2}\left(\frac{\bar{\theta}_{u_n}(c_{u_n,L},c_{u_n,U}) - a_{u_n}(c_{u_n,L},c_{u_n,U})}{\sigma_{u_n}(c_{u_n,L},c_{u_n,U})}\right) > \gamma^{\alpha*}_{u_n}(c_{u_n,L},c_{u_n,U})\right\} + o(1) \\
= \; & P_{u_n}\left\{h^{-d/2}\left(\frac{\bar{\theta}_{u_n}(c_{u_n,L},c_{u_n,U}) - a_{u_n}(c_{u_n,L},c_{u_n,U})}{\sigma_{u_n}(c_{u_n,L},c_{u_n,U})}\right) > \gamma^{\alpha}_{\infty}\right\} + o(1) = \alpha + o(1).
\end{aligned}$$

This completes the proof of Step 1.

Proof of (A.23) in Case 2: First, observe that

$$a^*_{u_n}(c_{u_n,L},c_{u_n,U}) \leq a^*_{u_n}(\hat{c}_{u_n}),$$

with probability approaching one by Lemma A3. Hence using this and (A.19),

$$\begin{aligned}
P_{u_n}\left\{\hat{\theta}_{u_n} > c^*_{u_n,\alpha,\eta}\right\} &= P_{u_n}\left\{h^{-d/2}\left(\hat{\theta}_{u_n} - a_{u_n}(c_{u_n,L},c_{u_n,U})\right) > h^{-d/2}\left(c^*_{u_n,\alpha,\eta} - a_{u_n}(c_{u_n,L},c_{u_n,U})\right)\right\} \\
&\leq P_{u_n}\left\{\begin{array}{c} h^{-d/2}\left(\bar{\theta}_{u_n}(c_{u_n,L},c_{u_n,U}) - a_{u_n}(c_{u_n,L},c_{u_n,U})\right) \\ > h^{-d/2}\left(h^{d/2}\eta + a^*_{u_n}(c_{u_n,L},c_{u_n,U}) - a_{u_n}(c_{u_n,L},c_{u_n,U})\right) \end{array}\right\} + o(1).
\end{aligned}$$

By Lemma A1, the leading probability is equal to

$$P_{u_n}\left\{h^{-d/2}\left(\bar{\theta}_{u_n}(c_{u_n,L},c_{u_n,U}) - a_{u_n}(c_{u_n,L},c_{u_n,U})\right) > \eta + o_P(1)\right\} + o(1).$$

Since $\eta > 0$ and $\lim_{n\to\infty}\sigma_{u_n}(c_{u_n,L},c_{u_n,U}) = 0$, the leading probability vanishes by Lemma C9(ii). ∎

Proof of Theorem 2. We focus on probabilities $P \in \mathcal{P}_n(\lambda_n, q_n) \cap \mathcal{P}_0$. Recalling the definition of $\mathbf{u}_{n,\tau}(x;\hat{\sigma}) \equiv [r_{n,j}v_{n,\tau,j}(x)/\hat{\sigma}_{\tau,j}(x)]_{j\in\mathbb{N}_J}$ and applying Lemma 1 along with the condition that

$$\sqrt{\log n}/c_{n,U} < \sqrt{\log n}/c_{n,L} \to 0,$$

as $n \to \infty$, we find that with probability approaching one,

$$\begin{aligned}\hat{\theta} &= \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,U},c_{n,L})} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x;\hat{\sigma})\right) dQ(x,\tau) \\ &= \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x;\hat{\sigma})\right) dQ(x,\tau) \\ &\quad + \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,U},c_{n,L}) \backslash B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x;\hat{\sigma})\right) dQ(x,\tau).\end{aligned}$$

Since under $P \in \mathcal{P}_0$, $\mathbf{u}_{n,\tau}(x;\hat{\sigma}) \leq 0$ for all $x \in \mathcal{S}$, with probability approaching one by Assumption 5, the last term multiplied by $h^{-d/2}$ is bounded by (from some large n on)

$$\begin{aligned}& h^{-d/2} \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(c_{n,U},c_{n,L}) \backslash B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)\right) dQ(x,\tau) \\ \leq\ & h^{-d/2} \sum_{A \in \mathcal{N}_J} \left(\sup_{(x,\tau) \in \mathcal{S}} ||\hat{\mathbf{s}}_\tau(x)|| \right)^p Q\left(B_{n,A}(c_{n,U},c_{n,L}) \backslash B_{n,A}(q_n)\right) \\ =\ & O_P\left(h^{-d/2} (\log n)^{p/2} \lambda_n\right) = o_P(1),\end{aligned}$$

where the second to the last equality follows because $Q\left(B_{n,A}(c_{n,U},c_{n,L}) \backslash B_{n,A}(q_n)\right) \leq \lambda_n$ by the definition of $\mathcal{P}_n(\lambda_n, q_n)$, and the last equality follows by (4.10).

On the other hand,

$$\begin{aligned}& h^{-d/2} \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x;\hat{\sigma})\right) dQ(x,\tau) \\ =\ & h^{-d/2} \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)\right) dQ(x,\tau) \\ & + h^{-d/2} \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_{n,\tau}(x;\hat{\sigma})\right) dQ(x,\tau) \\ & - h^{-d/2} \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)\right) dQ(x,\tau).\end{aligned}$$

From the definition of Λ_p in (4.1), the last difference (in absolute value) is bounded by

$$\begin{aligned}& C h^{-d/2} \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(q_n)} \left\|[\mathbf{u}_{n,\tau}(x;\hat{\sigma})]_A\right\| \left\|[\hat{\mathbf{s}}_\tau(x)]_A\right\|^{p-1} dQ(x,\tau) \\ & + C h^{-d/2} \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(q_n)} \left\|[\mathbf{u}_{n,\tau}(x;\hat{\sigma})]_A\right\| \left\|[\mathbf{u}_{n,\tau}(x;\hat{\sigma})]_A\right\|^{p-1} dQ(x,\tau),\end{aligned}$$

where $[a]_A$ is a vector a with the j-th entry is set to be zero for all $j \in \mathbb{N}_J \backslash A$ and $C > 0$ is a constant that does not depend on $n \geq 1$ or $P \in \mathcal{P}$. We have $\sup_{(x,\tau) \in B_{n,A}(q_n)} \left\|[\mathbf{u}_{n,\tau}(x;\hat{\sigma})]_A\right\| \leq$

$q_n(1+o_P(1))$, by the null hypothesis and by Assumption A5. Also, by Assumptions A3 and A5,

$$\sup_{(x,\tau)\in B_{n,A}(q_n)} \|[\hat{\mathbf{s}}_\tau(x)]_A\| = O_P\left(\sqrt{\log n}\right).$$

Therefore, we conclude that

$$\begin{aligned} & h^{-d/2}\sum_{A\in\mathcal{N}_J}\int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)+\mathbf{u}_{n,\tau}(x;\hat{\sigma})\right)dQ(x,\tau) \\ = \; & h^{-d/2}\sum_{A\in\mathcal{N}_J}\int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)\right)dQ(x,\tau) + O_P\left(h^{-d/2}q_n\{(\log n)^{(p-1)/2}+q_n^{p-1}\}\right). \end{aligned}$$

The last $O_P(1)$ term is $o_P(1)$ by the condition for q_n in (4.10). Thus we find that

$$\hat{\theta} = \bar{\theta}_n(q_n) + o_P(h^{d/2}), \tag{A.25}$$

where $\bar{\theta}_n(q_n) = \sum_{A\in\mathcal{N}_J}\int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)\right)dQ(x,\tau)$.

Now let us consider the bootstrap statistic. We write

$$\begin{aligned} \hat{\theta}^* &= \sum_{A\in\mathcal{N}_J}\int_{\hat{B}_A(\hat{c}_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}^*_\tau(x)\right)dQ(x,\tau) \\ &= \sum_{A\in\mathcal{N}_J}\int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}^*_\tau(x)\right)dQ(x,\tau) + \sum_{A\in\mathcal{N}_J}\int_{\hat{B}_A(\hat{c}_n)\backslash B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}^*_\tau(x)\right)dQ(x,\tau). \end{aligned}$$

By Lemma A3, we find that

$$\inf_{P\in\mathcal{P}} P\left\{\hat{B}_{n,A}(\hat{c}_n) \subset B_{n,A}(c_{n,U},c_{n,L})\right\} \to 1, \text{ as } n\to\infty,$$

so that

$$\sum_{A\in\mathcal{N}_J}\int_{\hat{B}_A(\hat{c}_n)\backslash B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}^*_\tau(x)\right)dQ(x,\tau) \le \sum_{A\in\mathcal{N}_J}\int_{B_{n,A}(c_{n,U},c_{n,L})\backslash B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}^*_\tau(x)\right)dQ(x,\tau),$$

with probability approaching one. The last term multiplied by $h^{-d/2}$ is bounded by

$$\begin{aligned} & h^{-d/2}\left(\sup_{(x,\tau)\in\mathcal{S}}||\hat{\mathbf{s}}^*_\tau(x)||\right)^p \sum_{A\in\mathcal{N}_J} Q\left(B_{n,A}(c_{n,U},c_{n,L})\backslash B_{n,A}(q_n)\right) \\ = \; & O_{P^*}\left(h^{-d/2}(\log n)^{p/2}\lambda_n\right) = o_{P^*}(1), \; \mathcal{P}_n(\lambda_n,q_n)\text{-uniformly}, \end{aligned}$$

where the second to the last equality follows by Assumption B2 and the definition of $\mathcal{P}_n(\lambda_n,q_n)$, and the last equality follows by (4.10). Thus, we conclude that

$$\frac{h^{-d/2}(\hat{\theta}^*-a_n(q_n))}{\sigma_n(q_n)} = \frac{h^{-d/2}\left(\bar{\theta}^*_n(q_n)-a_n(q_n)\right)}{\sigma_n(q_n)} + o_{P^*}(1), \; \mathcal{P}_n(\lambda_n,q_n)\text{-uniformly}, \tag{A.26}$$

where

$$\bar{\theta}^*(q_n) \equiv \sum_{A \in \mathcal{N}_J} \int_{B_{n,A}(q_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau^*(x)\right) dQ(x,\tau).$$

Using the same arguments, we also observe that

$$\hat{a}^* = \hat{a}^*(q_n) + o_P(h^{d/2}) = a_n(q_n) + o_P(h^{d/2}), \tag{A.27}$$

where the last equality uses Lemma A1. Let the $(1-\alpha)$-th percentile of the bootstrap distribution of $\bar{\theta}^*(q_n)$ be denoted by $\bar{c}_n^{\alpha*}(q_n)$. Then by (A.26), we have

$$\frac{h^{-d/2}\left(c_\alpha^* - a_n(q_n)\right)}{\sigma_n(q_n)} = \frac{h^{-d/2}\left(\bar{c}_n^{\alpha*}(q_n) - a_n(q_n)\right)}{\sigma_n(q_n)} + o_{P^*}(1),\ \mathcal{P}_n(\lambda_n, q_n)\text{-uniformly.} \tag{A.28}$$

By Lemma A4(ii) and by the condition that $\sigma_n(q_n) \geq \eta/\Phi^{-1}(1-\alpha)$, the leading term on the right hand side is equal to

$$\Phi^{-1}(1-\alpha) + o_{P^*}(1),\ \mathcal{P}_n(\lambda_n, q_n)\text{-uniformly.}$$

Note that

$$c_\alpha^* \geq h^{d/2}\eta + \hat{a}_n^* + o_P(h^{d/2}), \tag{A.29}$$

by the restriction $\sigma_n(q_n) \geq \eta/\Phi^{-1}(1-\alpha)$ in the definition of $\mathcal{P}_n(\lambda_n, q_n)$ and (A.27). Using this, and following the proof of Step 1 in the proof of Theorem 2, we deduce that

$$\begin{aligned}
& P\left\{h^{-d/2}\left(\frac{\hat{\theta} - a_n(q_n)}{\sigma_n(q_n)}\right) > h^{-d/2}\left(\frac{c_{\alpha,\eta}^* - a_n(q_n)}{\sigma_n(q_n)}\right)\right\} \\
= \ & P\left\{h^{-d/2}\left(\frac{\bar{\theta}_n(q_n) - a_n(q_n)}{\sigma_n(q_n)}\right) > h^{-d/2}\left(\frac{c_\alpha^* - a_n(q_n)}{\sigma_n(q_n)}\right)\right\} + o(1) \\
= \ & P\left\{h^{-d/2}\left(\frac{\bar{\theta}_n(q_n) - a_n(q_n)}{\sigma_n(q_n)}\right) > h^{-d/2}\left(\frac{\bar{c}_n^{\alpha*}(q_n) - a_n(q_n)}{\sigma_n(q_n)}\right)\right\} + o(1),
\end{aligned}$$

where the first equality uses (A.25), (A.29), and the second equality uses (A.28). Since $\sigma_n(q_n) \geq \eta/\Phi^{-1}(1-\alpha) > 0$ for all $P \in \mathcal{P}_n(\lambda_n, q_n) \cap \mathcal{P}_0$ by definition, using the same arguments in the proof of Lemma A4, we obtain that the last probability is equal to

$$\alpha + o(1),$$

uniformly over $P \in \mathcal{P}_n(\lambda_n, q_n) \cap \mathcal{P}_0$. ∎

Proof of Theorem 3. For any convex nonnegative map f on $\mathbf{R}^J$, we have $2f(b/2) \leq f(a+b) + f(-a)$. Hence we find that

$$\begin{aligned}\hat{\theta} &= \int \Lambda_p\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_\tau(x;\hat{\sigma})\right) dQ(x,\tau) \\ &\geq \frac{1}{2^{p-1}} \int \Lambda_p\left(\mathbf{u}_\tau(x;\hat{\sigma})\right) dQ(x,\tau) - \int \Lambda_p\left(-\hat{\mathbf{s}}_\tau(x)\right) dQ(x,\tau).\end{aligned}$$

From Assumption A3, the last term is $O_P((\log n)^{p/2})$. Using Assumption A3, we bound the leading integral from below by

$$\min_{j\in\mathbb{N}_J} r_{n,j}^p \left(\int \Lambda_p\left(\tilde{\mathbf{v}}_{n,\tau}(x)\right) dQ(x,\tau) \left\{ \frac{\int \Lambda_p\left(\mathbf{v}_{n,\tau}(x)\right) dQ(x,\tau)}{\int \Lambda_p\left(\tilde{\mathbf{v}}_{n,\tau}(x)\right) dQ(x,\tau)} - 1 \right\} + o_P(1) \right), \tag{A.30}$$

where $\mathbf{v}_{n,\tau}(x) \equiv [v_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x)]_{j\in\mathbb{N}_J}$ and $\tilde{\mathbf{v}}_{n,\tau}(x) \equiv [v_{\tau,j}(x)/\sigma_{n,\tau,j}(x)]_{j\in\mathbb{N}_J}$. Since

$$\text{liminf}_{n\to\infty} \int \Lambda_p\left(\tilde{\mathbf{v}}_{n,\tau}(x)\right) dQ(x,\tau) > 0,$$

we use Assumption C1 and apply the Dominated Convergence Theorem to write (A.30) as

$$\min_{j\in\mathbb{N}_J} r_{n,j}^p \int \Lambda_p\left(\tilde{\mathbf{v}}_{n,\tau}(x)\right) dQ(x,\tau) \left(1 + o_P(1)\right).$$

Since $\min_{j\in\mathbb{N}_J} r_{n,j} \to \infty$ as $n \to \infty$ and $\text{liminf}_{n\to\infty} \int \Lambda_p\left(\tilde{\mathbf{v}}_{n,\tau}(x)\right) dQ(x,\tau) > 0$, we have for any $M > 0$,

$$P\left\{ \frac{1}{2^{p-1}} \int \Lambda_p\left(\mathbf{u}_\tau(x;\hat{\sigma})\right) dQ(x,\tau) > M \right\} \to 1,$$

as $n \to \infty$. Also since $\sqrt{\log n}/\min_{j\in\mathbb{N}_J} r_{n,j} \to 0$ (Assumption A4(i)), Assumption A3 implies that for any $M > 0$,

$$P\left\{ \hat{\theta} > M \right\} \to 1.$$

Also, note that by Lemma A2(ii), $h^{-d/2}(c_\alpha^* - a_n)/\sigma_n = O_P(1)$. Hence

$$c_\alpha^* = a_n + O_P(h^{d/2}) = O_P(1).$$

Given that $c_\alpha^* = O_P(1)$ and $\hat{a}^* = O_P(1)$ by Lemma A1 and Assumption A6(i), we obtain that $P\{\hat{\theta} > c_{\alpha,\eta}^*\} \to 1$, as $n \to \infty$. ∎

Lemma A5. *Suppose that the conditions of Theorem 4 or Theorem 5 hold. Then as* $n \to \infty$*, the following holds: for any* $c_{n,1}, c_{n,2} > 0$ *such that*

$$\sqrt{\log n}/c_{n,2} \to 0,$$

as $n \to \infty$*. Then*

$$\inf_{P\in\mathcal{P}_n^0(\lambda_n)} P\left\{ \int_{\mathcal{S}\setminus B_n^0(c_{n,1},c_{n,2})} \Lambda_p\left(\hat{\mathbf{u}}_\tau(x)\right) dQ(x,\tau) = 0 \right\} \to 1.$$

Furthermore, we have for any $A \in \mathcal{N}_J$,

$$\inf_{P \in \mathcal{P}_n^0(\lambda_n)} P\left\{ \int_{B_{n,A}^0(c_{n,1},c_{n,2})} \left\{ \Lambda_p\left(\hat{\mathbf{u}}_\tau(x)\right) - \Lambda_{A,p}\left(\hat{\mathbf{u}}_\tau(x)\right) \right\} dQ(x,\tau) = 0 \right\} \to 1.$$

Proof of Lemma A5. Consider the first statement. Let λ be either $d/2$ or $d/4$. We write

$$\begin{aligned} & \int_{\mathcal{S}\setminus B_n^0(c_{n,1},c_{n,2})} \Lambda_p\left(\hat{\mathbf{u}}_\tau(x)\right) dQ(x,\tau) \\ = & \int_{\mathcal{S}\setminus B_n^0(c_{n,1},c_{n,2})} \Lambda_p\left(\hat{\mathbf{s}}_\tau(x) + (\mathbf{u}_\tau(x;\hat{\sigma})\right) dQ(x,\tau). \\ = & \int_{\mathcal{S}\setminus B_n^0(c_{n,1},c_{n,2})} \Lambda_p\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_\tau^0(x;\hat{\sigma}) + h^\lambda \delta_{\tau,\hat{\sigma}}(x)\right) dQ(x,\tau), \end{aligned}$$

where $\mathbf{u}_\tau^0(x;\hat{\sigma}) \equiv (r_{n,1}v_{n,\tau,1}^0(x)/\hat{\sigma}_{\tau,1}(x), \cdots, r_{n,J}v_{n,\tau,J}^0(x)/\hat{\sigma}_{\tau,J}(x))$ and

$$\delta_{\tau,\hat{\sigma}}(x) \equiv \left(\frac{\delta_{\tau,1}(x)}{\hat{\sigma}_{\tau,1}(x)}, \cdots, \frac{\delta_{\tau,J}(x)}{\hat{\sigma}_{\tau,J}(x)} \right). \tag{A.31}$$

Note that $\delta_{\tau,\hat{\sigma}}(x)$ is bounded with probability approaching one by Assumption A3. Also note that for each $j \in \mathbb{N}_J$,

(A.32)

$$\begin{aligned} \sup_{(x,\tau)\in\mathcal{S}} \left| \frac{r_{n,j}\{\hat{v}_{n,\tau,j}(x) - v_{n,\tau,j}^0(x)\}}{\hat{\sigma}_{\tau,j}(x)} \right| &\le \sup_{(x,\tau)\in\mathcal{S}} \left| \frac{r_{n,j}\{\hat{v}_{n,\tau,j}(x) - v_{n,\tau,j}(x)\}}{\hat{\sigma}_{\tau,j}(x)} \right| + h^\lambda \sup_{(x,\tau)\in\mathcal{S}} \left| \frac{\delta_{\tau,j}(x)}{\hat{\sigma}_{\tau,j}(x)} \right| \\ &= O_P\left(\sqrt{\log n} + h^\lambda\right) = O_P\left(\sqrt{\log n}\right), \end{aligned}$$

by Assumption A3. Hence we obtain the desired result, using the same arguments as in the proof of Lemma 1.

Given that we have (A.32), the proof of the second statement can be proceeded in the same way as the proof of the first statement. ∎

Recall the definitions of $\bar{\Lambda}_{x,\tau}(\mathbf{v})$ in (A.11). We define for $\mathbf{v} \in \mathbf{R}^J$, $\bar{\Lambda}_{x,\tau}^0(\mathbf{v})$ to be $\bar{\Lambda}_{x,\tau}(\mathbf{v})$ except that $B_{n,A}(c_{n,1}, c_{n,2})$ is replaced by $B_{n,A}^0(c_{n,1}, c_{n,2})$. Define for $\lambda \in \{0, d/4, d/2\}$,

$$\hat{\theta}_\delta(c_{n,1}, c_{n,2}; \lambda) \equiv \int \bar{\Lambda}_{x,\tau}^0\left(\hat{\mathbf{s}}_\tau(x) + h^\lambda \delta_{\tau,\sigma}(x)\right) dQ(x,\tau). \tag{A.33}$$

Let

$$\begin{aligned} a_{n,\delta}^R(c_{n,1}, c_{n,2}; \lambda) &\equiv \int \mathbf{E}\left[\bar{\Lambda}_{x,\tau}^0\left(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x) + h^\lambda \delta_{\tau,\sigma}(x) \right) \right] dQ(x,\tau), \\ \hat{\theta}_\delta^*(c_{n,1}, c_{n,2}; \lambda) &\equiv \int \bar{\Lambda}_{x,\tau}^0\left(\hat{\mathbf{s}}_\tau^*(x) + h^\lambda \delta_{\tau,\sigma}(x)\right) dQ(x,\tau), \end{aligned}$$

and

$$a_{n,\delta}^{R*}(c_{n,1}, c_{n,2}; \lambda) \equiv \int \mathbf{E}^*\left[\bar{\Lambda}_{x,\tau}^0\left(\sqrt{nh^d}\mathbf{z}_{N,\tau}^*(x) + h^\lambda \delta_{\tau,\sigma}(x) \right) \right] dQ(x,\tau). \tag{A.34}$$

We also define

$$a_{n,\delta}(c_{n,1}, c_{n,2}; \lambda) \equiv \int \mathbf{E}\left[\bar{\Lambda}^0_{x,\tau}(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0) + h^{\lambda}\delta_{\tau,\sigma}(x))\right] dQ(x,\tau).$$

When $c_{n,1} = c_{n,2} = c_n$, we simply write $a^R_{n,\delta}(c_n;\lambda)$, $a^{R*}_{n,\delta}(c_n;\lambda)$, and $a_{n,\delta}(c_n;\lambda)$, instead of writing $a^R_{n,\delta}(c_n, c_n;\lambda)$, $a^{R*}_{n,\delta}(c_n, c_n;\lambda)$, and $a_{n,\delta}(c_n, c_n;\lambda)$.

Lemma A6. *Suppose that the conditions of Assumptions A6(i) and B4 hold. Then for each $P \in \mathcal{P}$ such that the local alternatives in (6.2) hold with $b_{n,j} = r_{n,j}h^{-\lambda}$, $j = 1, \cdots, J$, for some $\lambda \in \{0, d/4, d/2\}$, and for each nonnegative sequences $c_{n,1}, c_{n,2}$,*

$$\begin{aligned} \left|a^R_{n,\delta}(c_{n,1}, c_{n,2};\lambda) - a_{n,\delta}(c_{n,1}, c_{n,2};\lambda)\right| &= o(h^{d/2}), \text{ and} \\ \left|a^{R*}_{n,\delta}(c_{n,1}, c_{n,2};\lambda) - a_{n,\delta}(c_{n,1}, c_{n,2};\lambda)\right| &= o_P(h^{d/2}). \end{aligned}$$

Proof of Lemma A6. The result follows immediately from Lemma D12 in Appendix D. ∎

Lemma A7. *Suppose that the conditions of Theorem 4 are satisfied. Then for each $\lambda \in \{0, d/4, d/2\}$, for each $P \in \mathcal{P}^0_n(\lambda_n)$ such that the local alternatives in (6.2) hold,*

$$\begin{aligned} h^{-d/2}\left(\frac{\bar{\theta}_{n,\delta}(c_{n,U}, c_{n,L};\lambda) - a^R_{n,\delta}(c_{n,U}, c_{n,L};\lambda)}{\sigma_n(c_{n,U}, c_{n,L})}\right) &\xrightarrow{d} N(0,1) \text{ and} \\ h^{-d/2}\left(\frac{\bar{\theta}^*_{n,\delta}(c_{n,U}, c_{n,L};\lambda) - a^{R*}_{n,\delta}(c_{n,U}, c_{n,L};\lambda)}{\sigma_n(c_{n,U}, c_{n,L})}\right) &\xrightarrow{d^*} N(0,1),\ \mathcal{P}^0_n(\lambda_n)\text{-uniformly.} \end{aligned}$$

Proof of Lemma A7. Note that by the definition of $\mathcal{P}^0_n(\lambda_n)$, we have

$$\liminf_{n\to\infty} \inf_{P\in\mathcal{P}^0_n(\lambda_n)} \sigma^2_n(c_{n,U}, c_{n,L}) \geq \frac{\eta}{\Phi^{-1}(1-\alpha)}.$$

Hence we can follow the proof of Lemma A2 to obtain the desired results. ∎

Proof of Theorem 4. Using Lemma A5, we find that

$$\hat{\theta} = \sum_{A\in\mathcal{N}_J} \int_{B^0_{n,A}(c_{n,U}, c_{n,L})} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_\tau(x;\hat{\sigma})\right) dQ(x,\tau)$$

with probability approaching one. We write the leading sum as

$$\sum_{A\in\mathcal{N}_J} \int_{B^0_{n,A}(0)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_\tau(x;\hat{\sigma})\right) dQ(x,\tau) + R_n,$$

where

$$R_n \equiv \sum_{A\in\mathcal{N}_J} \int_{B^0_{n,A}(c_{n,U}, c_{n,L})\backslash B^0_{n,A}(0)} \Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x) + \mathbf{u}_\tau(x;\hat{\sigma})\right) dQ(x,\tau).$$

We write $h^{-d/2}R_n$ as

$$h^{-d/2}\sum_{A\in\mathcal{N}_J}\int_{B^0_{n,A}(c_{n,U},c_{n,L})\backslash B^0_{n,A}(0)}\Lambda_{A,p}\left(\begin{array}{c}\hat{\mathbf{s}}_\tau(x)+\mathbf{u}^0_\tau(x;\hat{\sigma})\\ +h^{d/2}\delta_{\tau,\hat{\sigma}}(x)(1+o(1))\end{array}\right)dQ(x,\tau)$$

$$\leq\; h^{-d/2}\sum_{A\in\mathcal{N}_J}\int_{B^0_{n,A}(c_{n,U},c_{n,L})\backslash B^0_{n,A}(0)}\Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)+h^{d/2}\delta_{\tau,\hat{\sigma}}(x)(1+o(1))\right)dQ(x,\tau),$$

by Assumption C2. We bound the last sum as

$$Ch^{-d/2}\sum_{A\in\mathcal{N}_J}\left(\sup_{(x,\tau)\in\mathcal{S}}||\hat{\mathbf{s}}_\tau(x)||\right)^p Q\left(B^0_{n,A}(c_{n,U},c_{n,L})\backslash B^0_{n,A}(0)\right)=O_P\left(h^{-d/2}(\log n)^{p/2}\lambda_n\right)=o_P(1)$$

using Assumption A3 and the rate condition in (4.10). We conclude that

$$\text{(A.35)}\quad h^{-d/2}\hat{\theta} = h^{-d/2}\sum_{A\in\mathcal{N}_J}\int_{B^0_{n,A}(0)}\Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)+\mathbf{u}_\tau(x;\hat{\sigma})\right)dQ(x,\tau)+o_P(1)$$

$$= h^{-d/2}\sum_{A\in\mathcal{N}_J}\int_{B^0_{n,A}(0)}\Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)+h^{d/2}\delta_{\tau,\hat{\sigma}}(x)\right)dQ(x,\tau)+o_P(1),$$

where the second equality follows by Assumption C2 and by the definition of $B^0_{n,A}(0)$.

Fix small $\kappa>0$ and define

$$\delta^L_{\tau,\sigma,\kappa,j}(x)\equiv\begin{cases}\frac{\delta_{\tau,j}(x)}{(1+\kappa)\sigma_{n,\tau,j}(x)} & \text{if } \delta_{\tau,j}(x)\geq 0\\ \frac{\delta_{\tau,j}(x)}{(1-\kappa)\sigma_{n,\tau,j}(x)} & \text{if } \delta_{\tau,j}(x)<0\end{cases}\quad\text{and}$$

$$\delta^U_{\tau,\sigma,\kappa,j}(x)\equiv\begin{cases}\frac{\delta_{\tau,j}(x)}{(1-\kappa)\sigma_{n,\tau,j}(x)} & \text{if } \delta_{\tau,j}(x)\geq 0\\ \frac{\delta_{\tau,j}(x)}{(1+\kappa)\sigma_{n,\tau,j}(x)} & \text{if } \delta_{\tau,j}(x)<0\end{cases}.$$

Define $\delta^L_{\tau,\sigma,\kappa}(x)$ and $\delta^U_{\tau,\sigma,\kappa}(x)$ to be $\mathbf{R}^J$-valued maps whose j-th entries are given by $\delta^L_{\tau,\sigma,\kappa,j}(x)$ and $\delta^U_{\tau,\sigma,\kappa,j}(x)$ respectively. By construction, Assumptions A3 and C2(ii), we have

$$P\left\{\delta^L_{\tau,\sigma,\kappa}(x)\leq\delta_{\tau,\hat{\sigma}}(x)\leq\delta^U_{\tau,\sigma,\kappa}(x)\right\}\to 1,$$

as $n\to\infty$. Therefore, with probability approaching one,

$$\text{(A.36)}\quad\hat{\theta}_{\delta,L}(0;d/2) \equiv \sum_{A\in\mathcal{N}_J}\int_{B^0_{n,A}(0)}\Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)+h^{d/2}\delta^L_{\tau,\sigma,\kappa}(x)\right)dQ(x,\tau)$$

$$\leq \sum_{A\in\mathcal{N}_J}\int_{B^0_{n,A}(0)}\Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)+h^{d/2}\delta_{\tau,\hat{\sigma}}(x)\right)dQ(x,\tau)$$

$$\leq \sum_{A\in\mathcal{N}_J}\int_{B^0_{n,A}(0)}\Lambda_{A,p}\left(\hat{\mathbf{s}}_\tau(x)+h^{d/2}\delta^U_{\tau,\sigma,\kappa}(x)\right)dQ(x,\tau)\equiv\hat{\theta}_{\delta,U}(0;d/2).$$

We conclude from (A.35) that

$$\text{(A.37)}\quad\hat{\theta}_{\delta,L}(0;d/2)+o_P(h^{d/2})\leq\hat{\theta}\leq\hat{\theta}_{\delta,U}(0;d/2)+o_P(h^{d/2}).$$

As for the bootstrap counterpart, note that since $\delta_{\tau,j}(x)$ is bounded and $\sigma_{n,\tau,j}(x)$ is bounded away from zero uniformly over $(x,\tau) \in \mathcal{S}$ and $n \geq 1$, and hence

$$\sup_{(x,\tau)\in\mathcal{S}} \left| \frac{1}{h^{-d/2}} \frac{\delta_{\tau,j}(x)}{\sigma_{n,\tau,j}(x)} \right| \leq Ch^{d/2} \to 0, \tag{A.38}$$

as $n \to \infty$. By (A.38), the difference between $r_{n,j} v_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x)$ and $r_{n,j} v^0_{n,\tau,j}(x)/\sigma_{n,\tau,j}(x)$ vanishes uniformly over $(x,\tau) \in \mathcal{S}$. Therefore, combining this with Lemma A3, we find that

$$P\left\{ \hat{B}_n(\hat{c}_n) \subset B^0_n(c_{n,U}, c_{n,L}) \right\} \to 1, \tag{A.39}$$

as $n \to \infty$.

Now with probability approaching one,

$$\begin{aligned} \hat{\theta}^* &= \sum_{A\in\mathcal{N}_J} \int_{\hat{B}_A(\hat{c}_n)} \Lambda_{A,p}\left(\hat{\mathbf{s}}^*_\tau(x)\right) dQ(x,\tau) \\ &= \sum_{A\in\mathcal{N}_J} \int_{B^0_{n,A}(0)} \Lambda_{A,p}\left(\hat{\mathbf{s}}^*_\tau(x)\right) dQ(x,\tau) \\ &\quad + \sum_{A\in\mathcal{N}_J} \int_{\hat{B}_A(\hat{c}_n)\backslash B^0_{n,A}(0)} \Lambda_{A,p}\left(\hat{\mathbf{s}}^*_\tau(x)\right) dQ(x,\tau). \end{aligned} \tag{A.40}$$

As for the last sum, it is bounded by

$$\sum_{A\in\mathcal{N}_J} \int_{B^0_{n,A}(c_{n,U},c_{n,L})\backslash B^0_{n,A}(0)} \Lambda_{A,p}\left(\hat{\mathbf{s}}^*_\tau(x)\right) dQ(x,\tau),$$

with probability approaching one by (A.39). The above sum multiplied by $h^{-d/2}$ is bounded by

$$\begin{aligned} & h^{-d/2}\left(\sup_{(x,\tau)\in\mathcal{S}} ||\hat{\mathbf{s}}^*_\tau(x)|| \right)^p \sum_{A\in\mathcal{N}_J} Q\left(B^0_{n,A}(c_{n,U},c_{n,L})\backslash B^0_{n,A}(0)\right) \\ = \ & O_{P^*}\left(h^{-d/2}(\log n)^{p/2}\lambda_n\right) = o_{P^*}(1), \ \mathcal{P}\text{-uniformly}, \end{aligned}$$

by Assumption B2 and the rate condition for λ_n. Thus, we conclude that

$$\hat{\theta}^* = \bar{\theta}^*(0) + o_{P^*}(h^{d/2}), \ \mathcal{P}^0_n(\lambda_n)\text{-uniformly}, \tag{A.41}$$

where

$$\bar{\theta}^*(0) \equiv \sum_{A\in\mathcal{N}_J} \int_{B^0_{n,A}(0)} \Lambda_{A,p}\left(\hat{\mathbf{s}}^*_\tau(x)\right) dQ(x,\tau).$$

Let $\bar{c}^{\alpha*}_n(0)$ be the $(1-\alpha)$-th quantile of the bootstrap distribution of $\bar{\theta}^*(0)$ and let $\gamma^{\alpha*}_n(0)$ be the $(1-\alpha)$-th quantile of the bootstrap distribution of

$$h^{-d/2}\left(\frac{\bar{\theta}^*(0) - a^{R*}_n(0)}{\sigma_n(0)} \right). \tag{A.42}$$

By the definition of $\mathcal{P}_n^0(\lambda_n)$, we have $\sigma_n^2(0) > \eta/\Phi^{-1}(1-\alpha)$. Let $a_{\delta,U}^R(0;d/2)$ and $a_{\delta,L}^R(0;d/2)$ be $a_{n,\delta}^R(0;d/2)$ except that $\delta_{\tau,\sigma}$ is replaced by $\delta_{\tau,\sigma,\kappa}^U$ and $\delta_{\tau,\sigma,\kappa}^L$ respectively. Also, let $a_{\delta,U}(0;d/2)$ and $a_{\delta,L}(0;d/2)$ be $a_{n,\delta}(0;d/2)$ except that $\delta_{\tau,\sigma}$ is replaced by $\delta_{\tau,\sigma,\kappa}^U$ and $\delta_{\tau,\sigma,\kappa}^L$ respectively. We bound $P\{\hat{\theta} > c_{\alpha,\eta}^*\}$ by

$$\begin{aligned} & P\left\{h^{-d/2}\left(\frac{\hat{\theta}_{\delta,U}(0;d/2) - a_{\delta,U}^R(0;d/2)}{\sigma_n(0)}\right) > h^{-d/2}\left(\frac{c_\alpha^* - a_{\delta,U}^R(0;d/2)}{\sigma_n(0)}\right)\right\} + o(1) \\ = \quad & P\left\{h^{-d/2}\left(\frac{\hat{\theta}_{\delta,U}(0;d/2) - a_{\delta,U}^R(0;d/2)}{\sigma_n(0)}\right) > h^{-d/2}\left(\frac{\bar{c}_n^{\alpha *}(0) - a_{\delta,U}^R(0;d/2)}{\sigma_n(0)}\right)\right\} + o(1), \end{aligned}$$

where the equality uses (A.41). Then we observe that

$$\begin{aligned} \frac{\bar{c}_n^{\alpha *}(0) - a_{\delta,U}^R(0;d/2)}{\sigma_n(0)} &= \frac{\bar{c}_n^{\alpha *}(0) - a_n^{R*}(0)}{\sigma_n(0)} + \frac{a_n^{R*}(0) - a_{\delta,U}^R(0;d/2)}{\sigma_n(0)} \\ &= h^{d/2}\gamma_n^{\alpha *}(0) + \frac{a_n^{R*}(0) - a_{\delta,U}^R(0;d/2)}{\sigma_n(0)}. \end{aligned}$$

As for the last term, we use Lemmas A1 and A6 to deduce that

$$\begin{aligned} a_n^{R*}(0) - a_{\delta,U}^R(0;d/2) &= a_n^R(0) - a_{\delta,U}^R(0;d/2) + o_P(h^{d/2}) \\ &= a_n(0) - a_{\delta,U}(0;d/2) + o_P(h^{d/2}). \end{aligned}$$

As for $a_n(0) - a_{\delta,U}(0;d/2)$, we observe that

$$\begin{aligned} \text{(A.43)} \quad & \sigma_n(0)^{-1}h^{-d/2}\left\{\mathbf{E}\left[\Lambda_{A,p}\left(\mathbb{W}_{n,\tau,\tau}^{(1)}(x,0) + h^{d/2}\delta_{\tau,\sigma,\kappa}^U(x)\right)\right] - \mathbf{E}\left[\Lambda_{A,p}(\mathbb{W}_{n,\tau,\tau}^{(1)}(x,0))\right]\right\} \\ = \quad & \sigma_n(0)^{-1}h^{-d/2}\left\{\mathbf{E}\left[\Lambda_{A,p}\left(\mathbb{W}_{n,\tau,\tau}^{(1)}(x,0) + h^{d/2}\delta_{\tau,\sigma,\kappa}^U(x)\right)\right] - \mathbf{E}\left[\Lambda_{A,p}(\mathbb{W}_{n,\tau,\tau}^{(1)}(x,0))\right]\right\} \\ = \quad & \psi_{n,A,\tau}^{(1)}(\mathbf{0};x)^\top \delta_{\tau,\sigma,\kappa}^U(x) + O\left(h^{d/2}\right), \end{aligned}$$

so that

$$\begin{aligned} \frac{h^{-d/2}\left(a_n(0) - a_{\delta,U}(0)\right)}{\sigma_n(0)} &= -\sum_{A\in\mathcal{N}_J}\int \psi_{n,A,\tau}^{(1)}(\mathbf{0};x)^\top \delta_{\tau,\sigma,\kappa}^U(x)dQ(x,\tau) + o(1) \\ &= -\sum_{A\in\mathcal{N}_J}\int \psi_{A,\tau}^{(1)}(\mathbf{0};x)^\top \delta_{\tau,\sigma,\kappa}^U(x)dQ(x,\tau) + o(1), \end{aligned}$$

where the last equality follows by the Dominated Convergence Theorem. On the other hand, by Lemma A7, we have

$$h^{-d/2}\left(\frac{\hat{\theta}_{\delta,U}(0;d/2) - a_{\delta,U}^R(0;d/2)}{\sigma_n(0)}\right) \xrightarrow{d} N(0,1).$$

Since $\gamma_n^{\alpha *}(0) = \gamma_{\alpha,\infty} + o_P(1)$ by Lemma A4, we use this result to deduce that

$$\begin{aligned} & \lim_{n\to\infty} P\left\{ h^{-d/2}\left(\frac{\hat{\theta}_{\delta,U}(0;d/2) - a_{\delta,U}^R(0;d/2)}{\sigma_n(0)}\right) > h^{-d/2}\left(\frac{\bar{c}_n^{\alpha *}(0) - a_{\delta,U}^R(0;d/2)}{\sigma_n(0)}\right)\right\} \\ = & \ 1 - \Phi\left(z_{1-\alpha} - \sum_{A\in\mathcal{N}_J}\int \psi_{A,\tau}^{(1)}(\mathbf{0};x)^\top \delta_{\tau,\sigma,\kappa}^U(x) dQ(x,\tau)\right). \end{aligned}$$

Similarly, we also use (A.37) to bound $P\left\{\hat{\theta} > c_{\alpha,\eta}^*\right\}$ from below by

$$P\left\{ h^{-d/2}\left(\frac{\hat{\theta}_{\delta,L}(0;d/2) - a_{\delta,L}^R(0;d/2)}{\sigma_n(0)}\right) > h^{-d/2}\left(\frac{\bar{c}_n^{\alpha *}(0) - a_{\delta,L}^R(0;d/2)}{\sigma_n(0)}\right)\right\} + o(1),$$

and using similar arguments as before, we obtain that

$$\begin{aligned} & \lim_{n\to\infty} P\left\{ h^{-d/2}\left(\frac{\hat{\theta}_{\delta,L}(0;d/2) - a_{\delta,L}^R(0;d/2)}{\sigma_n(0)}\right) > h^{-d/2}\left(\frac{\bar{c}_n^{\alpha *}(0) - a_{\delta,L}^R(0;d/2)}{\sigma_n(0)}\right)\right\} \\ = & \ 1 - \Phi\left(z_{1-\alpha} - \sum_{A\in\mathcal{N}_J}\int \psi_{A,\tau}^{(1)}(\mathbf{0};x)^\top \delta_{\tau,\sigma,\kappa}^L(x) dQ(x,\tau)\right). \end{aligned}$$

We conclude from this and (A.36) that for any small $\kappa > 0$,

$$\begin{aligned} & 1 - \Phi\left(z_{1-\alpha} - \sum_{A\in\mathcal{N}_J}\int \psi_{A,\tau}^{(1)}(\mathbf{0};x)^\top \delta_{\tau,\sigma,\kappa}^L(x) dQ(x,\tau)\right) + o(1) \\ \leq & \ P\left\{\hat{\theta} > c_{\alpha,\eta}^*\right\} \leq 1 - \Phi\left(z_{1-\alpha} - \sum_{A\in\mathcal{N}_J}\int \psi_{A,\tau}^{(1)}(\mathbf{0};x)^\top \delta_{\tau,\sigma,\kappa}^U(x) dQ(x,\tau)\right) + o(1). \end{aligned}$$

Note that $\psi_{A,\tau}^{(1)}(\mathbf{0};x)^\top \delta_{\tau,\sigma,\kappa}^U(x)$ and $\psi_{A,\tau}^{(1)}(\mathbf{0};x)^\top \delta_{\tau,\sigma,\kappa}^L(x)$ are bounded maps in (x,τ) by the assumption of the theorem, and that

$$\lim_{\kappa\to 0} \delta_{\tau,\sigma,\kappa}^L(x) = \lim_{\kappa\to 0} \delta_{\tau,\sigma,\kappa}^U(x) = \delta_{\tau,\sigma}(x),$$

for each $(x,\tau) \in \mathcal{S}$. Hence by sending $\kappa \to 0$ and applying the Dominated Convergence Theorem to both the bounds above, we obtain the desired result. ∎

Proof of Theorem 5. First, observe that Lemma A5 continues to hold. This can be seen by following the proof of Lemma A5 and noting that (A.32) becomes here

$$\sup_{(x,\tau)\in\mathcal{S}} \left| \frac{r_{n,j}\{\hat{v}_{n,\tau,j}(x) - v_{n,\tau,j}^0(x)\}}{\hat{\sigma}_{\tau,j}(x)} \right| = O_P\left(\sqrt{\log n} + h^{d/4}\right) = O_P\left(\sqrt{\log n}\right),$$

yielding the same convergence rate. The rest of the proof is the same. Similarly, Lemma A6 continues to hold also under the modified local alternatives of (6.2) with $b_{n,j} = r_{n,j}h^{-d/4}$.

We define

$$\tilde{\delta}_{\tau,\sigma}(x) \equiv h^{-d/4}\delta_{\tau,\sigma}(x). \tag{A.44}$$

We follow the proof of Theorem 4 and take up arguments from (A.43). Observe that

$$\begin{aligned} & \sigma_n(0)^{-1}h^{-d/2}\left\{\mathbf{E}\left[\Lambda_{A,p}\left(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0)+h^{d/2}\tilde{\delta}_{\tau,\sigma}(x)\right)\right]-\mathbf{E}\left[\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0))\right]\right\} \\ = \; & \sigma_n(0)^{-1}h^{-d/2}\left\{\mathbf{E}\left[\Lambda_{A,p}\left(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0)+h^{d/2}\tilde{\delta}_{\tau,\sigma}(x)\right)\right]-\mathbf{E}\left[\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0))\right]\right\} \\ = \; & \psi^{(1)}_{n,A,\tau}(\mathbf{0};x)^\top\tilde{\delta}_{\tau,\sigma}(x)+h^{d/2}\tilde{\delta}_{\tau,\sigma}(x)^\top\psi^{(2)}_{n,A,\tau}(\mathbf{0};x)\tilde{\delta}_{\tau,\sigma}(x)/2. \end{aligned}$$

By the Dominated Convergence Theorem,

$$\begin{aligned} \int \psi^{(1)}_{n,A,\tau}(\mathbf{0};x)^\top\tilde{\delta}_{\tau,\sigma}(x)dQ(x,\tau) &= \int \psi^{(1)}_{A,\tau}(\mathbf{0};x)^\top\tilde{\delta}_{\tau,\sigma}(x)dQ(x,\tau)+o(1) \text{ and} \\ \int \psi^{(2)}_{n,A,\tau}(\mathbf{0};x)^\top\tilde{\delta}_{\tau,\sigma}(x)dQ(x,\tau) &= \int \psi^{(2)}_{A,\tau}(\mathbf{0};x)^\top\tilde{\delta}_{\tau,\sigma}(x)dQ(x,\tau)+o(1). \end{aligned}$$

Since $\sum_{A\in\mathcal{N}_J}\int \psi^{(1)}_{A,\tau}(\mathbf{0};x)^\top\tilde{\delta}_{\tau,\sigma}(x)dQ(x,\tau)=0$, by the condition for $\delta_{\tau,\sigma}(x)$ in the theorem,

$$\begin{aligned} & \sum_{A\in\mathcal{N}_J}\int h^{-d/2}\left\{\begin{array}{c}\mathbf{E}\left[\Lambda_{A,p}\left(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0)+h^{d/2}\tilde{\delta}_{\tau,\sigma}(x)\right)\right] \\ -\mathbf{E}\left[\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0))\right]\end{array}\right\}dQ(x,\tau) \\ = \; & \frac{1}{2}\sum_{A\in\mathcal{N}_J}\int \delta_{\tau,\sigma}(x)^\top\psi^{(2)}_{A,\tau}(\mathbf{0};x)\delta_{\tau,\sigma}(x)dQ(x,\tau)+o(1). \end{aligned}$$

Now we can use the above result by replacing $\delta_{\tau,\sigma}(x)$ by $\delta^U_{\tau,\sigma,\kappa}(x)$ and $\delta^L_{\tau,\sigma,\kappa}(x)$ and follow the proof of Theorem 4 to obtain the desired result. ∎

Appendix B. Proofs of Results for the Example in Section 5

We first offer a general asymptotic linear representation theorem for quantile regression functions that can be useful for other purposes. While the proof employs some arguments from Guerre and Sabbah (2012), the result is different from theirs. The main difference is that their result pays attention to uniformity in h over some range, while our result pays attention to uniformity in P.

Let $(B^\top, X^\top, L)^\top$, with $B \equiv (B_1,\cdots,B_{\bar{L}})^\top \in \mathbf{R}^{\bar{L}}$, and $X \in \mathbf{R}^d$, be a random vector such that the joint distribution of $(B^\top, X^\top)^\top$ is absolutely continuous with respect to Lebesgue measure and L is a discrete random variable taking values from $\mathbb{N}_L \equiv \{1,2,\cdots,\bar{L}\}$. For each $x \in \mathbf{R}^d$ and $k \in \mathbb{N}_L$, the conditional distribution of B_l given $(X,L)=(x,k)$ is the same across $l=1,\cdots,k$.

Let $q_k(\tau|x)$ denote the τ-th quantile of B_l conditional on $X=x$ and $L=k$, where $\tau\in(0,1)$. That is, $P\{B_l \le q_k(\tau|x)|X=x, L=k\}=\tau$ for all x in the support of X and all

$k \in \{1, \cdots, \bar{L}\}$. We write

$$B_l = q_k(\tau|X) + \varepsilon_{\tau,lk},\ \tau \in (0,1),\ \text{for all } k \in \{1, \cdots, \bar{L}\},$$

where $\varepsilon_{\tau,lk}$ is a continuous random variable such that the τ-th conditional quantile of $\varepsilon_{\tau,lk}$ given X and $L = k$ is equal to zero. Note that $q_k(\tau|x)$ is the same across $k = 1, \cdots, \bar{L}$, by our assumption.

Suppose that we are given a random sample $\{(B_i^\top, X_i^\top, L_i)^\top\}_{i=1}^n$ of $(B^\top, X^\top, L)^\top$. We use a local polynomial method, similar to Chaudhuri (1991a) and Chaudhuri (1991b). Assume that $q_k(\tau|x)$ is $(r+1)$-times continuously differentiable with respect to x, where $r \geq 1$. Then, we construct an estimator $\hat{\gamma}_{\tau,k}(x)$ as follows:

$$\hat{\gamma}_{\tau,k}(x) \equiv \text{argmin}_{\gamma \in \mathbf{R}^{|A_r|}} \sum_{i=1}^n 1\{L_i = k\} \sum_{l=1}^k l_\tau\left(B_{li} - \gamma^\top c(X_i - x)\right) K_h(X_i - x),$$

where $l_\tau(u) \equiv u[\tau - 1\{u \leq 0\}]$ for any $u \in \mathbf{R}$, $K_h(t) = K(t/h)/h^d$, K is a d-variate kernel function, and h is a bandwidth that goes to zero as $n \to \infty$.

We make the following assumptions.

Assumption QR1. *(i) There exists an integer $r \geq 1$ such that for all $(\tau, k) \in \mathcal{T} \times \mathbb{N}_L$, $q_k(\tau|\cdot)$ is $r+1$ times continuously differentiable on $\mathcal{S}_\tau(\varepsilon)$ with derivatives bounded uniformly over $(\tau, P) \in \mathcal{T} \times \mathcal{P}$.*
(ii) The density f of X is continuously differentiable on $\mathbf{R}$ with a derivative bounded uniformly over $P \in \mathcal{P}$.

Assumption QR2. *For each $k \in \mathbb{N}_L$, (i) $\inf_{x \in \mathcal{S}_\tau(\varepsilon)} f_{\tau,k}(0|x)$ is bounded away from zero uniformly over $(\tau, P) \in \mathcal{T} \times \mathcal{P}$, with $f_{\tau,k}(0|x)$ being the conditional density of $B_{li} - q_k(\tau|X_i)$ given $X_i = x$ and $L_i = k$. (ii) $\sup_{x \in \mathcal{S}_\tau(\varepsilon)} f_{\tau,k}(0|x)$ is bounded uniformly over $(\tau, P) \in \mathcal{T} \times \mathcal{P}$, and (iii) $f_{\tau,k}(\bar{\varepsilon}|x)$ is continuously differentiable in $(\bar{\varepsilon}, x)$ with a derivative bounded uniformly over $x \in \mathcal{S}_\tau(\varepsilon)$, $\tau \in \mathcal{T}$, and $P \in \mathcal{P}$. (iv) $P\{L_i = k|X_i = x\}$ is bounded away from zero uniformly over $x \in \mathcal{S}_\tau(\varepsilon)$, $\tau \in \mathcal{T}$, and $P \in \mathcal{P}$, and continuously differentiable in x with a derivative bounded uniformly over $x \in \mathcal{S}_\tau(\varepsilon)$, $\tau \in \mathcal{T}$, and $P \in \mathcal{P}$.*

Assumption QR3. *(i) K is compact-supported, nonnegative, bounded, and Lipschitz continuous on the interior of its support, $\int K(u)du = 1$, and $\int K(u)\,||u||^2 du > 0$. (ii) As $n \to \infty$, $n^{-1/2}h^{-d/2}\log n + \sqrt{n}h^{r+1} \to 0$.*

We define

$$\begin{aligned} \Delta_{x,\tau,lk,i} &\equiv B_{li} - \gamma_{\tau,k}^\top(x) c(X_i - x), \\ c_{h,x,i} &\equiv c\left((X_i - x)/h\right),\ \text{and } K_{h,x,i} \equiv K\left((X_i - x)/h\right), \end{aligned}$$

where we recall $c(z) = (z^u)_{u\in A_r}$, for $z \in \mathbf{R}^d$, and $\gamma_{\tau,k}(x) = (\gamma_{\tau,k,u}(x))_{u\in A_r}$ with

$$\gamma_{\tau,k,u}(x) = \frac{1}{u_1!\cdots u_d!} D^u q_k(\tau|x).$$

We also define for $a, b \in \mathbf{R}^{|A_r|}$,

$$\begin{aligned} \zeta_{n,x,\tau,k}(a,b) &\equiv \sum_{i=1}^{n} 1\{L_i = k\} \sum_{l=1}^{L_i} \left\{ \begin{array}{c} l_\tau\left(\Delta_{x,\tau,lk,i} - (a+b)^\top c_{h,x,i}/\sqrt{nh^d}\right) \\ -l_\tau\left(\Delta_{x,\tau,lk,i} - a^\top c_{h,x,i}/\sqrt{nh^d}\right) \end{array} \right\} K_{h,x,i}, \\ \psi_{n,x,\tau,k} &\equiv -\frac{1}{\sqrt{nh^d}} \sum_{i=1}^{n} 1\{L_i = k\} \sum_{l=1}^{L_i} \tilde{l}_\tau\left(\Delta_{x,\tau,lk,i}\right) c_{h,x,i} K_{h,x,i}, \end{aligned}$$

where we recall $\tilde{l}_\tau(x) \equiv \tau - 1\{x \le 0\}$. Define $\zeta^\Delta_{n,x,\tau,k}(a,b) \equiv \zeta_{n,x,\tau,k}(a,b) - b^\top \psi_{n,x,\tau,k}$.

Lemma QR1. *Suppose that Assumptions QR1-QR3 hold. Let $\{\delta_{1n}\}_{n=1}^\infty$ and $\{\delta_{2n}\}_{n=1}^\infty$ be positive sequences such that $\delta_{1n} = O_P(1)$ and $\delta_{2n} \le \delta_{1n}$ from some large n on. Then for each $k \in \mathbb{N}_L$, the following holds uniformly over $P \in \mathcal{P}$:*
(i)

$$\begin{aligned} &\mathbf{E}\left[\sup_{a,b:||a||\le\delta_{1n},||b||\le\delta_{2n}} \sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)} |\zeta^\Delta_{n,x,\tau,k}(a,b) - \mathbf{E}[\zeta^\Delta_{n,x,\tau,k}(a,b)]|\right] \\ &= O\left(\frac{\delta_{2n}\sqrt{\log n}}{n^{1/4}h^{d/4}}\right). \end{aligned}$$

(ii)

$$\mathbf{E}\left[\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)} \|\psi_{n,x,\tau,k}\|\right] = O\left(\sqrt{\log n}\right).$$

(iii)

$$\begin{aligned} &\sup_{a,b:||a||\le\delta_{1n},||b||\le\delta_{2n}} \sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)} \left|\mathbf{E}[\zeta^\Delta_{n,x,\tau,k}(a,b)] - \frac{b^\top M_{n,\tau,k}(x)(b+2a)}{2}\right| \\ &= O\left(\frac{\delta_{2n}\delta_{1n}^2}{n^{1/2}h^{d/2}} + \delta_{2n}\delta_{1n}h^{r+1}\right), \end{aligned}$$

where we recall the definition of $M_{n,\tau,k}(x)$ as

$$M_{n,\tau,k}(x) \equiv k\int P\{L_i = k|X_i = x+th\} f_{\tau,k}(0|x+th) f(x+th) K(t) c(t) c^\top(t) dt.$$

Proof of Lemma QR1. (i) Define

$$\delta_{n,\tau,k}(x_1; x) \equiv \{q_k(\tau|x_1) - \gamma_{\tau,k}(x_1)^\top c(x_1 - x)\} 1\{|x_1 - x| \le h\}, \tag{B.1}$$

where the dependence on P is through $q_k(\tau|x_1)$ and $\gamma_{\tau,k}(x_1)$. We also let

$$\delta_{n,\tau,k}(x_1) \equiv \sup_{\tau\in\mathcal{T},\ x\in\mathcal{S}_\tau(\varepsilon)} \sup_{P\in\mathcal{P}} |\delta_{n,\tau,k}(x_1;x)|. \tag{B.2}$$

It is not hard to see that

$$\sup\nolimits_{\tau\in\mathcal{T},\ x_1\in\mathcal{S}_\tau(\varepsilon)} |\delta_{n,\tau,k}(x_1)| = O(h^{r+1}), \tag{B.3}$$

because $q_k(\tau|x_1) - \gamma_{\tau,k}(x_1)^\top c(x_1 - x)$ is a residual from the Taylor expansion of $q_k(\tau|x_1)$ and $\mathcal{X}$ is bounded, and the derivatives from the Taylor expansion are bounded uniformly over $P \in \mathcal{P}$.

Let $f^{\Delta}_{\tau,k,x}(t|x')$ be the conditional density of $\Delta_{x,\tau,lk,i}$ given $X_i = x'$. For all $x' \in \mathbf{R}^d$ such that $|x - x'| \leq h$, we have

(B.4)

$$\begin{aligned} f^{\Delta}_{\tau,k,x}(t|x') &= \frac{\partial}{\partial t} P\left\{\Delta_{x,\tau,lk,i} \leq t | X_i = x'\right\} \\ &= \frac{\partial}{\partial t} P\left\{B_{li} - q_k(\tau|X_i) \leq t - \delta_{n,\tau,k}(x';x) | X_i = x'\right\} = f_{\tau,k}(t - \delta_{n,\tau,k}(x';x)|x'). \end{aligned}$$

Since $f_{\tau,k}(\cdot|x')$ is bounded uniformly over $x' \in \mathcal{S}_\tau(\varepsilon)$ and over $\tau \in \mathcal{T}$ (Assumption QR2(iii)), we conclude that for some $C > 0$ that does not depend on $P \in \mathcal{P}$,

$$\sup_{\tau\in\mathcal{T}} \sup_{x',x\in\mathcal{S}_\tau(\varepsilon)} f^{\Delta}_{\tau,k,x}(t|x') < C. \tag{B.5}$$

We will use the results in (B.3) and (B.5) later.

Following the identity in Knight (1998, see the proof of Theorem 1), we write

$$l_\tau(x - y) - l_\tau(x) = -y \cdot \bar{l}_\tau(x) + \mu(x, y),$$

where $\mu(x, y) \equiv y \int_0^1 \{1\{x \leq ys\} - 1\{x \leq 0\}\} ds$ and

$$\bar{l}_\tau(x) \equiv \tau - 1\{x \leq 0\} + (1/2) \cdot 1\{x = 0\}.$$

Write $\zeta^{\Delta}_{n,x,\tau,k}(a, b) - \mathbf{E}[\zeta^{\Delta}_{n,x,\tau,k}(a, b)]$ as

$$\sum_{i=1}^n \left\{G_{n,x,\tau,k}(S_i; a, b) - \mathbf{E}\left[G_{n,x,\tau,k}(S_i; a, b)\right]\right\},$$

where $S_i \equiv (Y_i^\top, X_i^\top, L_i)^\top$, $Y_i = (Y_{1,i}, \cdots, Y_{\bar{L},i})^\top$, and

$$G_{n,x,\tau}(S_i; a, b) \equiv \int_0^1 g_{n,x,\tau,k}(S_i; s, b, a) ds \tag{B.6}$$

and $g_{n,x,\tau,k}(S_i; s, b, a)$ is defined to be

$$1\{L_i = k\} \sum_{l=1}^{k} \left(\begin{array}{c} 1\left\{\Delta_{x,\tau,lk,i} - a^\top c_{h,x,i}/\sqrt{nh^d} \le (sb)^\top c_{h,x,i}/\sqrt{nh^d}\right\} \\ -1\left\{\Delta_{x,\tau,lk,i} - a^\top c_{h,x,i}/\sqrt{nh^d} \le 0\right\} \end{array} \right) \frac{b^\top c_{h,x,i} K_{h,x,i}}{\sqrt{nh^d}}.$$

Let $\mathcal{G}_n \equiv \{G_{n,x,\tau,k}(\cdot; a, b) : (a, b, x) \in [-\delta_{1n}, \delta_{1n}]^{r+1} \times [-\delta_{2n}, \delta_{2n}]^{r+1} \times \mathcal{S}_\tau(\varepsilon), \tau \in \mathcal{T}\}$,

$$\begin{aligned} \mathcal{G}_{1n} &\equiv \{\lambda_{\tau,1n}(\cdot; a, x) : (a, x) \in [-\delta_{1n}, \delta_{1n}]^{r+1} \times \mathcal{S}_\tau(\varepsilon), \tau \in \mathcal{T}\} \\ \mathcal{G}_{2n} &\equiv \{b^\top \lambda_{\tau,2n}(\cdot; x) : (b, x) \in [-\delta_{2n}, \delta_{2n}]^{r+1} \times \mathcal{S}_\tau(\varepsilon), \tau \in \mathcal{T}\} \text{ and} \\ \mathcal{G}_{3n} &\equiv \{\lambda_{\tau,3n}(\cdot; x) : x \in \mathcal{S}_\tau(\varepsilon), \tau \in \mathcal{T}\}, \end{aligned}$$

where

$$\begin{aligned} \lambda_{\tau,1n}(S_i; a, x) &\equiv (\Delta_{x,\tau,lk,i} - a^\top \lambda_{\tau,2n}(S_i; x))_{l=1}^{\bar{L}} \\ \lambda_{\tau,2n}(S_i; x) &\equiv c_{h,x,i}/\sqrt{nh^d} \text{ and } \lambda_{\tau,3n}(S_i; x) \equiv K_{h,x,i}. \end{aligned}$$

First, we compute the entropy bound for $\mathcal{G}_n$. We focus on $\mathcal{G}_{1n}$ first. There exists $C > 0$ that does not depend on $P \in \mathcal{P}$, such that for any $\tau \in \mathcal{T}$, any (a, x) and (a', x') in $[-\delta_{1n}, \delta_{1n}]^{r+1} \times \mathcal{S}_\tau(\varepsilon)$, and any $\tau, \tau' \in \mathcal{T}$,

$$|\lambda_{\tau,1n}(S_i; a, x) - \lambda_{\tau',1n}(S_i; a', x')| \le \frac{C}{n^{1/2} h^{r+d/2}} \{||a - a'|| + |\tau - \tau'| + ||x - x'||\}.$$

Since $[-\delta_{1n}, \delta_{1n}]^{r+1} \times \mathcal{S}_\tau(\varepsilon)$ is bounded in the Euclidean space uniformly in $\tau \in \mathcal{T}$, there is $C > 0$ such that for all $\varepsilon \in (0, 1]$,

$$\log N(\varepsilon, \mathcal{G}_{1n}, ||\cdot||_\infty) \le -C \log(\varepsilon \min\{n^{-1/2} h^{-r-d/2}, 1\}),$$

where $||\cdot||_\infty$ denotes the usual supremum norm. Applying similar arguments to $\mathcal{G}_{2n}$ and $\mathcal{G}_{3n}$, we conclude that

$$\log N(\varepsilon, \mathcal{G}_{mn}, ||\cdot||_\infty) \le C - C \log(\varepsilon/n),\ m = 1, 2, 3, \tag{B.7}$$

for some $C > 0$.

Define for $x \in \mathbf{R}$, $\delta > 0$,

$$\begin{aligned} 1_\delta^L(x) &\equiv (1 - \min\{x/\delta, 1\})\, 1\{0 < x\} + 1\{x \le 0\} \text{ and} \\ 1_\delta^U(x) &\equiv (1 - \min\{(x/\delta) + 1, 1\})\, 1\{0 < x + \delta\} + 1\{x + \delta \le 0\}. \end{aligned}$$

We also define for $x, y, z \in \mathbf{R}$,

$$\begin{aligned}\mu(x,y,z) &\equiv zy\int_0^1 \{1\{x \le ys\} - 1\{x \le 0\}\}ds,\\ \mu_\delta^U(x,y,z) &\equiv zy\int_0^1 \{1\{x \le ys\} - 1_\delta^U(x)\}ds, \text{ and}\\ \mu_\delta^L(x,y,z) &\equiv zy\int_0^1 \{1\{x \le ys\} - 1_\delta^L(x)\}ds.\end{aligned}$$

Then observe that

$$\begin{aligned}\mu_\delta^L(x,y,z) &\le \mu(x,y,z) \le \mu_\delta^U(x,y,z) \quad \text{(B.8)}\\ \left|\mu_\delta^U(x,y,z) - \mu(x,y,z)\right| &\le |zy|1\{|x| < \delta\}\\ \left|\mu_\delta^L(x,y,z) - \mu(x,y,z)\right| &\le |zy|1\{|x| < \delta\}\\ \left|\mu_\delta^U(x,y,z) - \mu_\delta^U(x',y',z')\right| &\le C\{|y-y'| + |z-z'| + |x-x'|/\delta\}, \text{ and}\\ \left|\mu_\delta^L(x,y,z) - \mu_\delta^L(x',y',z')\right| &\le C\{|y-y'| + |z-z'| + |x-x'|/\delta\},\end{aligned}$$

for any $y, y', x, x', z, z' \in \mathbf{R}$. Define

$$\begin{aligned}\mathcal{G}_{n,\delta}^U &\equiv \left\{\mu_\delta^U(g_1(S_i), g_2(S_i), g_3(S_i)) : g_m \in \mathcal{G}_{mn},\ m = 1,2,3\right\}, \text{ and}\\ \mathcal{G}_{n,\delta}^L &\equiv \left\{\mu_\delta^L(g_1(S_i), g_2(S_i), g_3(S_i)) : g_m \in \mathcal{G}_{mn},\ m = 1,2,3\right\}.\end{aligned}$$

From (B.8) and (B.7), we find that there exists $C > 0$ such that for each $\delta > 0$ and $\varepsilon > 0$,

$$\begin{aligned}\log N_{[]}(C\varepsilon, \mathcal{G}_{n,\delta}^U, L_p(P)) &\le C - C\log(\varepsilon\delta/n) \text{ and} \quad \text{(B.9)}\\ \log N_{[]}(C\varepsilon, \mathcal{G}_{n,\delta}^L, L_p(P)) &\le C - C\log(\varepsilon\delta/n).\end{aligned}$$

We fix $\varepsilon > 0$, set $\delta = \varepsilon$, and take brackets $[g_{1,L}^{(\varepsilon)}, g_{1,U}^{(\varepsilon)}], \cdots, [g_{N,L}^{(\varepsilon)}, g_{N,U}^{(\varepsilon)}]$ and $[\tilde{g}_{1,L}^{(\varepsilon)}, \tilde{g}_{1,U}^{(\varepsilon)}], \cdots, [\tilde{g}_{N,L}^{(\varepsilon)}, \tilde{g}_{N,U}^{(\varepsilon)}]$ such that

$$\begin{aligned}\mathbf{E}\left(|g_{s,U}^{(\varepsilon)}(S_i) - g_{s,L}^{(\varepsilon)}(S_i)|^2\right) &\le \varepsilon^2 \text{ and} \quad \text{(B.10)}\\ \mathbf{E}\left(|\tilde{g}_{s,U}^{(\varepsilon)}(S_i) - \tilde{g}_{s,L}^{(\varepsilon)}(S_i)|^2\right) &\le \varepsilon^2,\end{aligned}$$

and for any $g \in \mathcal{G}_n^U$ and $\tilde{g} \in \mathcal{G}_n^L$, there exists $s \in \{1, \cdots, N\}$ such that $g_{s,L}^{(\varepsilon)} \le g \le g_{s,U}^{(\varepsilon)}$ and $\tilde{g}_{s,L}^{(\varepsilon)} \le \tilde{g} \le \tilde{g}_{s,U}^{(\varepsilon)}$. Without loss of generality, we assume that $g_{s,L}^{(\varepsilon)}, g_{s,U}^{(\varepsilon)} \in \mathcal{G}_n^U$ and $\tilde{g}_{s,L}^{(\varepsilon)}, \tilde{g}_{s,U}^{(\varepsilon)} \in \mathcal{G}_n^L$. By the first inequality in (B.8), we find that the brackets $[\tilde{g}_{s,L}^{(\varepsilon)}, g_{s,U}^{(\varepsilon)}],\ k = 1, \cdots, N$, cover $\mathcal{G}_n$. Hence by putting $\delta = \varepsilon$ in (B.9) and redefining constants, we conclude that for some $C > 0$

$$\log N_{[]}(C\varepsilon, \mathcal{G}_n, L_p(P)) \le C - C\log(\varepsilon/n), \quad \text{(B.11)}$$

for all $\varepsilon > 0$.

Now, observe that

$$\sup_{b:||b||\leq\delta_{2n},\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\left|\frac{b^\top c_{h,x,i}K_{h,x,i}}{\sqrt{nh^d}}\right|\leq\frac{\bar{c}||K||_\infty\delta_{2n}}{\sqrt{nh^d}}, \tag{B.12}$$

where $\bar{c}>0$ is the diameter of the compact support of K.

For any $g\in\mathcal{G}_n/\bar{L}$ and any $m\geq 1$, $\mathbf{E}\left[|g(S_i)|^m|X_i,L_i=k\right]$ is bounded by

$$\left|\frac{b^\top c_{h,x,i}K_{h,x,i}}{\sqrt{nh^d}}\right|^m.$$

Therefore, by (B.12), for some constants $C_1,C_2>0$, it is satisfied that for any $m\geq 2$,

$$\sup_{P\in\mathcal{P}}\mathbf{E}\left[|g(S_i)|^m\right]\leq C_1\left(\frac{\delta_{2n}}{\sqrt{nh^d}}\right)^m\cdot\sup_{P\in\mathcal{P}}P\left\{\max_{s=1,\cdots,d}|X_{is}-x|\leq h/2\right\}\leq C_2 b_n^{m-2}s_n^2,$$

where

$$b_n\equiv\frac{\delta_{2n}}{\sqrt{nh^d}}\text{ and }s_n\equiv\frac{\delta_{2n}}{n^{3/4}h^{d/4}}. \tag{B.13}$$

By (B.8), (B.10), and (B.12), and the definition of b_n and s_n in (B.13), there exist constants $C_1,C_2>0$ such that for all $m\geq 2$,

$$\begin{aligned}\mathbf{E}\left(|g_{s,U}^{(\varepsilon)}(S_i)-\tilde{g}_{s,L}^{(\varepsilon)}(S_i)|^m\right) &= \mathbf{E}\left(|g_{s,U}^{(\varepsilon)}(S_i)-\tilde{g}_{s,L}^{(\varepsilon)}(S_i)|^{m-2}|g_{s,U}^{(\varepsilon)}(S_i)-\tilde{g}_{s,L}^{(\varepsilon)}(S_i)|^2\right)\\ &\leq C_1\cdot b_n^{m-2}\cdot\mathbf{E}\left(|g_{s,U}^{(\varepsilon)}(S_i)-\tilde{g}_{s,L}^{(\varepsilon)}(S_i)|^2\right)\\ &\leq 2C_1\cdot b_n^{m-2}\cdot\mathbf{E}\left(|g_{s,U}^{(\varepsilon)}(S_i)-g_{s,L}^{(\varepsilon)}(S_i)|^2\right)\\ &\quad +2C_1\cdot b_n^{m-2}\cdot\mathbf{E}\left(|g_{s,L}^{(\varepsilon)}(S_i)-\tilde{g}_{s,L}^{(\varepsilon)}(S_i)|^2\right)\\ &\leq 2C_2\cdot b_n^{m-2}\cdot\{\varepsilon^2+b_n^2\varepsilon\}\leq 2C_2\cdot b_n^{m-2}\cdot\varepsilon.\end{aligned}$$

(The term $b_n^2\varepsilon$ is obtained by chaining the second and third inequalities of (B.8) and using the fact that $\delta=\varepsilon$ and the uniform bound in (B.5). The last inequality follows because $b_n\to 0$ as $n\to\infty$.) We define $\bar{\varepsilon}=\varepsilon^{1/2}$ and bound the last term by $C_3 b_n^{m-2}\bar{\varepsilon}^2$, for some $C_3>0$, because $b_n\leq 1$ from some large n on. The entropy bound in (B.11) as a function of $\bar{\varepsilon}$ remains the same except for a different constant $C>0$ there.

Now by Theorem 6.8 of Massart (2007) and (B.11), there exist $C_1,C_2>0$ such that

$$\begin{aligned}&\sup_{P\in\mathcal{P}}\mathbf{E}\left[\sup_{a,b:||a||\leq\delta_{1n},||b||\leq\delta_{2n},\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}|\zeta_{n,x,\tau,k}^{\Delta}(a,b)-\mathbf{E}[\zeta_{n,x,\tau,k}^{\Delta}(a,b)]|\right]\\ \leq\;& C_1\sqrt{n}\int_0^{s_n}\sqrt{n\wedge\left\{-\log\left(\frac{\varepsilon}{n}\right)\right\}}d\varepsilon+C_1(b_n+s_n)\log n\\ \leq\;& C_2 s_n\sqrt{n\log n}+C_2 b_n\log n=O\left(\frac{\delta_{2n}\sqrt{\log n}}{n^{1/4}h^{d/4}}\right),\end{aligned}$$

where the last equality follows by the definitions of b_n and s_n in (B.13) and by Assumption QR3(ii).

(ii) Define $\lambda_{\tau,4n}(S_i;x) \equiv \Delta_{x,\tau,lk,i}$ and $\mathcal{L}_{k,1} \equiv \{\tilde{l}_\tau(\lambda_{\tau,4n}(\cdot;x)) : \tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)\}$, and $\mathcal{L}_{k,2} \equiv \{\lambda_{\tau,2n}(\cdot;x)\lambda_{\tau,3n}(\cdot;x) : \tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)\}$. We write

$$\psi_{n,x,\tau,k} = \{\psi_{n,x,\tau,k} - \mathbf{E}\left[\psi_{n,x,\tau,k}\right]\} + \mathbf{E}\left[\psi_{n,x,\tau,k}\right].$$

The leading term is an empirical process indexed by the functions in $\mathcal{L}_k \equiv \mathcal{L}_{k,1} \cdot \mathcal{L}_{k,2}$. Approximating the indicator function in $\tilde{l}_\tau$ by upper and lower Lipschitz functions and following similar arguments in the proof of (i), we find that

$$\sup_{P \in \mathcal{P}} \log N_{[]}(\varepsilon, \mathcal{L}_k, L_p(P)) \leq C - C\log\varepsilon + C\log n,$$

for some constant $C > 0$. Note that we can take a constant function C as an envelope of $\mathcal{L}_k$. Then we follow the proof of Lemma 2 to obtain that

$$\mathbf{E}\left[\sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} \|\{\psi_{n,x,\tau,k} - \mathbf{E}\left[\psi_{n,x,\tau,k}\right]\}\|\right] = O(\sqrt{\log n}), \text{ uniformly in } P \in \mathcal{P}.$$

By using (B.3) and (B.4), we find that

$$\mathbf{E}\left[\psi_{n,x,\tau,k}\right] = O(h^{r+1}).$$

Since $\sqrt{n}h^{r+1} \to 0$, we obtain the desired result.

(iii) Recall the definition of $g_{n,x,\tau,k}(S_i;s,b,a)$ in the proof of Lemma QR1(i). We write

$$\mathbf{E}[\zeta^{\Delta}_{n,x,\tau,k}(a,b)] = n\int_0^1 \mathbf{E}\left[g_{n,x,\tau,k}(S_i;s,b,a)\right] ds.$$

Using change of variables, we rewrite

$$\int_0^1 \mathbf{E}\left[g_{n,x,\tau,k}(S_i;s,b,a)\right] ds = kP\left\{L_i = k|X_i\right\} \cdot \phi_n(X_i;a,b),$$

where

$$\phi_n(X_i;a,b) = \int_{a^\top c_{h,x,i}/\sqrt{nh^d}}^{(b+a)^\top c_{h,x,i}/\sqrt{nh^d}} \left\{\begin{array}{c} F_{\tau,k}\left(u - \delta_{n,\tau,k}(X_i;x)|X_i\right) \\ -F_{\tau,k}\left(-\delta_{n,\tau,k}(X_i;x)|X_i\right) \end{array}\right\} du \cdot K_{h,x,i}.$$

By expanding the difference, we have

$$\phi_n(X_i;a,b) = \int_{a^\top c_{h,x,i}/\sqrt{nh^d}}^{(b+a)^\top c_{h,x,i}/\sqrt{nh^d}} udu \cdot f_{\tau,k}\left(-\delta_{n,\tau,k}(X_i;x)|X_i\right) \cdot K_{h,x,i} + R_{n,x,i}(a,b),$$

where $R_{n,x,i}(a,b)$ denotes the remainder term in the expansion. As for the leading integral,

$$\int_{a^\top c_{h,x,i}/\sqrt{nh^d}}^{(b+a)^\top c_{h,x,i}/\sqrt{nh^d}} udu = \frac{1}{2nh^d}\left\{b^\top c_{h,x,i}c_{h,x,i}^\top(b+2a)\right\}.$$

Hence, for any sequences a_n, b_n, we can write $\mathbf{E}[\zeta^{\Delta}_{n,x,\tau,k}(a_n, b_n)]$ as

$$\begin{aligned}
&\frac{1}{2}b_n^\top h^{-d}\mathbf{E}\left[P\left\{L_i = k|X_i\right\} f_{\tau,k}\left(-\delta_{n,\tau,k}(X_i;x)|X_i\right) c_{h,x,i}c_{h,x,i}^\top \cdot K_{h,x,i}\right](b_n + 2a_n) \\
&+nk\mathbf{E}\left[P\left\{L_i = k|X_i\right\} R_{n,x,i}(a_n, b_n)\right] \\
= \; &\frac{1}{2}b_n^\top M_{n,\tau,k}(x)(b_n + 2a_n) + nk\mathbf{E}\left[P\left\{L_i = k|X_i\right\} R_{n,x,i}(a_n, b_n)\right],
\end{aligned}$$

where the last $O(h^{r+1}a_n b_n)$ term is due to (B.3). We can bound

$$\begin{aligned}
nk\left|\mathbf{E}\left[P\left\{L_i = k|X_i\right\} R_{n,x,i}(a_n, b_n)\right]\right| &\leq C_1 nk\mathbf{E}\left[\int_{a_n^\top c_{h,x,i}/\sqrt{nh^d}}^{(b_n+a_n)^\top c_{h,x,i}/\sqrt{nh^d}} u^2 du \cdot K_{h,x,i}\right] + O(h^{r+1}a_n b_n) \\
&\leq \frac{C_2 b_n a_n^2}{n^{1/2}h^{d/2}} + O(h^{r+1}a_n b_n),
\end{aligned}$$

where $C_1 > 0$ and $C_2 > 0$ are constants that do not depend on n or $P \in \mathcal{P}$. ∎

Lemma QR2. *Suppose that Assumptions QR1-QR3 hold. Then, for each $k \in \mathbb{N}_L$,*

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\left\|\sqrt{nh^d}H(\hat{\gamma}_{\tau,k}(x) - \gamma_{\tau,k}(x)) - M_{n,\tau,k}^{-1}(x)\psi_{n,x,\tau,k}\right\| \\
= \; &O_P\left(\frac{\log^{1/2} n}{n^{1/4}h^{d/4}}\right), \; \mathcal{P}\text{-uniformly.}
\end{aligned}$$

Proof of Lemma QR2. (i) Let

(B.14)
$$\begin{aligned}
\tilde{u}_{n,x,\tau} &\equiv -M_{n,\tau,k}^{-1}(x)\psi_{n,x,\tau,k}, \\
\tilde{\psi}_{n,x,\tau,k}(b) &\equiv b^\top\psi_{n,x,\tau,k} + b^\top M_{n,\tau,k}(x)b/2, \text{ and} \\
\tilde{\psi}_{n,x,\tau,k}(a,b) &\equiv \tilde{\psi}_{n,x,\tau,k}(a+b) - \tilde{\psi}_{n,x,\tau,k}(a).
\end{aligned}$$

For any $a \in \mathbf{R}^{|A_r|}$, we can write

(B.15)
$$\begin{aligned}
\tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, a - \tilde{u}_{n,x,\tau}) &= \tilde{\psi}_{n,x,\tau,k}(a) - \tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}) \\
&= \left(a - \tilde{u}_{n,x,\tau}\right)^\top M_{n,\tau,k}(x)\left(a - \tilde{u}_{n,x,\tau}\right)/2 \\
&\geq C_1||a - \tilde{u}_{n,x,\tau}||^2,
\end{aligned}$$

where $C_1 > 0$ is a constant that does not depend on $\tau \in \mathcal{T}$, $x \in \mathcal{S}_\tau(\varepsilon)$ or $P \in \mathcal{P}$. The last inequality uses Assumption QR2(ii) and the fact that K is a nonnegative map that is not constant at zero.

Let

$$\hat{u}_{n,x,\tau} \equiv \sqrt{nh^d}H(\hat{\gamma}_{\tau,k}(x) - \gamma_{\tau,k}(x)),$$

where $x \in \mathcal{S}_\tau(\varepsilon)$ and $\tau \in \mathcal{T}$. Since $\zeta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b)$ is convex in b, we have for any $0 < \delta \leq l$ and for any $b \in \mathbf{R}^{|A_r|}$ such that $||b|| = 1$,

$$\begin{aligned} (\delta/l)\zeta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, lb) &\geq \zeta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, \delta b) \\ &\geq \tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, \delta b) - \Delta_{n,k}(\delta), \end{aligned} \tag{B.16}$$

where

$$\Delta_{n,k}(\delta) \equiv \sup_{b \in \mathbf{R}^{|A_r|}:||b|| \leq 1} |\zeta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, \delta b) - \tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, \delta b)|.$$

Therefore, if $||\hat{u}_{n,x,\tau} - \tilde{u}_{n,x,\tau}|| \geq \delta$, we replace b by $\hat{u}^\Delta_{n,x,\tau} = (\hat{u}_{n,x,\tau} - \tilde{u}_{n,x,\tau})/||\hat{u}_{n,x,\tau} - \tilde{u}_{n,x,\tau}||$ and l by $||\hat{u}_{n,x,\tau} - \tilde{u}_{n,x,\tau}||$ in (B.14), and use (B.16) to obtain that

$$\begin{aligned} 0 &\geq \zeta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, ||\hat{u}_{n,x,\tau} - \tilde{u}_{n,x,\tau}||\hat{u}^\Delta_{n,x,\tau}) \\ &\geq \zeta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, \delta\hat{u}^\Delta_{n,x,\tau}) \\ &\geq \tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, \delta\hat{u}^\Delta_{n,x,\tau}) - \Delta_{n,k}(\delta) \\ &\geq C_1\delta^2||\hat{u}^\Delta_{n,x,\tau}||^2 - \Delta_{n,k}(\delta) = C_1\delta^2 - \Delta_{n,k}(\delta), \end{aligned} \tag{B.17}$$

for all $\delta \leq ||\hat{u}_{n,x,\tau} - \tilde{u}_{n,x,\tau}||$, where the first inequality follows because $\zeta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, ||\hat{u}_{n,x,\tau} - \tilde{u}_{n,x,\tau}||b)$ is minimized at $b = \hat{u}^\Delta_{n,x,\tau}$ by the definition of local polynomial estimation, the fourth inequality follows from (B.15), and the last equality follows because $||\hat{u}^\Delta_{n,x,\tau}||^2 = 1$.

We take large $M > 0$ and let

$$\delta_{1n} = M\sqrt{\log n} \text{ and } \delta_{2n} = \frac{M\sqrt{\log n}}{n^{1/4}h^{d/4}}. \tag{B.18}$$

If $\delta_{2n} \leq ||\hat{u}_{n,x,\tau} - \tilde{u}_{n,x,\tau}||$, we have

$$C_1\delta^2_{2n} \leq \Delta_{n,k}(\delta_{2n}),$$

from (B.17). We let

$$1_n \equiv 1\left\{\sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} ||\tilde{u}_{n,x,\tau}|| \leq M\delta_{1n}\right\}.$$

Then we write

$$P\left\{\inf_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} ||\hat{u}_{n,x,\tau} - \tilde{u}_{n,x,\tau}||^2 \geq \delta^2_{2n}\right\} \leq P\left\{\Delta_{n,k}(\delta_{2n})1_n \geq \delta^2_{2n}\right\} + \mathbf{E}\left[1 - 1_n\right].$$

Now, we show that the first probability vanishes as $n \to \infty$. For each $b \in \mathbf{R}^{|A_r|}$, using the definition of $\tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b) = \tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau} + b) - \tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau})$, we write

$$\begin{aligned}
\tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b) &= \tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau} + b) - \tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}) \\
&= b^\top \psi_{n,x,\tau,k} + (\tilde{u}_{n,x,\tau} + b)^\top M_{n,x,\tau}(\tilde{u}_{n,x,\tau} + b)/2 - \tilde{u}_{n,x,\tau}^\top M_{n,x,\tau}\tilde{u}_{n,x,\tau}/2 \\
&= b^\top \psi_{n,x,\tau,k} + b^\top M_{n,x,\tau} b/2 + b^\top M_{n,x,\tau}\tilde{u}_{n,x,\tau} \\
&= b^\top M_{n,x,\tau} b/2.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\zeta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b) - \tilde{\psi}_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b) &= \zeta^\Delta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b) - \mathbf{E}\left[\zeta^\Delta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b)\right] \\
&\quad + \mathbf{E}\left[\zeta^\Delta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b)\right] - b^\top M_{n,x,\tau} b/2 + b^\top \psi_{n,x,\tau,k} \\
&= \zeta^\Delta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b) - \mathbf{E}\left[\zeta^\Delta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b)\right] \\
&\quad + \mathbf{E}\left[\zeta^\Delta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b)\right] - b^\top M_{n,x,\tau}(b + 2\tilde{u}_{n,x,\tau})/2.
\end{aligned}$$

By Lemma QR1(i),

$$\begin{aligned}
&\sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} \sup_{b \in \mathbf{R}^{|A_r|}: ||b|| \le \delta_{2n}} \left|\zeta^\Delta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b) - \mathbf{E}\left[\zeta^\Delta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b)\right]\right| \\
&= O_P\left(\frac{\delta_{2n}\sqrt{\log n}}{n^{1/4}h^{d/4}} + \frac{\delta_{2n}\log n}{n^{1/2}h^{d/2}}\right) = O_P\left(\frac{\delta_{2n}\sqrt{\log n}}{n^{1/4}h^{d/4}}\right),
\end{aligned}$$

by the definition in (B.18). And by Lemma QR1(iii),

$$\begin{aligned}
&\sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} \sup_{b \in \mathbf{R}^{|A_r|}: ||b|| \le \delta_{2n}} \left|\mathbf{E}\left[\zeta^\Delta_{n,x,\tau,k}(\tilde{u}_{n,x,\tau}, b)\right] - b^\top M_{n,x,\tau}(b + 2\tilde{u}_{n,x,\tau})/2\right| \\
&= O\left(\frac{\delta_{2n}\log n}{n^{1/2}h^{d/2}} + \delta_{2n}\sqrt{\log n}\, h^{r+1}\right) = O\left(\frac{\delta_{2n}\log n}{n^{1/2}h^{d/2}}\right),
\end{aligned}$$

by the definition in (B.18) and Assumption QR3(ii). Thus we conclude that

$$|\Delta_{n,k}(\delta_{2n})| = O_P\left(\frac{\delta_{2n}\sqrt{\log n}}{n^{1/4}h^{d/4}}\right), \tag{B.19}$$

where the last O_P term is uniform over $P \in \mathcal{P}$. We deduce from (B.19) that

$$\sup_{P \in \mathcal{P}} P\left\{\Delta_{n,k}(\delta_{2n}) 1\left\{\sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} ||\tilde{u}_{n,x,\tau}|| \le \delta_{1n}\right\} \ge \delta_{2n}^2\right\} \to 0 \text{ as } n \to \infty$$

and as $M \uparrow \infty$. The proof is completed because

$$\sup_{P \in \mathcal{P}} P\left\{\sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} ||\tilde{u}_{n,x,\tau}|| > \delta_{1n}\right\} \to 0,$$

as $n \to \infty$ and as $M \uparrow \infty$ by Lemma QR1(ii). Thus, we conclude that

$$||\hat{u}_{n,x,\tau} - \tilde{u}_{n,x,\tau}|| = O_P\left(\frac{\sqrt{\log n}}{n^{1/4}h^{d/4}}\right),$$

uniformly in $P \in \mathcal{P}$. ∎

For $z = (x,\tau) \in \mathcal{Z}$, define $\Delta^*_{x,\tau,lk,i} \equiv Y^*_{l,i} - \gamma^\top_{\tau,k}(x)c(X^*_i - x)$,

$$\begin{aligned}
\zeta^*_{n,x,\tau,k}(a,b) &\equiv \sum_{i=1}^n 1\{L_i = k\}\sum_{l=1}^k \left\{\begin{array}{c} l_\tau\left(\Delta^*_{x,\tau,lk,i} - (a+b)^\top c^*_{h,x,i}/\sqrt{nh^d}\right) \\ -l_\tau\left(\Delta^*_{x,\tau,lk,i} - b^\top c^*_{h,x,i}/\sqrt{nh^d}\right)\end{array}\right\} K_{h,x}\left(X^*_i\right), \\
\psi^*_{n,x,\tau,k} &\equiv -\frac{1}{\sqrt{nh^d}}\sum_{i=1}^n 1\{L_i = k\}\sum_{l=1}^k \tilde{l}_\tau\left(\Delta^*_{x,\tau,lk,i}\right)\cdot c^*_{h,x,i}\cdot K^*_{h,x,i},
\end{aligned}$$

where $c^*_{h,x,i}$ and $K^*_{h,x,i}$ are $c_{h,x,i}$ and $K_{h,x,i}$ except that X_i is replaced by X^*_i. We also define $\zeta^{\Delta*}_{n,x,\tau,k}(a,b) \equiv \zeta^*_{n,x,\tau,k}(a,b) - b^\top\psi^*_{n,x,\tau,k}$. The following lemma is the bootstrap analogue of Lemma QR1.

Lemma QR3. *Suppose that Assumptions QR1-QR3 hold. Let $\{\delta_{1n}\}_{n=1}^\infty$ and $\{\delta_{2n}\}_{n=1}^\infty$ be positive sequences such that $\delta_{1n} = O(1)$ and $\delta_{2n} \le \delta_{1n}$ from some large n on. Then for each $k \in \mathbb{N}_L$, the following holds uniformly over $P \in \mathcal{P}$:*
(i)

$$\begin{aligned}
&\mathbf{E}^*\left[\sup_{a,b:||a||\le\delta_{1n},||b||\le\delta_{2n}}\ \sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}|\zeta^{\Delta*}_{n,x,\tau,k}(a,b) - \mathbf{E}^*[\zeta^{\Delta*}_{n,x,\tau,k}(a,b)]|\right] \\
&= O_P\left(\frac{\delta_{2n}\sqrt{\log n}}{n^{1/4}h^{d/4}}\right).
\end{aligned}$$

(ii)

$$\begin{aligned}
&\sup_{a,b:||a||\le\delta_{1n},||b||\le\delta_{2n}}\ \sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\left|\mathbf{E}^*[\zeta^{\Delta*}_{n,x,\tau,k}(a,b)] - \frac{b^\top M_{n,\tau,k}(x)(b+2a)}{2}\right| \\
&= O_P\left(\frac{\delta_{2n}\sqrt{\log n}}{n^{1/4}h^{d/4}} + \delta_{2n}\delta_{1n}h^{r+1}\right).
\end{aligned}$$

Proof of Lemma QR3. (i) Similarly as in the proof of Lemma QR1(i), we rewrite $\zeta^{\Delta*}_{n,x,\tau,k}(a,b) - \mathbf{E}^*[\zeta^{\Delta*}_{n,x,\tau,k}(a,b)]$ as

$$\sum_{i=1}^n \left\{G_{n,x,\tau,k}(S^*_i;a,b) - \mathbf{E}\left[G_{n,x,\tau,k}(S^*_i;a,b)\right]\right\},$$

where $S^*_i = (Y^{*\top}_i, X^{*\top}_i)^\top$. Let $\pi = (x,\tau,s,a,b)$ and $\Pi_n = \mathcal{S}(\varepsilon)\times\mathcal{T}\times[0,1]\times[-\delta_{1n},\delta_{1n}]^{r+1}\times[-\delta_{2n},\delta_{2n}]^{r+1}$, where $\mathcal{S}(\varepsilon) = \{(x,\tau) \in \mathcal{X}\times\mathcal{T} : x \in \mathcal{S}_\tau(\varepsilon)\}$. Using Proposition 2.5 of Giné

(1997),

$$
\begin{aligned}
&\mathbf{E}\left[\mathbf{E}^*\left[\sup_{a,b:||a||\le\delta_{1n},||b||\le\delta_{2n}}\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}|\zeta^{\Delta*}_{n,x,\tau,k}(a,b)-\mathbf{E}^*[\zeta^{\Delta*}_{n,x,\tau,k}(a,b)]|\right]\right]\\
&\le\ C\mathbf{E}\left[\mathbf{E}_{N_i}\left(\sup_{\pi\in\Pi_n}\left|\sum_{i=1}^n(N_i-1)\left\{g_{n,x,\tau,k}(S_i;s,b,a)-\frac{1}{n}\sum_{i=1}^n g_{n,x,\tau,k}(S_i;s,b,a)\right\}\right|\right)\right],
\end{aligned}
$$

where $\{N_i\}_{i=1}^n$ are i.i.d. Poisson random variables with mean 1 independent of $\{(Y_i^\top,X_i^\top)^\top\}_{i=1}^\infty$, $\mathbf{E}_{N_i}$ denotes expectation only with respect to the distribution of $\{N_i\}_{i=1}^n$, and $g_n(\cdot;s,b,a)$ is as defined in the proof of Lemma QR1(i). Here the constant $C>0$ does not depend on $P\in\mathcal{P}$. We can bound the above by

$$
\begin{aligned}
&C\mathbf{E}\left[\sup_{\pi\in\Pi_n}\left|\sum_{i=1}^n(N_i-1)\left(g_{n,x,\tau,k}(S_i;s,b,a)-\mathbf{E}\left[g_{n,x,\tau,k}(S_i;s,b,a)\right]\right)\right|\right]\\
&+C\mathbf{E}\left(\left|\sum_{i=1}^n(N_i-1)\right|\right)\times\mathbf{E}\left(\sup_{\pi\in\Pi_n}\left|\frac{1}{n}\sum_{i=1}^n g_{n,x,\tau,k}(S_i;s,b,a)-\mathbf{E}\left[g_{n,x,\tau,k}(S_i;s,b,a)\right]\right|\right).
\end{aligned}
$$

The leading expectation is bounded by $O(\delta_{2n}\sqrt{\log n}/(n^{1/4}h^{d/4}))$ similarly as in the proof of Lemma QR1(i). And the product of the two expectations in the second term is bounded by

$$
\begin{aligned}
&O(\sqrt{n})\times\mathbf{E}\left(\sup_{\pi\in\Pi_n}\left|\frac{1}{n}\sum_{i=1}^n\left\{g_{n,x,\tau,k}(S_i;s,b,a)-\mathbf{E}\left[g_{n,x,\tau,k}(S_i;s,b,a)\right]\right\}\right|\right)\\
&=\ O\left(\delta_{2n}\sqrt{\log n}/(n^{1/4}h^{d/4})\right),
\end{aligned}
$$

where the constant $C>0$ does not depend on $P\in\mathcal{P}$, and the last equality follows similarly as in the proof of Lemma QR1(i).

(ii) Note that

$$
\mathbf{E}^*[\zeta^{\Delta*}_{n,x,\tau,k}(a,b)]=\mathbf{E}^*[\zeta^{\Delta*}_{n,x,\tau,k}(a,b)]-\mathbf{E}[\zeta^{\Delta}_{n,x,\tau,k}(a,b)]+\mathbf{E}[\zeta^{\Delta}_{n,x,\tau,k}(a,b)]. \tag{B.20}
$$

The difference between the first two terms on the right hand side is

$$
O_P\left(\frac{\delta_{2n}\sqrt{\log n}}{n^{1/4}h^{d/4}}+\frac{\delta_{2n}\log n}{n^{1/2}h^{d/2}}\right)=O_P\left(\frac{\delta_{2n}\sqrt{\log n}}{n^{1/4}h^{d/4}}\right),
$$

uniformly in $P\in\mathcal{P}$, as we have seen in (i). We apply Lemma QR1(iii) to the last expectation in (B.20) to obtain the desired result. ∎

Lemma QR4. *Suppose that Assumptions QR1-QR3 hold. Then for each $k \in \mathbb{N}_J$,*

$$\sup_{(x,\tau)\in\mathcal{X}_1\times\mathcal{T}} \left\| \sqrt{nh^d} H(\hat{\gamma}^*_{\tau,k}(x) - \hat{\gamma}_{\tau,k}(x)) - M^{-1}_{n,\tau,k}(x)\psi^*_{n,x,\tau,k} \right\|$$
$$= O_{P^*}\left(\frac{\log^{1/2} n}{n^{1/4}h^{d/4}}\right), \ \mathcal{P}\text{-uniformly.}$$

Proof of Lemma QR4. The proof uses Lemma QR3 in the same way as the proof of Lemma QR2 uses Lemma QR1. Details are omitted. ∎

Proof of Theorem AUC. First, let us turn to Assumption A1. Since $\sqrt{nh^d} \times o_P(n^{-1/2}) = o_P(h^{d/2})$ by Assumption AUC-4, it suffices to consider $\hat{v}_{\tau,2}(x)$ that uses $\underline{b}$ instead of $\hat{\underline{b}}$. Hence By Lemma QR2, the asymptotic linear representation in Assumption A1 follows. The error rate $o_P(\sqrt{h^d})$ is satisfied, because

$$h^{-d/2}\left(\frac{\log^{1/2} n}{n^{1/4}h^{d/4}}\right) = n^{-1/4}h^{-3d/4}\log^{1/2} n \to 0,$$

by Assumption AUC-3(ii) and the condition $r > 3d/2 - 1$. Assumption A2 follows because both $\beta_{n,x,\tau,1}(S_i, z)$ and $\beta_{n,x,\tau,2}(S_i, z)$ have a multiplicative component of $K(z)$ which has a compact support by Assumption AUC-3(i). As for Assumption A3, we use Lemma 2 in Appendix A. First define

$$\begin{aligned} e_{x,\tau,k,li} &\equiv 1\{L_i = k\}\tilde{l}_\tau\left(B_{li} - \gamma^\top_{\tau,k}(x)\cdot H \cdot c\left(\frac{X_i - x}{h}\right)\right) \text{ and} \\ \xi_{x,\tau,k,i} &\equiv \mathbf{e}_1^\top M^{-1}_{n,\tau,k}(x) c\left(\frac{X_i - x}{h}\right) K\left(\frac{X_i - x}{h}\right) \end{aligned}$$

First observe that for each fixed $x_2 \in \mathbf{R}^d, \tau_2 \in \mathcal{T}$, and $\lambda > 0$,

$$\begin{aligned} \text{(B.21)} \quad & \mathbf{E}\left[\sup_{||x_2 - x_3|| + ||\tau_2 - \tau_3|| \le \lambda}\left(\alpha_{n,x_2,\tau_2,2}\left(Y_i, \frac{X_i - x_2}{h}\right) - \alpha_{n,x_3,\tau_3,2}\left(Y_i, \frac{X_i - x_3}{h}\right)\right)^2\right] \\ \le\ & 2\sum_{l=1}^k \mathbf{E}\left[\mathbf{E}\left[\sup_{||x_2 - x_3|| + ||\tau_2 - \tau_3|| \le \lambda}(e_{x_2,\tau_2,k,li} - e_{x_3,\tau_3,k,li})^2 | X_i\right]\xi^2_{x_2,\tau_2,k,i}\right] \\ & + 2\sum_{l=1}^k \mathbf{E}\left[\sup_{||x_2 - x_3|| + ||\tau_2 - \tau_3|| \le \lambda}(\xi_{x_2,\tau_2,k,i} - \xi_{x_3,\tau_3,k,i})^2\right]. \end{aligned}$$

Using Lipschitz continuity of the conditional density of B_{li} given $L_i = k$ and $X_i = x$ in (x,τ) (Assumption AUC-2(iii)) and Lipschitz continuity of $\gamma_{\tau,k}(x)$ in (x,τ) (Assumption AUC-1(i)), we find that the first term is bounded by $Ch^{-s_1}\lambda$ for some $C > 0$ and $s_1 > 0$. Since

$$M_{n,\tau,k}(x) = kP\{L_i = k | X_i = x\} f_{\tau,k}(0|x) f(x) \int K(t)c(t)c(t)^\top dt + o(1),$$

we find that $M_{n,\tau,k}^{-1}(x)$ is Lipschitz continuous in (x,τ) by Assumptions AUC-1(i), AUC-2(iii)(iv) and AUC-3(i). Hence the last term in (B.21) is also bounded by $Ch^{-s_2}\lambda^2$ for some $C > 0$ and $s_2 > 0$. Therefore, the condition in (4.11) holds with

$$b_{n,ij}(x,\tau) = \alpha_{n,x,\tau,2}\left(Y_i, \frac{X_i - x}{h}\right).$$

Also, observe that

$$\mathbf{E}\left[\left|\alpha_{n,x,\tau,2}\left(Y_i, \frac{X_i - x}{h}\right)\right|^4\right] \leq C,$$

because $\alpha_{n,x,\tau,2}(\cdot,\cdot)$ is uniformly bounded. We also obtain the same result for $\alpha_{n,x,\tau,3}(\cdot,\cdot)$. Thus the conditions of Lemma 2 are satisfied with $b_{n,ij}(x,\tau)$ taken to be $\beta_{n,x,\tau,1}(Y_i,(X_i-x)/h)$ or $\beta_{n,x,\tau,2}(Y_i,(X_i-x)/h)$. Now Assumption 3 follows from Lemma 2. Since we are taking $\hat{\sigma}_{\tau,j}(x) = 1$, it suffices to take $\sigma_{n,\tau,j}(x) = 1$ in Assumption A5. Assumption A6(i) is satisfied because $\beta_{n,x,\tau,j}$ is bounded. Assumption B1 follows by Lemma QR4. Assumption B4 follows by taking $M \to \infty$, because $\beta_{n,x,\tau,j}$ is bounded. ∎

Appendix C. Proofs of Auxiliary Results for Lemmas A2(i), Lemma A4(i), and Theorem 1

The eventual result in this appendix is Lemma C9 which is used to show the asymptotic normality of the location-scale normalized representation of $\hat{\theta}$ and its bootstrap version, and to establish its asymptotic behavior in the degenerate case. For this, we first prepare Lemmas C1-C3. To obtain uniformity that covers the case of degeneracy, this paper uses a method of regularization, where the covariance matrix of random quantities is added by a diagonal matrix of small diagonal elements. The regularized random quantities having this covariance matrix does not suffer from degeneracy in the limit, even when the original quantities have covariate matrix that is degenerate in the limit. Thus, for these regularized quantities, we can obtain uniform asymptotic theory using an appropriate Berry-Esseen-type bound. Then, we need to deal with the difference between the regularized covariance matrix and the original one. Lemma C1 is a simple result of linear algebra that is used to control this discrepancy.

Lemma C2 has two sub-results from which one can deduce a uniform version of Levy's continuity theorem. We have not seen any such results in the literature or monographs, so we provide its full proof. The result has two functions. First, the result enables one to deduce convergence in distribution in terms of convergence of cumulative distribution functions and convergence in distribution in terms of convergence of characteristic functions in a manner that is uniform over a given collection of probabilities. The original form of convergence in distribution due to the Poissonization method in Giné, Mason, and Zaitsev (2003) is convergence of characteristic functions. Certainly pointwise in P, this convergence implies convergence of cumulative distribution functions, but it is not clear under what conditions this implication is uniform over a given class of probabilities. Lemma C2 essentially clarifies this issue.

Lemma C3 is an extension of de-Poissonization lemma that appeared in Beirlant and Mason (1995). The proof is based on the proof of their same result in Giné, Mason, and Zaitsev (2003), but involves a substantial modification, because unlike their results, we need a version that holds uniformly over $P \in \mathcal{P}$. This de-Poissonization lemma is used to transform the asymptotic distribution theory for the Poissonized version of the test statistic into that for the original test statistic.

Lemmas C4-C5 establish some moment bounds for a normalized sum of independent quantities. This moment bound is later used to control a Berry-Esseen-type bound, when we approximate those sums by corresponding centered normal random vectors.

Lemma C6 obtains an approximate version for the scale normalizer σ_n. The approximate version involves a functional of a Gaussian random vector, which stems from approximating a normalized sum of independent random vectors by a Gaussian random vector through using a Berry-Esseen-type bound. For this result, we use the regularization method that we mentioned before. Due to the regularization, we are able to cover the degenerate case eventually.

Lemma C7 is an auxiliary result that is used to establish Lemma C9 in combination with the de-Poissonization lemma (Lemma C3). And Lemma C8 establishes asymptotic normality of the Poissonized version of the test statistics. The asymptotic normality for the Poissonized statistic involves the discretization of the integrals, thereby, reducing the integral to a sum of 1-dependent random variables, and then applies the Berry-Esseen-type bound in Shergin (1993). Note that by the moment bound in Lemmas C4-C5 that is uniform over $P \in \mathcal{P}$, we obtain the asymptotic approximation that is uniform over $P \in \mathcal{P}$. The lemma also presents a corresponding result for the degenerate case.

Finally, Lemma C9 combines the asymptotic distribution theory for the Poissonized test statistic in Lemma C7 with the de-Poissonization lemma (Lemma C3) to obtain the asymptotic distribution theory for the original test statistic. The result of Lemma C9 is used to establish the asymptotic normality result in Lemma A7.

The following lemma provides some inequality of matrix algebra.

Lemma C1. *For any $J \times J$ positive semidefinite symmetric matrix Σ and any $\varepsilon > 0$,*

$$\left\| (\Sigma + \varepsilon I)^{1/2} - \Sigma^{1/2} \right\| \leq \sqrt{J\varepsilon},$$

where $\|A\| = \sqrt{tr(AA')}$ for any square matrix A.

Remark 1. The main point of Lemma C1 is that the bound $\sqrt{J\varepsilon}$ is independent of the matrix Σ. Such a uniform bound is crucially used for the derivation of asymptotic validity of the test uniform in $P \in \mathcal{P}$.

Proof of Lemma C1. First observe that

$$\begin{aligned} & tr\{(\Sigma + \varepsilon I)^{1/2} - \Sigma^{1/2}\}^2 \\ = \; & tr\,(2\Sigma + \varepsilon I) - 2tr((\Sigma + \varepsilon I)^{1/2}\,\Sigma^{1/2}). \end{aligned} \tag{C.1}$$

Since $\Sigma \leq \Sigma + \varepsilon I$, we have $\Sigma^{1/2} \leq (\Sigma + \varepsilon I)^{1/2}$. For any positive semidefinite matrices A and B, $tr(AB) \geq 0$ (see e.g. Abadir and Magnus (2005)). Therefore, $tr(\Sigma) \leq tr((\Sigma + \varepsilon I)^{1/2}\,\Sigma^{1/2})$. From (C.1), we find that

$$\begin{aligned} & tr\,(2\Sigma + \varepsilon I) - 2tr((\Sigma + \varepsilon I)^{1/2}\,\Sigma^{1/2}) \\ \leq \; & tr\,(2\Sigma + \varepsilon I) - 2tr(\Sigma) = \varepsilon J. \end{aligned}$$

∎

The following lemma can be used to derive a version of Levy's Continuity Theorem that is uniform in $P \in \mathcal{P}$.

Lemma C2. *Suppose that $V_n \in \mathbf{R}^d$ is a sequence of random vectors and $V \in \mathbf{R}^d$ is a random vector. We assume without loss of generality that V_n and V live on the same measure space $(\Omega, \mathcal{F})$, and $\mathcal{P}$ is a given collection of probabilities on $(\Omega, \mathcal{F})$. Furthermore define*

$$\begin{aligned} \varphi_n(t) &\equiv \mathbf{E}\left[\exp(it^\top V_n)\right], \; \varphi(t) \equiv \mathbf{E}\left[\exp(it^\top V)\right], \\ F_n(t) &\equiv P\left\{V_n \leq t\right\}, \; \text{and } F(t) \equiv P\left\{V \leq t\right\}. \end{aligned}$$

(i) Suppose that the distribution $P \circ V^{-1}$ is uniformly tight in $\{P \circ V^{-1} : P \in \mathcal{P}\}$. Then for any continuous function f on $\mathbf{R}^d$ taking values in $[-1, 1]$ and for any $\varepsilon \in (0, 1]$, we have

$$\sup_{P \in \mathcal{P}} |\mathbf{E}f(V_n) - \mathbf{E}f(V)| \leq \varepsilon^{-d} C_d \sup_{P \in \mathcal{P}} \sup_{t \in \mathbf{R}^d} |F_n(t) - F(t)| + 4\varepsilon,$$

where $C_d > 0$ is a constant that depends only on d.
(ii) Suppose that $\sup_{P \in \mathcal{P}} \mathbf{E}||V||^2 < \infty$. If

$$\sup_{P \in \mathcal{P}} \sup_{u \in \mathbf{R}^d} |\varphi_n(u) - \varphi(u)| \to 0, \text{ as } n \to \infty,$$

then for each $t \in \mathbf{R}^d$,

$$\sup_{P \in \mathcal{P}} |F_n(t) - F(t)| \to 0, \text{ as } n \to \infty.$$

On the other hand, if for each $t \in \mathbf{R}^d$,

$$\sup_{P \in \mathcal{P}} |F_n(t) - F(t)| \to 0, \text{ as } n \to \infty,$$

then for each $u \in \mathbf{R}^d$,

$$\sup_{P \in \mathcal{P}} |\varphi_n(u) - \varphi(u)| \to 0, \text{ as } n \to \infty.$$

Proof of Lemma C2. (i) The proof uses arguments in the proof of Lemma 2.2 of van der Vaart (1998). Take a large compact rectangle $B \subset \mathbf{R}^d$ such that $P\{V \notin B\} < \varepsilon$. Since the distribution of V is tight uniformly over $P \in \mathcal{P}$, we can take such B independently of $P \in \mathcal{P}$. Take a partition $B = \cup_{j=1}^{J_\varepsilon} B_j$ and points $x_j \in B_j$ such that $J_\varepsilon \leq C_{d,1}\varepsilon^{-d}$, and $|f(x) - f_\varepsilon(x)| < \varepsilon$ for all $x \in B$, where $C_{d,1} > 0$ is a constant that depends only on d, and

$$f_\varepsilon(x) \equiv \sum_{j=1}^{J_\varepsilon} f(x_j) 1\{x \in B_j\}.$$

Thus we have

$$\begin{aligned}
|\mathbf{E}f(V_n) - \mathbf{E}f(V)| &\leq |\mathbf{E}f(V_n) - \mathbf{E}f_\varepsilon(V_n)| + |\mathbf{E}f_\varepsilon(V_n) - \mathbf{E}f_\varepsilon(V)| + |\mathbf{E}f_\varepsilon(V) - \mathbf{E}f(V)| \\
&\leq 2\varepsilon + P\{V_n \notin B\} + P\{V \notin B\} + |\mathbf{E}f_\varepsilon(V_n) - \mathbf{E}f_\varepsilon(V)| \\
&\leq 4\varepsilon + |P\{V_n \notin B\} - P\{V \notin B\}| + |\mathbf{E}f_\varepsilon(V_n) - \mathbf{E}f_\varepsilon(V)| \\
&= 4\varepsilon + |P\{V_n \in B\} - P\{V \in B\}| + |\mathbf{E}f_\varepsilon(V_n) - \mathbf{E}f_\varepsilon(V)|.
\end{aligned}$$

The second inequality following by $P\{V \notin B\} < \varepsilon$. As for the last term, we let

$$b_n \equiv \sup_{P \in \mathcal{P}} \sup_{t \in \mathbf{R}^d} |F_n(t) - F(t)|,$$

and observe that

$$\begin{aligned}
|\mathbf{E}f_\varepsilon(V_n) - \mathbf{E}f_\varepsilon(V)| &\leq \sum_{j=1}^{J_\varepsilon} |P\{V_n \in B_j\} - P\{V \in B_j\}| \, |f(x_j)| \\
&\leq \sum_{j=1}^{J_\varepsilon} |P\{V_n \in B_j\} - P\{V \in B_j\}| \leq C_{d,2} b_n J_\varepsilon,
\end{aligned}$$

where $C_{d,2} > 0$ is a constant that depends only on d. The last inequality follows because for any rectangle B_j, we have $|P\{V_n \in B_j\} - P\{V \in B_j\}| \leq C_{d,2} b_n$ for some $C_{d,2} > 0$. We conclude that

$$|\mathbf{E}f(V_n) - \mathbf{E}f(V)| \leq 4\varepsilon + C_{d,2}\left(C_{d,1}\varepsilon^{-d} + 1\right) b_n \leq 4\varepsilon + C_d \varepsilon^{-d} b_n,$$

where $C_d = C_{d,2}\{C_{d,1} + 1\}$. The last inequality follows because $\varepsilon \leq 1$.

(ii) We show the first statement. We first show that under the stated condition, the sequence $\{P \circ V_n^{-1}\}_{n=1}^{\infty}$ is uniformly tight uniformly over $P \in \mathcal{P}$. That is, for any $\varepsilon > 0$, we show there exists a compact set $B \subset \mathbf{R}^d$ such that for all $n \geq 1$,

$$\sup_{P \in \mathcal{P}} P\left\{V_n \in \mathbf{R}^d \backslash B\right\} < \varepsilon.$$

For this, we assume $d = 1$ without loss of generality, let P_n denote the distribution of V_n and consider the following: (using arguments in the proof of Theorem 3.3.6 of Durrett (2010))

$$\begin{aligned} P\left\{|V_n| > \frac{2}{u}\right\} &\leq 2\int_{|x|>2/u}\left(1 - \frac{1}{|ux|}\right) dP_n(x) \\ &\leq 2\int\left(1 - \frac{\sin ux}{ux}\right) dP_n(x) \\ &= \frac{1}{u}\int_{-u}^{u}(1 - \varphi_n(t))\, dt. \end{aligned}$$

Define $\bar{e}_n \equiv \sup_{P \in \mathcal{P}} \sup_{t \in \mathbf{R}} |\varphi_n(t) - \varphi(t)|$. Using Theorem 3.3.8 of Durrett (2010), we bound the last term by

$$\begin{aligned} 2\bar{e}_n + \frac{1}{u}\int_{-u}^{u}(1 - \varphi(t))\, dt &\leq 2\bar{e}_n + \left|\frac{1}{u}\int_{-u}^{u}\left(-it\mathbf{E}V + \frac{t^2\mathbf{E}V^2}{2}\right) dt\right| \\ &\quad +2\left|\frac{1}{u}\int_{-u}^{u} t^2\mathbf{E}V^2 dt\right|. \end{aligned}$$

The supremum of the right hand side terms over $P \in \mathcal{P}$ vanishes as we send $n \to \infty$ and then $u \downarrow 0$, by the assumption that $\sup_{P \in \mathcal{P}}\mathbf{E}|V|^2 < \infty$. Hence the sequence $\{P \circ V_n^{-1}\}_{n=1}^{\infty}$ is uniformly tight uniformly over $P \in \mathcal{P}$.

Now, for each $t \in \mathbf{R}^d$, there exists a subsequence $\{n'\} \subset \{n\}$ and $\{P_{n'}\} \subset \mathcal{P}$ such that

$$\text{(C.2)} \qquad \limsup_{n \to \infty} \sup_{P \in \mathcal{P}} |F_n(t) - F(t)| = \lim_{n' \to \infty} |F_{n'}(t; P_{n'}) - F(t; P_{n'})|\,,$$

where

$$F_n(t; P_n) = P_n\{V_n \leq t\} \text{ and } F(t; P_n) = P_n\{V \leq t\}\,.$$

(Hence, $F(t; P_n)$ is the cdf of distribution $P_n \circ V^{-1}$.)

Since $\{P_{n'} \circ V_{n'}^{-1}\}_{n'=1}^{\infty}$ is uniformly tight (as shown above), there exists a subsequence $\{n'_k\} \subset \{n'\}$ such that

$$\text{(C.3)} \qquad F_{n'_k}(t; P_{n'_k}) \to F^*(t), \text{ as } k \to \infty,$$

for some cdf F^*. Also $\{P_{n'} \circ V^{-1}\}_{n'=1}^{\infty}$ is uniformly tight (because $\sup_{P \in \mathcal{P}} \mathbf{E}||V||^2 < \infty$), $\{P_{n'_k} \circ V^{-1}\}_{k=1}^{\infty}$ is uniformly tight and hence there exists a further subsequence $\{n'_{k_j}\} \subset \{n'_k\}$

such that

$$F(t;P_{n'_{k_j}}) \to F^{**}(t), \text{ as } j \to \infty, \tag{C.4}$$

for some cdf F^{**}. Since $\{n'_{k_j}\} \subset \{n'_k\}$, we have from (C.3),

$$F_{n'_{k_j}}(t;P_{n'_{k_j}}) \to F^{*}(t), \text{ as } j \to \infty. \tag{C.5}$$

By the condition of (ii), we have

$$\left|\varphi_{n'_{k_j}}(u;P_{n'_{k_j}}) - \varphi(u;P_{n'_{k_j}})\right| \to 0, \text{ as } j \to \infty, \tag{C.6}$$

where

$$\varphi_n\left(u;P_n\right) = \mathbf{E}_{P_n}\left(\exp\left(iuV_n\right)\right) \text{ and } \varphi\left(u;P_n\right) = \mathbf{E}_{P_n}\left(\exp\left(iuV\right)\right),$$

and $\mathbf{E}_{P_n}$ represents expectation with respect to the probability measure P_n. Furthermore, by (C.4) and (C.5), and Levy's Continuity Theorem,

$$\lim_{j\to\infty} \varphi_{n'_{k_j}}(u;P_{n'_{k_j}}) \text{ and } \lim_{j\to\infty} \varphi(u;P_{n'_{k_j}})$$

exist and coincide by (C.6). Therefore, for all $t \in \mathbf{R}^d$,

$$F^{**}(t) = F^{*}(t).$$

In other words,

$$\lim_{n'\to\infty} \left|F_{n'}\left(t;P_{n'}\right) - F\left(t;P_{n'}\right)\right| = \lim_{n'\to\infty} \left|F_{n'_{k_j}}\left(t;P_{n'_{k_j}}\right) - F\left(t;P_{n'_{k_j}}\right)\right| = 0.$$

Therefore, the first statement of (ii) follows by the last limit applied to (C.2).

Let us turn to the second statement. Again, we show that $\{P \circ V_n^{-1}\}_{n=1}^{\infty}$ is uniformly tight uniformly in $P \in \mathcal{P}$. Note that given a large rectangle B,

$$P\left\{V_n \in \mathbf{R}^d \backslash B\right\} \le \left|P\left\{V_n \in \mathbf{R}^d \backslash B\right\} - P\left\{V \in \mathbf{R}^d \backslash B\right\}\right| + P\left\{V \in \mathbf{R}^d \backslash B\right\}.$$

There exists N such that for all $n \ge N$, the first difference vanishes as $n \to \infty$, uniformly in $P \in \mathcal{P}$, by the condition of the lemma. As for the second term, we bound it by

$$P\left\{V_j > a_j,\ j = 1, \cdots, d\right\} \le \sum_{j=1}^{d} \frac{\mathbf{E}V_j^2}{a_j},$$

where V_j is the j-th entry of V and $B = \times_{j=1}^{d}[a_j, b_j]$, $b_j < 0 < a_j$. By taking a_j's large enough, we make the last bound arbitrarily small independently of $P \in \mathcal{P}$, because $\sup_{P\in\mathcal{P}}\mathbf{E}V_j^2 < \infty$ for each $j = 1, \cdots, d$. Therefore, $\{P \circ V_n^{-1}\}_{n=1}^{\infty}$ is uniformly tight uniformly in $P \in \mathcal{P}$.

Now, we turn to the proof of the second statement of (ii). For each $u \in \mathbf{R}^d$, there exists a subsequence $\{n'\} \subset \{n\}$ and $\{P_{n'}\} \subset \mathcal{P}$ such that

$$\limsup_{n\to\infty} \sup_{P\in\mathcal{P}} |\varphi_n(u) - \varphi(u)| = \lim_{n'\to\infty} |\varphi_{n'}(u; P_{n'}) - \varphi(u; P_{n'})|,$$

where $\varphi_n(u; P_n) = \mathbf{E}_{P_n} \exp(iu^\top V_n)$ and $\varphi(u; P_n) = \mathbf{E}_{P_n} \exp(iu^\top V)$. By the condition in the second statement of (ii), for each $t \in \mathbf{R}^d$,

$$\lim_{n'\to\infty} |F_{n'}(t; P_{n'}) - F(t; P_{n'})| = 0. \tag{C.7}$$

Since $\{P_{n'} \circ V_{n'}^{-1}\}_{n'=1}^{\infty}$ is uniformly tight (as shown above), there exists a subsequence $\{n'_k\} \subset \{n'\}$ such that $F_{n'_k}(t; P_{n'_k}) \to F^*(t)$, as $k \to \infty$, and hence by Levy's Continuity Theorem, we have $\varphi_{n'_k}(u; P_{n'_k}) \to \varphi^*(u)$, as $k \to \infty$. Similarly, we also have $\varphi(u; P_{n'_k}) \to \varphi^{**}(u)$, as $k \to \infty$. By (C.7), we have $F^*(t) = F^{**}(t)$ and $\varphi^*(u) = \varphi^{**}(u)$. Therefore,

$$\lim_{n'\to\infty} |\varphi_{n'}(u; P_{n'}) - \varphi(u; P_{n'})| = \lim_{n'\to\infty} \left|\varphi_{n'_{k_j}}\left(u; P_{n'_{k_j}}\right) - \varphi\left(u; P_{n'_{k_j}}\right)\right| = 0.$$

Thus we arrive at the desired result. ∎

The following lemma offers a version of the de-Poissonization lemma of Beirlant and Mason (1995) (see Theorem 2.1 on page 5). In contrast to the result of Beirlant and Mason (1995), the version here is uniform in $P \in \mathcal{P}$.

Lemma C3. *Let $N_{1,n}(\alpha)$ and $N_{2,n}(\alpha)$ be independent Poisson random variables with $N_{1,n}(\alpha)$ being Poisson $(n(1-\alpha))$ and $N_{2,n}(\alpha)$ being Poisson $(n\alpha)$, where $\alpha \in (0,1)$. Denote $N_n(\alpha) = N_{1,n}(\alpha) + N_{2,n}(\alpha)$ and set*

$$U_n(\alpha) = \frac{N_{1,n}(\alpha) - n(1-\alpha)}{\sqrt{n}} \text{ and } V_n(\alpha) = \frac{N_{2,n}(\alpha) - n\alpha}{\sqrt{n}}.$$

Let $\{S_n\}_{n=1}^{\infty}$ be a sequence of random variables and $\mathcal{P}$ be a given set of probabilities P on a measure space on which $(S_n, U_n(\alpha_P), V_n(\alpha_P))$ lives, where $\alpha_P \in (0,1)$ is a quantity that may depend on $P \in \mathcal{P}$ and for some $\varepsilon > 0$,

$$\varepsilon \le \inf_{P\in\mathcal{P}} \alpha_P \le \sup_{P\in\mathcal{P}} \alpha_P \le 1 - \varepsilon. \tag{C.8}$$

Furthermore, assume that for each $n \ge 1$, the random vector $(S_n, U_n(\alpha_P))$ is independent of $V_n(\alpha_P)$ with respect to each $P \in \mathcal{P}$. Let for $t_1, t_2 \in \mathbf{R}^2$,

$$b_{n,P}(t_1, t_2; \sigma_P) \equiv \left|P\{S_n \le t_1, U_n(\alpha_P) \le t_2\} - P\{\sigma_P \mathbb{Z}_1 \le t_1, \sqrt{1-\alpha_P}\mathbb{Z}_2 \le t_2\}\right|,$$

where $\mathbb{Z}_1$ and $\mathbb{Z}_2$ are independent standard normal random variables and $\sigma_P^2 > 0$ for each $P \in \mathcal{P}$. (Note that $\inf_{P\in\mathcal{P}} \sigma_P^2$ is allowed to be zero.)

(i) As $n \to \infty$,

$$\sup_{P \in \mathcal{P}} \sup_{t \in \mathbf{R}} \left| \mathbf{E}[\exp(itS_n)|N_n(\alpha_P) = n] - \exp\left(-\frac{\sigma_P^2 t^2}{2}\right) \right|$$
$$\leq 2\varepsilon + \left(4C_d \sup_{P \in \mathcal{P}} a_{n,P}(\varepsilon)\right) \sqrt{\frac{2\pi}{\varepsilon}},$$

where $a_{n,P}(\varepsilon) \equiv \varepsilon^{-d} b_{n,P} + \varepsilon, b_{n,P} \equiv \sup_{t_1, t_2 \in \mathbf{R}} b_{n,P}(t_1, t_2; \sigma_P)$*, and* ε *is the constant in (C.8).*
(ii) Suppose further that for all $t_1, t_2 \in \mathbf{R}$*, as* $n \to \infty$,

$$\sup_{P \in \mathcal{P}} b_{n,P}(t_1, t_2; 0) \to 0.$$

Then, for all $t \in \mathbf{R}$*, we have as* $n \to \infty$,

$$\sup_{P \in \mathcal{P}} |\mathbf{E}[\exp(itS_n)|N_n(\alpha_P) = n] - 1| \to 0.$$

Remark 2. While the proof of Lemma C3 follows that of Lemma 2.4 of Giné, Mason, and Zaitsev (2003), it is worth noting that in contrast to Lemma 2.4 of Giné, Mason, and Zaitsev (2003) or Theorem 2.1 of Beirlant and Mason (1995), Lemma C3 gives an explicit bound for the difference between the conditional characteristic function of S_n given $N_n(\alpha_P) = n$ and the characteristic function of $N(0, \sigma_P^2)$. Under the stated conditions, (in particular (C.8)), the explicit bound is shown to depend on $P \in \mathcal{P}$ only through $b_{n,P}$. Thus in order to obtain a bound uniform in $P \in \mathcal{P}$, it suffices to control α_P and $b_{n,P}$ uniformly in $P \in \mathcal{P}$.

Proof of Lemma C3. (i) Let $\phi_{n,P}(t, u) = \mathbf{E}[\exp(itS_n + iuU_n(\alpha_P))]$ and

$$\phi_P(t, u) = \exp(-(\sigma_P^2 t^2 + (1 - \alpha_P)u^2)/2).$$

By the condition of the lemma and Lemma C2(i), we have for any $\varepsilon > 0$,

$$\text{(C.9)} \qquad \begin{aligned} |\phi_{n,P}(t, u) - \phi_P(t, u)| &\leq (\varepsilon^{-d} C_d b_{n,P} + 4\varepsilon) \\ &\leq 4\varepsilon^{-d} C_d b_{n,P} + 4\varepsilon = 4C_d a_{n,P}(\varepsilon). \end{aligned}$$

Note that $a_{n,P}(\varepsilon)$ depends on $P \in \mathcal{P}$ only through $b_{n,P}$.

Following the proof of Lemma 2.4 of Giné, Mason, and Zaitsev (2003), we have

$$\begin{aligned} \psi_{n,P}(t) &= \mathbf{E}[\exp(itS_n)|N_n(\alpha_P) = n] \\ &= \frac{1}{\sqrt{2\pi}} (1 + o(1)) \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \phi_{n,P}(t, v) \mathbf{E}\left[\exp(ivV_n(\alpha_P))\right] dv, \end{aligned}$$

uniformly over $P \in \mathcal{P}$. Note that the equality comes after applying Sterling's formula to $2\pi P\{N_n(\alpha_P) = n\}$ and change of variables from u to $v/\sqrt{n}$. (See the proof of Lemma 2.4 of Giné, Mason, and Zaitsev (2003).) The distribution of $N_n(\alpha_P)$, being Poisson (n), does not depend on the particular choice of $\alpha_P \in (0, 1)$, and hence the $o(1)$ term is $o(1)$ uniformly over $t \in \mathbf{R}$ and over $P \in \mathcal{P}$. We follow the proof of Theorem 3 of Feller (1966, p.517) to observe that there exists $n_0 > 0$ such that uniformly over $\alpha \in [\varepsilon, 1-\varepsilon]$,

$$\left\{ \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \left| \mathbf{E} \exp(ivV_n(\alpha)) - \exp(-\alpha v^2/2) \right| dv + \int_{|v|>\pi\sqrt{n}} \exp\left(-\alpha v^2/2\right) dv \right\} < \varepsilon,$$

for all $n > n_0$. Note that the distribution of $V_n(\alpha_P)$ depends on $P \in \mathcal{P}$ only through $\alpha_P \in [\varepsilon, 1-\varepsilon]$ and ε does not depend on P. Since there exists n_1 such that for all $n > n_1$,

$$\sup_{P\in\mathcal{P}} \int_{|v|>\pi\sqrt{n}} \exp\left(-\alpha_P v^2/2\right) dv < \varepsilon,$$

the previous inequality implies that for all $n > \max\{n_0, n_1\}$,

$$\begin{aligned}
\text{(C.10)} \qquad & \sup_{P\in\mathcal{P}} \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \left| \phi_{n,P}(t,u) \left(\mathbf{E} \exp(iuV_n(\alpha_P)) - \exp(-\alpha_P u^2/2) \right) \right| du \\
\leq\ & \sup_{P\in\mathcal{P}} \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \left(\sup_{P\in\mathcal{P}} |\phi_{n,P}(t,u)| \right) |\mathbf{E} \exp(iuV_n(\alpha_P)) - \exp(-\alpha_P u^2/2)| du \\
\leq\ & \sup_{P\in\mathcal{P}} \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} |\mathbf{E} \exp(iuV_n(\alpha_P)) - \exp(-\alpha_P u^2/2)| du \leq \varepsilon.
\end{aligned}$$

By (C.9) and (C.10),

$$\begin{aligned}
& \sup_{P\in\mathcal{P}} \left| \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \phi_{n,P}(t,u) \mathbf{E}\left[\exp(iuV_n(\alpha_P))\right] du - \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \phi_P(t,u) \exp\left(-\alpha_P u^2/2\right) du \right| \\
\leq\ & \sup_{P\in\mathcal{P}} \sup_{\alpha\in[\varepsilon,1-\varepsilon]} \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \left| \phi_{n,P}(t,u) \left(\mathbf{E} \exp(iuV_n(\alpha)) - \exp(-\alpha u^2/2) \right) \right| du \\
& + \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \sup_{P\in\mathcal{P}} \sup_{\alpha\in[\varepsilon,1-\varepsilon]} |\phi_{n,P}(t,u) - \phi_P(t,u)| \exp(-\alpha u^2/2) du \\
\leq\ & \varepsilon + \left(4C_d \sup_{P\in\mathcal{P}} a_{n,P}(\varepsilon) \right) \sup_{\alpha\in[\varepsilon,1-\varepsilon]} \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \exp(-\alpha u^2/2) du \\
\leq\ & \varepsilon + \left(4C_d \sup_{P\in\mathcal{P}} a_{n,P}(\varepsilon) \right) \sup_{\alpha\in[\varepsilon,1-\varepsilon]} \sqrt{\frac{2\pi}{\alpha}} = \varepsilon + \left(4C_d \sup_{P\in\mathcal{P}} a_{n,P}(\varepsilon) \right) \sqrt{\frac{2\pi}{\varepsilon}}
\end{aligned}$$

as $n \to \infty$. Since

$$\exp\left(-\frac{\sigma_P^2 t^2}{2}\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \phi_P(t,u) \exp\left(-\frac{\alpha_P u^2}{2}\right) du,$$

and from some large n on that does not depend on $P \in \mathcal{P}$,

$$\begin{aligned}&\left|\int_{-\infty}^{\infty} \phi_P(t,u)\exp\left(-\frac{\alpha_P u^2}{2}\right)du - \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}} \phi_P(t,u)\exp\left(-\frac{\alpha_P u^2}{2}\right)du\right| \\ &= \exp\left(-\frac{\sigma_P^2 t^2}{2}\right)\left|\int_{-\infty}^{\infty}\exp\left(-\frac{u^2}{2}\right)du - \int_{-\pi\sqrt{n}}^{\pi\sqrt{n}}\exp\left(-\frac{u^2}{2}\right)du\right| < \varepsilon,\end{aligned}$$

we conclude that for each $t \in \mathbf{R}$,

$$\left|\psi_{n,P}(t) - \exp\left(-\frac{\sigma_P^2 t^2}{2}\right)\right| \leq 2\varepsilon + \left(4C_d \sup_{P\in\mathcal{P}} a_{n,P}(\varepsilon)\right)\sqrt{\frac{2\pi}{\varepsilon}},$$

as $n \to \infty$. Since the right hand side does not depend on $t \in \mathbf{R}$ and $P \in \mathcal{P}$, we obtain the desired result.

(ii) By the condition of the lemma and Lemma C2(ii), we have for any $t, u \in \mathbf{R}$,

$$\sup_{P\in\mathcal{P}} |\phi_{n,P}(t,u) - \phi_P(0,u)| \to 0,$$

as $n \to \infty$. The rest of the proof is similar to that of (i). We omit the details. ∎

Define for $x \in \mathcal{X}$, $\tau_1, \tau_2 \in \mathcal{T}$, and $j,k \in \mathbb{N}_J$,

$$k_{n,\tau,j,m}(x) \equiv \frac{1}{h^d}\mathbf{E}\left[\left|\beta_{n,x,\tau,j}\left(Y_{ij}, \frac{X_i - x}{h}\right)\right|^m\right].$$

Lemma C4. *Suppose that Assumption A6(i) holds. Then for all $m \in [2, M]$, (with $M > 0$ being the constant that appears in Assumption A6(i)), there exists $C_1 \in (0,\infty)$ that does not depend on n such that for each $j \in \mathbb{N}_J$,*

$$\sup_{\tau\in\mathcal{T}, x\in\mathcal{S}_\tau(\varepsilon)} \sup_{P\in\mathcal{P}} k_{n,\tau,j,m}(x) \leq C_1.$$

Proof of Lemma C4. The proof can be proceeded by using Assumption A6(i) and following the proof of Lemma 4 of Lee, Song, and Whang (2013). ∎

Let N be a Poisson random variable with mean n and independent of $(Y_i^\top, X_i^\top)_{i=1}^\infty$. Also, let $\beta_{n,x,\tau}(Y_i, (X_i - x)/h)$ be the J-dimensional vector whose j-th entry is equal to $\beta_{n,x,\tau,j}(Y_{ij}, (X_i - x)/h)$. We define

$$\begin{aligned}\mathbf{z}_{N,\tau}(x) &\equiv \frac{1}{nh^d}\sum_{i=1}^{N}\beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right) - \frac{1}{h^d}\mathbf{E}\beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right) \text{ and} \\ \mathbf{z}_{n,\tau}(x) &\equiv \frac{1}{nh^d}\sum_{i=1}^{n}\beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right) - \frac{1}{h^d}\mathbf{E}\beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right).\end{aligned}$$

Let N_1 be a Poisson random variable with mean 1, independent of $(Y_i^\top, X_i^\top)_{i=1}^\infty$. Define

$$\begin{aligned} q_{n,\tau}(x) &\equiv \frac{1}{\sqrt{h^d}} \sum_{1 \le i \le N_1} \left\{ \beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right) - \mathbf{E}\beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right) \right\} \text{ and} \\ \bar{q}_{n,\tau}(x) &\equiv \frac{1}{\sqrt{h^d}} \left\{ \beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right) - \mathbf{E}\beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right) \right\}. \end{aligned}$$

Lemma C5. *Suppose that Assumption A6(i) holds. Then for any $m \in [2, M]$ (with $M > 0$ being the constant in Assumption A6(i))*

$$\begin{aligned} \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[||q_{n,\tau}(x)||^m\right] &\le \bar{C}_1 h^{d(1-(m/2))} \text{ and} \\ \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[||\bar{q}_{n,\tau}(x)||^m\right] &\le \bar{C}_2 h^{d(1-(m/2))}, \end{aligned} \tag{C.11}$$

where $\bar{C}_1, \bar{C}_2 > 0$ are constants that depend only on m.

If furthermore, $\limsup_{n\to\infty} n^{-(m/2)+1} h^{d(1-(m/2))} < C$ *for some constant $C > 0$, then*

$$\begin{aligned} \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[||n^{1/2}h^{d/2}\mathbf{z}_{N,\tau}(x)||^m\right] &\le \left(\frac{15m}{\log m}\right)^m \max\left\{\bar{C}_1, 2\bar{C}_1 C\right\} \text{ and} \\ \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[||n^{1/2}h^{d/2}\mathbf{z}_{n,\tau}(x)||^m\right] &\le \left(\frac{15m}{\log m}\right)^m \max\left\{\bar{C}_2, 2\bar{C}_2 C\right\}, \end{aligned} \tag{C.12}$$

where $\bar{C}_1, \bar{C}_2 > 0$ are the constants that appear in (C.11).

Proof of Lemma C5. Let $q_{n,\tau,j}(x)$ be the j-th entry of $q_{n,\tau}(x)$. For the first statement of the lemma, it suffices to observe that for some positive constants C_1 and $\bar{C}$,

$$\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[|q_{n,\tau,j}(x)|^m\right] \le \frac{C_1 h^d k_{n,\tau,j,m}}{h^{dm/2}} \le \bar{C} h^{d(1-(m/2))}, \tag{C.13}$$

where the first inequality uses the definition of $k_{n,\tau,j,m}$, and the last inequality uses Lemma C4 and the fact that $m \in [2, M]$. The second statement in (C.11) follows similarly.

We consider the statements in (C.12). We consider the first inequality in (C.12). Let $z_{N,\tau,j}(x)$ be the j-th entry of $\mathbf{z}_{N,\tau}(x)$. Then using Rosenthal's inequality (e.g. (2.3) of Giné, Mason, and Zaitsev (2003)), we find that

$$\begin{aligned} &\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}[|\sqrt{nh^d} z_{N,\tau,j}(x)|^m] \\ &\le \left(\frac{15m}{\log m}\right)^m \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \max\left\{ \left(\mathbf{E}q_{n,\tau,j}^2(x)\right)^{m/2}, n^{-m/2+1}\mathbf{E}|q_{n,\tau,j}(x)|^m \right\}. \end{aligned}$$

Since $\mathbf{E}q_{n,\tau,j}^2(x) \leq (\mathbf{E}|q_{n,\tau,j}(x)|^m)^{2/m}$, by (C.13), the last term is bounded by

$$\begin{aligned} &\left(\frac{15m}{\log m}\right)^m \max\left\{\bar{C}, \bar{C}n^{-(m/2)+1}h^{d(1-(m/2))}\right\} \\ \leq \quad &\left(\frac{15m}{\log m}\right)^m \max\left\{\bar{C}, 2\bar{C}C\right\}, \end{aligned}$$

from some large n on by the condition $\limsup_{n\to\infty} n^{-(m/2)+1}h^{d(1-(m/2))} < C$.

As for the second inequality in (C.12), for some $C > 0$, we use the second inequality in (C.11) and use Rosenthal's inequality in the same way as before, to obtain the inequality. ∎

The following lemma offers a characterization of the scale normalizer of our test statistic. For $A, A' \subset \mathbb{N}_J$, define $\zeta_{n,\tau}(x) \equiv \sqrt{nh^d}\mathbf{z}_{N,\tau}(x)$,

$$\text{(C.14)} \qquad \begin{aligned} C_{n,\tau,\tau',A,A'}^R(x,x') &\equiv h^{-d}Cov\left(\Lambda_{A,p}\left(\zeta_{n,\tau}(x)\right), \Lambda_{A',p}\left(\zeta_{n,\tau'}(x')\right)\right), \text{ and} \\ C_{n,\tau,\tau',A,A'}(x,u) &\equiv Cov\left(\Lambda_{A,p}\left(\mathbb{W}_{n,\tau,\tau'}^{(1)}(x,u)\right), \Lambda_{A',p}\left(\mathbb{W}_{n,\tau,\tau'}^{(2)}(x,u)\right)\right), \end{aligned}$$

where we recall that $[\mathbb{W}_{n,\tau_1,\tau_2}^{(1)}(x,u)^\top, \mathbb{W}_{n,\tau_1,\tau_2}^{(2)}(x,u)^\top]^\top$ is a mean zero $\mathbf{R}^{2J}$-valued Gaussian random vector whose covariance matrix is given by (6.9).

Then for Borel sets $B, B' \subset \mathcal{S}$ and $A, A' \subset \mathbb{N}_J$, let

$$\sigma_{n,A,A'}^R(B,B') \equiv \int_{B'}\int_B C_{n,\tau,\tau',A,A'}^R(x,x')dQ(x,\tau)dQ(x',\tau')$$

and

$$\text{(C.15)} \qquad \sigma_{n,A,A'}(B,B') \equiv \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_\tau\cap B'_{\tau'}}\int_{\mathcal{U}} C_{n,\tau,\tau',A,A'}(x,u)dudxd\tau d\tau',$$

where $B_\tau \equiv \{x \in \mathcal{X} : (x,\tau) \in B\}$ and $B'_{\tau'} \equiv \{x \in \mathcal{X} : (x,\tau') \in B'\}$.

The lemma below shows that $\sigma_{n,A,A'}^R(B,B')$ and $\sigma_{n,A,A'}(B,B')$ are asymptotically equivalent uniformly in $P \in \mathcal{P}$. We introduce some notation. Recall the definition of $\Sigma_{n,\tau_1,\tau_2}(x,u)$, which is found below (6.7). Define for $\bar{\varepsilon} > 0$,

$$\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) \equiv \begin{bmatrix} \Sigma_{n,\tau_1,\tau_1}(x,0) + \bar{\varepsilon}I_J & \Sigma_{n,\tau_1,\tau_2}(x,u) \\ \Sigma_{n,\tau_1,\tau_2}(x,u) & \Sigma_{n,\tau_2,\tau_2}(x+uh,0) + \bar{\varepsilon}I_J \end{bmatrix},$$

where I_J is the J dimensional identity matrix. Certainly $\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)$ is positive definite. We define

$$\xi_{N,\tau_1,\tau_2}(x,u;\eta_1,\eta_2) \equiv \sqrt{nh^d}\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^{-1/2}(x,u)\begin{bmatrix} \mathbf{z}_{N,\tau_1}(x;\eta_1) \\ \mathbf{z}_{N,\tau_2}(x+uh;\eta_2) \end{bmatrix},$$

where $\eta_1 \in \mathbf{R}^J$ and $\eta_2 \in \mathbf{R}^J$ are random vectors that are independent, and independent of $(Y_i^\top, X_i^\top)_{i=1}^\infty$, each following $N(0,\bar{\varepsilon}I_J)$, and $\mathbf{z}_{N,\tau}(x;\eta_1) \equiv \mathbf{z}_{N,\tau}(x) + \eta_1/\sqrt{nh^d}$. We are prepared to state the lemma.

Lemma C6. *Suppose that Assumption A6(i) holds and that $nh^d \to \infty$, as $n \to \infty$, and*

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

for some constant $C > 0$ and some $m \in [2(p+1), M]$.

Then for any sequences of Borel sets B_n, $B_n' \subset \mathcal{S}$ and for any $A, A' \subset \mathbb{N}_J$,

$$\sigma^R_{n,A,A'}(B_n, B_n') = \sigma_{n,A,A'}(B_n, B_n') + o(1),$$

where $o(1)$ vanishes uniformly in $P \in \mathcal{P}$ as $n \to \infty$.

Remark 3. The main innovative element of Lemma C6 is that the result does not require that $\sigma_{n,A,A'}(B_n, B_n')$ be positive for each finite n or positive in the limit. Hence the result can be applied to the case where the scale normalizer $\sigma^R_{n,A,A'}(B_n, B_n')$ is degenerate (either in finite samples or asymptotically).

Proof of Lemma C6. Define $B_{n,\tau} \equiv \{x \in \mathcal{X} : (x, \tau) \in B_n\}$, $w_{\tau,B_n}(x) \equiv 1_{B_{n,\tau}}(x)$. For a given $\bar{\varepsilon} > 0$, let

$$\begin{aligned} g_{1n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) &\equiv h^{-d} Cov(\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_1}(x;\eta_1)), \Lambda_{A',p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_2}(x+uh;\eta_2))), \\ g_{2n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) &\equiv Cov(\Lambda_{A,p}(\mathbb{Z}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x)), \Lambda_{A',p}(\mathbb{Z}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x+uh))), \end{aligned}$$

and $\left(\mathbb{Z}^\top_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x), \mathbb{Z}^\top_{n,\tau_1,\tau_2,\bar{\varepsilon}}(v)\right)^\top$ is a centered normal $\mathbf{R}^{2J}$-valued random vector with the same covariance matrix as that of $[\sqrt{nh^d}\mathbf{z}^\top_{N,\tau_1}(x;\eta_1), \sqrt{nh^d}\mathbf{z}^\top_{N,\tau_2}(v;\eta_2)]^\top$. Then we define

$$\sigma^R_{n,A,A',\bar{\varepsilon}}(B_n, B_n') \equiv \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_{n,\tau_1}}\int_{\mathcal{U}} g_{1n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) w_{\tau_1,B_n}(x) w_{\tau_2,B_n'}(x+uh) du dx d\tau_1 d\tau_2,$$

and

$$\sigma_{n,A,A',\bar{\varepsilon}}(B_n, B_n') \equiv \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_{n,\tau_1}\cap B'_{n,\tau_2}}\int_{\mathcal{U}} C_{n,\tau_1,\tau_2,A,A',\bar{\varepsilon}}(x,u) du dx d\tau_1 d\tau_2,$$

where

$$C_{n,\tau_1,\tau_2,A,A',\bar{\varepsilon}}(x,u) \equiv Cov\left(\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)), \Lambda_{A',p}(\mathbb{W}^{(2)}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u))\right), \tag{C.16}$$

and, with $\mathbb{Z} \sim N(0, I_{2J})$,

$$\begin{bmatrix} \mathbb{W}^{(1)}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) \\ \mathbb{W}^{(2)}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) \end{bmatrix} \equiv \tilde{\Sigma}^{1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\mathbb{Z}. \tag{C.17}$$

Thus, $\sigma^R_{n,A,A',\bar{\varepsilon}}(B_n, B_n')$ and $\sigma_{n,A,A',\bar{\varepsilon}}(B_n, B_n')$ are "regularized" versions of $\sigma^R_{n,A,A'}(B_n, B_n')$ and $\sigma_{n,A,A'}(B_n, B_n')$. We also define

$$\tau_{n,A,A',\bar{\varepsilon}}(B_n, B_n') \equiv \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_{n,\tau_1}}\int_{\mathcal{U}} g_{2n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) w_{\tau_1,B_n}(x) w_{\tau_2,B_n'}(x+uh) du dx d\tau_1 d\tau_2.$$

Then it suffices for the lemma to show the following two statements.

Step 1: As $n \to \infty$,

$$\begin{aligned}\sup_{P \in \mathcal{P}} \left|\sigma^R_{n,A,A',\bar{\varepsilon}}(B_n, B'_n) - \tau_{n,A,A',\bar{\varepsilon}}(B_n, B'_n)\right| &\to 0, \text{ and} \\ \sup_{P \in \mathcal{P}} \left|\tau_{n,A,A',\bar{\varepsilon}}(B_n, B'_n) - \sigma_{n,A,A',\bar{\varepsilon}}(B_n, B'_n)\right| &\to 0.\end{aligned}$$

Step 2: For some $C > 0$ that does not depend on $\bar{\varepsilon}$ or n,

$$\begin{aligned}\sup_{P \in \mathcal{P}} |\sigma^R_{n,A,A',\bar{\varepsilon}}(B_n, B'_n) - \sigma^R_{n,A,A'}(B_n, B'_n)| &\leq C\sqrt{\bar{\varepsilon}}, \text{ and} \\ \sup_{P \in \mathcal{P}} |\sigma_{n,A,A',\bar{\varepsilon}}(B_n, B'_n) - \sigma_{n,A,A'}(B_n, B'_n)| &\leq C\sqrt{\bar{\varepsilon}}.\end{aligned}$$

Then the desired result follows by sending $n \to \infty$ and then $\bar{\varepsilon} \downarrow 0$, while chaining Steps 1 and 2.

Proof of Step 1: We first focus on the first statement. For any vector $\mathbf{v} = [\mathbf{v}_1^\top, \mathbf{v}_2^\top]^\top \in \mathbf{R}^{2J}$, we define

$$\begin{aligned}\tilde{\Lambda}_{A,p,1}(\mathbf{v}) &\equiv \Lambda_{A,p}\left(\left[\tilde{\Sigma}^{1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\mathbf{v}\right]_1\right), \\ \tilde{\Lambda}_{A',p,2}(\mathbf{v}) &\equiv \Lambda_{A',p}\left(\left[\tilde{\Sigma}^{1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\mathbf{v}\right]_2\right),\end{aligned}$$

and

$$C_{n,p}(\mathbf{v}) \equiv \tilde{\Lambda}_{A,p,1}(\mathbf{v})\,\tilde{\Lambda}_{A',p,2}(\mathbf{v}), \tag{C.18}$$

where $[a]_1$ of a vector $a \in \mathbf{R}^{2J}$ indicates the vector of the first J entries of a, and $[a]_2$ the vector of the remaining J entries of a. By Theorem 9 of Magnus and Neudecker (2001, p. 208),

$$\begin{aligned}\lambda_{\min}\left(\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\right) &\geq \lambda_{\min}\left(\left[\begin{array}{cc} \Sigma_{n,\tau_1,\tau_2}(x,0) & \Sigma_{n,\tau_1,\tau_2}(x,u) \\ \Sigma^\top_{n,\tau_1,\tau_2}(x,u) & \Sigma_{n,\tau_2,\tau_2}(x+uh,0)\end{array}\right]\right) \\ &\quad + \lambda_{\min}\left(\left[\begin{array}{cc} \bar{\varepsilon}I_J & 0 \\ 0 & \bar{\varepsilon}I_J\end{array}\right]\right) \\ &\geq \lambda_{\min}\left(\left[\begin{array}{cc} \bar{\varepsilon}I_J & 0 \\ 0 & \bar{\varepsilon}I_J\end{array}\right]\right) = \bar{\varepsilon}.\end{aligned} \tag{C.19}$$

Let $q_{n,\tau,j}(x;\eta_{1j}) \equiv p_{n,\tau,j}(x) + \eta_{1j}$, where

$$p_{n,\tau,j}(x) \equiv \frac{1}{\sqrt{h^d}} \sum_{1 \leq i \leq N_1} \left\{\beta_{n,x,\tau,j}\left(Y_{ij}, \frac{X_i - x}{h}\right) - \mathbf{E}\left[\beta_{n,x,\tau,j}\left(Y_{ij}, \frac{X_i - x}{h}\right)\right]\right\},$$

η_{1j} is the j-th entry of η_1, and N_1 is a Poisson random variable with mean 1 and $((\eta_{1j})_{j\in\mathbb{N}_J}, N_1)$ is independent of $\{(Y_i^\top, X_i^\top)\}_{i=1}^\infty$. Let $p_{n,\tau}(x)$ be the column vector of entries $p_{n,\tau,j}(x)$ with j running in the set $\mathbb{N}_J$. Let $[p_{n,\tau_1}^{(i)}(x), p_{n,\tau_2}^{(i)}(x+uh)]$ be i.i.d. copies of $[p_{n,\tau_1}(x), p_{n,\tau_2}(x+uh)]$ and $\eta_1^{(i)}$ and $\eta_2^{(i)}$ be also i.i.d. copies of η_1 and η_2. Define

$$q_{n,\tau,1}^{(i)}(x) \equiv p_{n,\tau}^{(i)}(x) + \eta_1^{(i)} \text{ and } q_{n,\tau,2}^{(i)}(x+uh) \equiv p_{n,\tau}^{(i)}(x+uh) + \eta_2^{(i)}.$$

Note that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{bmatrix} q_{n,\tau_1,1}^{(i)}(x) \\ q_{n,\tau_2,2}^{(i)}(x+uh) \end{bmatrix} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{bmatrix} p_{n,\tau_1}^{(i)}(x) \\ p_{n,\tau_2}^{(i)}(x+uh) \end{bmatrix} + \frac{1}{\sqrt{n}}\sum_{i=1}^n \begin{bmatrix} \eta_1^{(i)} \\ \eta_2^{(i)} \end{bmatrix}.$$

The last sum has the same distribution as $[\eta_1^\top, \eta_2^\top]^\top$ and the leading sum on the right-hand side has the same distribution as that of $[\mathbf{z}_{N,\tau_1}^\top(x), \mathbf{z}_{N,\tau_2}^\top(x+uh)]^\top$. Therefore, we conclude that

$$\xi_{N,\tau_1,\tau_2}(x,u;\eta_1,\eta_2) \overset{d}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{W}_{n,\tau_1,\tau_2}^{(i)}(x,u),$$

where

$$\tilde{W}_{n,\tau_1,\tau_2}^{(i)}(x,u) \equiv \tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^{-1/2}(x,u) \begin{bmatrix} q_{n,\tau_1,1}^{(i)}(x) \\ q_{n,\tau_2,2}^{(i)}(x+uh) \end{bmatrix}.$$

Now we invoke the Berry-Esseen-type bound of Sweeting (1977, Theorem 1) to prove Step 1. By Lemma C5, we deduce that

$$\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}||q_{n,\tau,1}^{(i)}(x)||^3 \le Ch^{-d/2}, \tag{C.20}$$

for some $C > 0$. Also, recall the definition of $\rho_{n,\tau_1,\tau_1,j,j}(x,0)$ in (6.7) and note that

$$\begin{aligned} &\sup_{\tau\in\mathcal{T}} \sup_{(x,u)\in\mathcal{S}_\tau(\varepsilon)\times\mathcal{U}} \sup_{P\in\mathcal{P}} tr\left(\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\right) \\ \le\ & \sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)} \sup_{P\in\mathcal{P}} \sum_{j\in J} \left(\rho_{n,\tau_1,\tau_1,j,j}(x,0) + \rho_{n,\tau_2,\tau_2,j,j}(x,0) + 2\bar{\varepsilon}\right) \le C, \end{aligned} \tag{C.21}$$

for some $C > 0$ that depends only on J and $\bar{\varepsilon}$ by Lemma C4. Observe that by the definition of $C_{n,p}$ in (C.18), and (C.21),

$$\sup_{\mathbf{v}\in\mathbf{R}^{2J}} \frac{|C_{n,p}(\mathbf{v}) - C_{n,p}(0)|}{1+||\mathbf{v}||^{2p+2}\min\{||\mathbf{v}||,1\}} \le C.$$

We find that for each $u \in \mathcal{U}$, $||\tilde{W}^{(i)}_{n,\tau_1,\tau_2}(x,u)||^2$ is equal to

$$tr\left(\tilde{\Sigma}^{-1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\left[\begin{array}{c} q^{(i)}_{n,\tau_1,1}(x) \\ q^{(i)}_{n,\tau_1,2}(x+uh)\end{array}\right]\left[\begin{array}{c} q^{(i)}_{n,\tau_2,1}(x) \\ q^{(i)}_{n,\tau_2,2}(x+uh)\end{array}\right]^\top \tilde{\Sigma}^{-1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\right) \tag{C.22}$$
$$\leq \lambda_{\max}\left(\tilde{\Sigma}^{-1}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\right) tr\left(\left[\begin{array}{c} q^{(i)}_{n,\tau_1,1}(x) \\ q^{(i)}_{n,\tau_1,2}(x+uh)\end{array}\right]\left[\begin{array}{c} q^{(i)}_{n,\tau_2,1}(x) \\ q^{(i)}_{n,\tau_2,2}(x+uh)\end{array}\right]^\top\right).$$

Therefore, $\mathbf{E}||\tilde{W}^{(i)}_{n,\tau_1,\tau_2}(x,u)||^3$ is bounded by

$$\lambda^{3/2}_{\max}\left(\tilde{\Sigma}^{-1}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\right)\mathbf{E}\left\|\left[\begin{array}{c} q^{(i)}_{n,\tau_1,1}(x) \\ q^{(i)}_{n,\tau_2,2}(x+uh)\end{array}\right]\right\|^3.$$

From (C.19),

$$\lambda^{3/2}_{\max}(\tilde{\Sigma}^{-1}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)) = \lambda^{-3/2}_{\min}(\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)) \leq \bar{\varepsilon}^{-3/2}.$$

Therefore, we conclude that

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}}\sup_{(x,u)\in\mathcal{S}_\tau(\varepsilon)\times\mathcal{U}}\sup_{P\in\mathcal{P}}\mathbf{E}||\tilde{W}^{(i)}_{n,\tau_1,\tau_2}(x,u)||^3 \\
\leq\ & C_1\bar{\varepsilon}^{-3/2}\cdot\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\sup_{P\in\mathcal{P}}\mathbf{E}||q^{(i)}_{n,\tau_1,1}(x)||^3 \\
&+C_1\bar{\varepsilon}^{-3/2}\cdot\sup_{\tau\in\mathcal{T}}\sup_{(x,u)\in\mathcal{S}_\tau(\varepsilon)\times\mathcal{U}}\sup_{P\in\mathcal{P}}\mathbf{E}||q^{(i)}_{n,\tau_2,2}(x+uh)||^3 \leq C_2\bar{\varepsilon}^{-3/2}/\sqrt{h^d},
\end{aligned}$$

where $C_1 > 0$ and $C_2 > 0$ are constants depending only on J, and the last bound follows by (C.20). Therefore, by Theorem 1 of Sweeting (1977), we find that with $\bar{\varepsilon} > 0$ fixed and $n \to \infty$,

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}}\sup_{(x,u)\in\mathcal{S}_\tau(\varepsilon)\times\mathcal{U}}\sup_{P\in\mathcal{P}}\left|\mathbf{E}C_{n,p}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{W}^{(i)}_{n,\tau_1,\tau_2}(x,u)\right)-\mathbf{E}C_{n,p}\left(\tilde{\mathbb{Z}}_{n,\tau_1,\tau_2}(x,u)\right)\right| \\
&= O\left(n^{-1/2}h^{-d/2}\right) = o(1),
\end{aligned} \tag{C.23}$$

where $\tilde{\mathbb{Z}}_{n,\tau_1,\tau_2}(x,u) = [\mathbb{Z}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x)^\top, \mathbb{Z}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x+uh)^\top]^\top$.

Using similar arguments, we also deduce that for $j = 1, 2$, and $A \subset \mathbb{N}_J$,

$$\sup_{\tau\in\mathcal{T}}\sup_{(x,u)\in\mathcal{S}_\tau(\varepsilon)\times\mathcal{U}}\sup_{P\in\mathcal{P}}\left|\mathbf{E}\tilde{\Lambda}_{A,p,j}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\tilde{W}^{(i)}_{n,\tau_1,\tau_2}(x,u)\right)-\mathbf{E}\tilde{\Lambda}_{A,p,j}\left(\tilde{\mathbb{Z}}_{n,\tau_1,\tau_2}(x,u)\right)\right| = o(1).$$

For some $C > 0$,

$$\begin{aligned}
&\sup_{\tau\in\mathcal{T}}\sup_{(x,u)\in\mathcal{S}_\tau(\varepsilon)\times\mathcal{U}}\sup_{P\in\mathcal{P}} Cov\left(\Lambda_p(\mathbb{Z}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x)), \Lambda_p(\mathbb{Z}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x+uh))\right) \\
\leq\ & \sup_{\tau\in\mathcal{T}}\sup_{(x,u)\in\mathcal{S}_\tau(\varepsilon)\times\mathcal{U}}\sup_{P\in\mathcal{P}}\sqrt{\mathbf{E}\left\|\mathbb{Z}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x)\right\|^{2p}}\sqrt{\mathbf{E}\left\|\mathbb{Z}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x+uh)\right\|^{2p}} < C.
\end{aligned}$$

The last inequality follows because $\mathbb{Z}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x)$ and $\mathbb{Z}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x+uh)$ are centered normal random vectors with a covariance matrix that has a finite Euclidean norm by Lemma C4. Hence we apply the Dominated Convergence Theorem to deduce the first statement of Step 1 from (C.23).

We turn to the second statement of Step 1. The statement immediately follows because for each $u \in \mathcal{U}$, the covariance matrix of $\tilde{\Sigma}^{-1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\xi_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)$ is equal to the covariance matrix of $[\mathbb{W}^{(1)\top}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u), \mathbb{W}^{(2)\top}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)]^\top and$

$$\left|w_{\tau_1,B_n}(x)w_{\tau_2,B_n'}(x+uh) - w_{\tau_1,B_n}(x)w_{\tau_2,B_n'}(x)\right| \to 0,$$

as $n \to \infty$, for each $u \in \mathcal{U}$, and for almost every $x \in \mathcal{X}$ (with respect to Lebesgue measure.)

Proof of Step 2: We consider the first statement. First, we write

$$\begin{aligned}(\text{C.24}) \qquad & \left|\left(\sigma^R_{n,A,A',\bar{\varepsilon}}(B_n,B_n')\right)^2 - \left(\sigma^R_{n,A,A'}(B_n,B_n')\right)^2\right| \\ \leq\ & \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_n}\int_{\mathcal{U}} \left|\Delta^\eta_{n,\tau_1,\tau_2,1}(x,u)\right| w_{\tau_1,B_n}(x)w_{\tau_2,B_n'}(x+uh)dudxd\tau_1 d\tau_2 \\ & + \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_n}\int_{\mathcal{U}} \left|\Delta^\eta_{n,\tau_1,\tau_2,2}(x,u)\right| w_{\tau_1,B_n}(x)w_{\tau_2,B_n'}(x+uh)dudxd\tau_1 d\tau_2,\end{aligned}$$

where

$$\begin{aligned}\Delta^\eta_{n,\tau_1,\tau_2,1}(x,u) &= \mathbf{E}\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_1}(x))\mathbf{E}\Lambda_{A',p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_2}(x+uh)) \\ &\quad -\mathbf{E}\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_1}(x;\eta_1))\mathbf{E}\Lambda_{A',p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_2}(x+uh;\eta_2)),\end{aligned}$$

and

$$\begin{aligned}\Delta^\eta_{n,\tau_1,\tau_2,2}(x,u) &= \mathbf{E}\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_1}(x))\Lambda_{A',p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_2}(x+uh)) \\ &\quad -\mathbf{E}\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_1}(x;\eta_1))\Lambda_{A',p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_2}(x+uh;\eta_2)).\end{aligned}$$

By Hölder inequality, for $C>0$ that depends only on P,

$$\left|\Delta^\eta_{n,\tau_1,\tau_2,2}(x,u)\right| \leq CA_{1n}(x,u) + CA_{2n}(x,u),$$

where, if $p=1$ then we set $s=2$, and $q=1$, and if $p>1$, we set $s=(p+1)/(p-1)$ and $q=(1-1/s)^{-1}$,

$$\begin{aligned}A_{1n}(x,u) &= (nh^d)^p \left\{\mathbf{E}\left\|\mathbf{z}_{N,\tau_1}(x) - \mathbf{z}_{N,\tau_1}(x;\eta_1)\right\|^{2q}\right\}^{\frac{1}{2q}} \\ &\quad \times\left(\left\{\mathbf{E}\left\|\mathbf{z}_{N,\tau_1}(x)\right\|^{2s(p-1)}\right\}^{\frac{1}{2s}} + \left\{\mathbf{E}\left\|\mathbf{z}_{N,\tau_1}(x;\eta_1)\right\|^{2s(p-1)}\right\}^{\frac{1}{2s}}\right) \\ &\quad \times\sqrt{\mathbf{E}\left(\left\|\mathbf{z}_{N,\tau_2}(x+uh)\right\|^{2p}\right)},\end{aligned}$$

and

$$\begin{aligned} A_{2n}(x,u) &= (nh^d)^p \left\{ \mathbf{E} \left\| \mathbf{z}_{N,\tau_2}(x+uh) - \mathbf{z}_{N,\tau_2}(x+uh;\eta_2) \right\|^{2q} \right\}^{\frac{1}{2q}} \\ &\quad \times \left(\left\{ \mathbf{E} \left\| \mathbf{z}_{N,\tau_2}(x+uh) \right\|^{2s(p-1)} \right\}^{\frac{1}{2s}} + \left\{ \mathbf{E} \left\| \mathbf{z}_{N,\tau_2}(x+uh;\eta_2) \right\|^{2s(p-1)} \right\}^{\frac{1}{2s}} \right) \\ &\quad \times \sqrt{\mathbf{E} \left(\left\| \mathbf{z}_{N,\tau_1}(x;\eta_1) \right\|^{2p} \right)}. \end{aligned}$$

Now,

$$\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E} \left\| \sqrt{nh^d} \{ \mathbf{z}_{N,\tau}(x) - \mathbf{z}_{N,\tau}(x;\eta_1) \} \right\|^{2q} = \mathbf{E} \left\| \sqrt{\bar{\varepsilon}} \mathbb{Z} \right\|^{2q} = C\bar{\varepsilon}^q,$$

where $\mathbb{Z} \in \mathbf{R}^J$ is a centered normal random vector with identity covariance matrix I_J. Also, we deduce that for some $C > 0$,

$$\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E} \left\| \sqrt{nh^d} \mathbf{z}_{N,\tau}(x) \right\|^{2s(p-1)} \leq C,$$

by (C.12) of Lemma C5 and by the fact that $2s(p-1) = 2(p+1) \leq M$. Similarly, from some large n on,

$$\begin{aligned} &\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E} \left(\left\| \sqrt{nh^d} \mathbf{z}_{N,\tau}(x+uh;\eta_2) \right\|^{2p} \right) \\ \leq &\sup_{\tau\in\mathcal{T}, x\in\mathcal{S}_\tau(\varepsilon)} \sup_{P\in\mathcal{P}} \mathbf{E} \left(\left\| \sqrt{nh^d} \mathbf{z}_{N,\tau}(x;\eta_2) \right\|^{2p} \right) < C, \end{aligned}$$

for some $C > 0$. Thus we conclude that for some $C > 0$,

$$\sup_{(\tau_1,\tau_2)\in\mathcal{T}\times\mathcal{T}} \sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \left(A_{1n}(x,u) + A_{2n}(x,u) \right) \leq C\sqrt{\bar{\varepsilon}},$$

and that for some $C > 0$,

$$\sup_{(\tau_1,\tau_2)\in\mathcal{T}\times\mathcal{T}} \sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \left| \Delta^{\eta}_{n,\tau_1,\tau_2,2}(x,u) \right| \leq C\sqrt{\bar{\varepsilon}}.$$

Using similar arguments, we also find that for some $C > 0$,

$$\sup_{(\tau_1,\tau_2)\in\mathcal{T}\times\mathcal{T}} \sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \left| \Delta^{\eta}_{n,\tau_1,\tau_2,1}(x,u) \right| \leq C\sqrt{\bar{\varepsilon}}.$$

Therefore, there exist $C_1 > 0$ and $C_2 > 0$ such that from some large n on,

$$\begin{aligned} &\sup_{P\in\mathcal{P}} \left| \sigma^2_{n,A,A',\bar{\varepsilon}}(B_n, B_n') - \sigma^2_{n,A,A'}(B_n, B_n') \right| \\ \leq &\; C_1\sqrt{\bar{\varepsilon}} \int_{\mathcal{T}} \int_{\mathcal{T}} \int_{B_n} \int_{\mathcal{U}} w_{\tau_1,B_n}(x) w_{\tau_2,B_n'}(x+uh) du dx d\tau_1 d\tau_2. \end{aligned}$$

Since the last multiple integral is finite, we obtain the first statement of Step 2.

We turn to the second statement of Step 2. Similarly as before, we write

$$
\begin{aligned}
&\left|\sigma^2_{n,A,A',\bar{\varepsilon}}(B_n, B_n') - \sigma^2_{n,A,A'}(B_n, B_n')\right| \\
\leq\ & \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_n}\int_{\mathcal{U}} \left|\Delta^{\eta}_{1,\tau_1,\tau_2}(x,u)\right| w_{\tau_1,B_n}(x) w_{\tau_2,B_n'}(x+uh) du dx d\tau_1 d\tau_2 \\
&+ \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_n}\int_{\mathcal{U}} \left|\Delta^{\eta}_{2,\tau_1,\tau_2}(x,u)\right| w_{\tau_1,B_n}(x) w_{\tau_2,B_n'}(x+uh) du dx d\tau_1 d\tau_2,
\end{aligned}
$$

where

$$
\begin{aligned}
\Delta^{\eta}_{1,\tau_1,\tau_2}(x,u) \ =\ & \mathbf{E}\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau_1,\tau_2}(x,u))\mathbf{E}\Lambda_{A',p}(\mathbb{W}^{(2)}_{n,\tau_1,\tau_2}(x,u)) \\
& - \mathbf{E}\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u))\mathbf{E}\Lambda_{A',p}(\mathbb{W}^{(2)}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)),
\end{aligned}
$$

and

$$
\begin{aligned}
\Delta^{\eta}_{2,\tau_1,\tau_2}(x,u) \ =\ & \mathbf{E}\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau_1,\tau_2}(x,u))\Lambda_{A',p}(\mathbb{W}^{(2)}_{n,\tau_1,\tau_2}(x,u)) \\
& - \mathbf{E}\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u))\Lambda_{A',p}(\mathbb{W}^{(2)}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)).
\end{aligned}
$$

Now, observe that for $C > 0$ that does not depend on $\bar{\varepsilon}$, we have by Lemma C1(i),

$$
\sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \left\|\tilde{\Sigma}^{1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) - \begin{bmatrix} \Sigma_{n,\tau_1}(x,0) & \Sigma_{n,\tau_1,\tau_2}(x,u) \\ \Sigma_{n,\tau_1,\tau_2}(x,u) & \Sigma_{n,\tau_2}(x+uh) \end{bmatrix}^{1/2}\right\| \leq C\sqrt{\bar{\varepsilon}}.
$$

Using this, recalling the definitions of $\mathbb{W}^{(1)}_{n,\tau_1,\tau_2}(x,u)$ and $\mathbb{W}^{(2)}_{n,\tau_1,\tau_2}(x,u)$ in (C.17), and following the previous arguments, we obtain the second statement of Step 2. ∎

Lemma C7. *Suppose that for some small $\nu_1 > 0$, $n^{-1/2}h^{-d-\nu_1} \to 0$, as $n \to \infty$ and the conditions of Lemma C6 hold. Then there exists $C > 0$ such that for any sequence of Borel sets $B_n \subset \mathcal{S}$, and $A \subset \mathbb{N}_J$, from some large n on,*

$$
\begin{aligned}
& \sup_{P\in\mathcal{P}} \mathbf{E}\left[\left|h^{-d/2}\int_{B_n}\left\{\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{n,\tau}(x)) - \mathbf{E}\left[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x))\right]\right\} dQ(x,\tau)\right|\right] \\
\leq\ & C\sqrt{Q(B_n)}.
\end{aligned}
$$

Remark 4. The result is in the same spirit as Lemma 6.2 of Giné, Mason, and Zaitsev (2003). (Also see Lemma A8 of Lee, Song and Whang (2013).) However, unlike these results, the location normalization here involves $\mathbf{E}[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x))]$ instead of $\mathbf{E}[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{n,\tau}(x))]$. We can obtain the same result with $\mathbf{E}[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x))]$ replaced by $\mathbf{E}[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{n,\tau}(x))]$, but with a stronger bandwidth condition.

Like Lemma C6, the result of Lemma C7 does not require that the quantities $\sqrt{nh^d}\mathbf{z}_{n,\tau}(x)$ and $\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)$ have a (pointwise in x) nondegenerate limit distribution.

Proof of Lemma C7. As in the proof of Lemma A8 of Lee, Song, and Whang (2013), it suffices to show that there exists $C > 0$ such that C does not depend on n and for any Borel set $B \subset \mathbf{R}$,

Step 1:

$$\sup_{P \in \mathcal{P}} \mathbf{E}\left[\left| h^{-d/2} \int_{B_n} \left\{ \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{n,\tau}(x)) - \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)) \right\} dQ(x,\tau) \right|\right] \leq CQ(B_n), \text{ and}$$

Step 2:

$$\sup_{P \in \mathcal{P}} \mathbf{E}\left[\left| h^{-d/2} \int_{B_n} \left\{ \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)) - \mathbf{E}\left[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x))\right] \right\} dQ(x,\tau) \right|\right] \leq C\sqrt{Q(B_n)}.$$

By chaining Steps 1 and 2, we obtain the desired result.

Proof of Step 1: Similarly as in (2.13) of Horváth (1991), we first write

$$\mathbf{z}_{n,\tau}(x) = \mathbf{z}_{N,\tau}(x) + \mathbf{v}_{n,\tau}(x) + \mathbf{s}_{n,\tau}(x), \tag{C.25}$$

where, for $\beta_{n,x,\tau}(Y_i, (X_i - x)/h)$ defined prior to Lemma C5,

$$\begin{aligned} \mathbf{v}_{n,\tau}(x) &\equiv \left(\frac{n-N}{n}\right) \cdot \frac{1}{h^d}\mathbf{E}\left[\beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right)\right] \text{ and} \\ \mathbf{s}_{n,\tau}(x) &\equiv \frac{1}{nh^d}\sum_{i=N+1}^{n}\left\{\beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right) - \mathbf{E}\left[\beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right)\right]\right\}, \end{aligned}$$

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

Using (C.25), we deduce that for some $C_1, C_2 > 0$ that depend only on P,

$$\begin{aligned} &\int_{B_n} \left|\Lambda_{A,p}\left(\mathbf{z}_{n,\tau}(x)\right) - \Lambda_{A,p}\left(\mathbf{z}_{N,\tau}(x)\right)\right| dQ(x,\tau) \\ \leq\ & C_1 \int_{B_n} \|\mathbf{v}_{n,\tau}(x)\| \left(\|\mathbf{z}_{n,\tau}(x)\|^{p-1} + \|\mathbf{z}_{N,\tau}(x)\|^{p-1}\right) dQ(x,\tau) \\ & + C_2 \int_{B_n} \|\mathbf{s}_{n,\tau}(x)\| \left(\|\mathbf{z}_{n,\tau}(x)\|^{p-1} + \|\mathbf{z}_{N,\tau}(x)\|^{p-1}\right) dQ(x,\tau) \\ \equiv\ & D_{1n} + D_{2n}, \text{ say.} \end{aligned} \tag{C.26}$$

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

Claim 1: $\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}[||\mathbf{v}_{n,\tau}(x)||^2] = O(n^{-1})$, and

Claim 2: $\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}[||\mathbf{s}_{n,\tau}(x)||^2] = O(n^{-3/2}h^{-d})$.

Proof of Claim 1: First, note that

$$\sup_{(x,\tau)\in\mathcal{S}} \mathbf{E}\left[||\mathbf{v}_{n,\tau}(x)||^2\right] \leq \mathbf{E}\left|\frac{n-N}{n}\right|^2 \cdot \sup_{(x,\tau)\in\mathcal{S}} \left\|\frac{1}{h^d}\mathbf{E}\left[\beta_{n,x,\tau}\left(Y_i, \frac{X_i - x}{h}\right)\right]\right\|^2.$$

Since $\mathbf{E}|n^{-1/2}(n-N)|^2$ does not depend on the joint distribution of (Y_i, X_i), $\mathbf{E}|n^{-1/2}(n-N)|^2 \leq O(1)$ uniformly over $P \in \mathcal{P}$. Combining this with the second statement of (C.12), the product on the right hand side becomes $O(n^{-1})$ uniformly over $P \in \mathcal{P}$.

Proof of Claim 2: Let $\eta_1 \in \mathbf{R}^J$ be the random vector defined prior to Lemma C6, and define

$$\mathbf{s}_{n,\tau}(x;\eta_1) \equiv \mathbf{s}_{n,\tau}(x) + \frac{(N-n)\eta_1}{n^{3/2}h^{d/2}}.$$

Note that

$$\mathbf{E}\left\|\mathbf{s}_{n,\tau}(x)\right\|^2 \leq 2\mathbf{E}\left\|\mathbf{s}_{n,\tau}(x;\eta_1)\right\|^2 + \frac{2}{n^2h^d}\mathbf{E}\left\|\frac{(N-n)\eta_1}{\sqrt{n}}\right\|^2. \tag{C.27}$$

As for the last term, since N and η_1 are independent, it is bounded by

$$\frac{1}{n^2h^d}\left(\mathbf{E}\left|\frac{N-n}{\sqrt{n}}\right|^2\right)\cdot\mathbf{E}\left\|\eta_1\right\|^2 \leq \frac{C\bar{\varepsilon}}{n^2h^d} = O(n^{-2}h^{-d-\nu_1}),$$

from some large n on.

As for the leading expectation on the right hand side of (C.27), we write

$$\begin{aligned}\mathbf{E}\left\|\sqrt{nh^d}\mathbf{s}_{n,\tau}(x;\eta_1)\right\|^2 &= \mathbf{E}\left\|\frac{1}{\sqrt{n}}\sum_{i=N+1}^{n} q_{n,\tau,1}^{(i)}(x)\right\|^2 \\ &= \frac{1}{n}\sum_{j=1}^{J}\bar{\sigma}_{n,\tau,j}^2(x)\mathbf{E}\left(\sum_{i=N+1}^{n}\frac{q_{n,\tau,1,j}^{(i)}(x)}{\bar{\sigma}_{n,\tau,j}(x)}\right)^2,\end{aligned}$$

where $q_{n,\tau,1}^{(i)}(x)$'s $(i=1,2,\cdots)$ are i.i.d. copies of $q_{n,\tau}(x)+\eta_1$ and $q_{n,\tau,1,j}^{(i)}(x)$ is the j-th entry of $q_{n,\tau,1}^{(i)}(x)$, and $\bar{\sigma}_{n,\tau,j}^2(x) \equiv Var(q_{n,\tau,1,j}^{(i)}(x))$. Recall that $q_{n,\tau}(x)$ was defined prior to Lemma C5. Now we apply Lemma 1(i) of Horváth (1991) to deduce that

$$\begin{aligned}&\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}\left(\sum_{i=N+1}^{n}\frac{q_{n,\tau,1,j}^{(i)}(x)}{\bar{\sigma}_{n,\tau,j}(x)}\right)^2 \\ \leq\ & \mathbf{E}|N-n|\cdot\mathbf{E}|\mathbb{Z}_1|^2 + C\mathbf{E}\left|N-n\right|^{1/2}\cdot\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}\left|\frac{q_{n,\tau,1,j}^{(i)}(x)}{\bar{\sigma}_{n,\tau,j}(x)}\right|^3 \\ & +C\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}\left|\frac{q_{n,\tau,1,j}^{(i)}(x)}{\bar{\sigma}_{n,\tau,j}(x)}\right|^4,\end{aligned}$$

for some $C>0$, where $\mathbb{Z}_1 \sim N(0,1)$.

First, observe that $\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\bar{\sigma}_{n,\tau,j}(x) < \infty$ by Lemma C5, and

$$\inf_{(x,\tau)\in\mathcal{S}}\inf_{P\in\mathcal{P}}\bar{\sigma}_{n,\tau,j}(x) > \bar{\varepsilon} > 0, \tag{C.28}$$

due to the additive term η_1 in $q_{n,\tau}(x)+\eta_1$. Let η_{1j} be the j-th entry of η_1. We apply Lemma C5 to deduce that for some $C>0$, from some large n on,

$$\text{(C.29)} \quad \begin{aligned} \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}|(q_{n,\tau,j}(x)+\eta_{1j})/\bar{\sigma}_{n,\tau,j}(x)|^3 &\leq Ch^{-(d/2)-(\nu_1/2)} \text{ and} \\ \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}|(q_{n,\tau,j}(x)+\eta_{1j})/\bar{\sigma}_{n,\tau,j}(x)|^4 &\leq Ch^{-d-\nu_1}. \end{aligned}$$

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

$$\text{(C.30)} \quad \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left(\sum_{i=N+1}^{n} \frac{q_{n,\tau,1,j}^{(i)}(x)}{\bar{\sigma}_{n,\tau,j}(x)}\right)^2 \leq \frac{C}{\bar{\varepsilon}^4}\left\{n^{1/2}+n^{1/4}h^{-(d/2)-(\nu_1/2)}+h^{-d-\nu_1}\right\}.$$

This implies that for some $C>0$, (with $\bar{\varepsilon}>0$ fixed while $n\to\infty$)

$$\text{(C.31)} \quad \begin{aligned} &\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left\|\sqrt{nh^d}\mathbf{s}_{n,\tau}(x)\right\|^2 \\ \leq\ & O\left(n^{-1}h^{-\nu_1}\right) + O\left(n^{-1/2}+n^{-3/4}h^{-(d/2)-(\nu_1/2)}+n^{-1}h^{-d-\nu_1}\right) \\ =\ & O\left(n^{-1}h^{-\nu_1}\right) + O(n^{-1/2}) = O(n^{-1/2}), \end{aligned}$$

since $n^{-1/2}h^{-d-\nu_1}\to 0$. Hence, we obtain Claim 2.

Using Claim 1 and the second statement of Lemma C5, we deduce that

$$\begin{aligned} \sup_{P\in\mathcal{P}} \mathbf{E}\left[n^{p/2}h^{d(p-1)/2}D_{1n}\right] &\leq C_1 Q(B_n) \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \sqrt{\mathbf{E}\left\|\sqrt{n}\mathbf{v}_{n,\tau}(x)\right\|^2} \\ &\quad \times\sqrt{\mathbf{E}\left\|\sqrt{nh^d}\mathbf{z}_{n,\tau}(x)\right\|^{2p-2}+\mathbf{E}\left\|\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)\right\|^{2p-2}} \\ &\leq C_2 Q(B_n), \end{aligned}$$

for $C_1, C_2>0$. Similarly, we can see that

$$\sup_{P\in\mathcal{P}} \mathbf{E}\left[n^{p/2}h^{d(p-1)/2}D_{2n}\right] = O(n^{-1/2}h^{-d}) = o(1),$$

using Claim 2 and the second statement of Lemma C5. Thus, we obtain Step 1.

Proof of Step 2: We can follow the proof of Lemma C6 to show that

$$\begin{aligned} &\mathbf{E}\left[h^{-d/2}\int_{B_n}\left(\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)) - \mathbf{E}\left[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x))\right]\right)dQ(x,\tau)\right]^2 \\ =\ & \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_{n,\tau_1}\cap B_{n,\tau_2}}\int_{\mathcal{U}} C_{n,\tau_1,\tau_2,A,A'}(x,u)dudxd\tau_1 d\tau_2 + o(1), \end{aligned}$$

where $C_{n,\tau_1,\tau_2,A,A'}(x,u)$ is defined in (C.14) and $o(1)$ is uniform over $P \in \mathcal{P}$. Now, observe that

$$\begin{aligned} & \sup_{(\tau_1,\tau_2)\in\mathcal{T}\times\mathcal{T}} \sup_{u\in\mathcal{U}} \sup_{x\in\mathcal{X}} \sup_{P\in\mathcal{P}} |C_{n,\tau_1,\tau_2,A,A'}(x,u)| \\ \leq & \sup_{(\tau_1,\tau_2)\in\mathcal{T}\times\mathcal{T}} \sup_{u\in\mathcal{U}} \sup_{x\in\mathcal{X}} \sup_{P\in\mathcal{P}} \sqrt{\mathbf{E}||\mathbb{W}_{n,\tau_1,\tau_2}^{(1)}(x,u)||^{2p}\mathbf{E}||\mathbb{W}_{n,\tau_1,\tau_2}^{(2)}(x,u)||^{2p}} < \infty. \end{aligned}$$

Therefore,

$$\begin{aligned} & \mathbf{E}\left[\left|h^{-d/2}\int_{B_n}\left(\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)) - \mathbf{E}\left[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x))\right]\right)dQ(x,\tau)\right|\right] \\ \leq & \sqrt{\int_{\mathcal{T}}\int_{\mathcal{T}}\int_{\mathcal{U}}\int_{B_{n,\tau_1}\cap B_{n,\tau_2}} C_{n,\tau_1,\tau_2,A,A'}(x,u)dxdud\tau_1 d\tau_2} + o(1) \\ \leq & C\sqrt{\int_{\mathcal{T}}\int_{\mathcal{T}}\int_{\mathcal{U}}\int_{B_{n,\tau_1}\cap B_{n,\tau_2}} dxdud\tau_1 d\tau_2} + o(1), \end{aligned}$$

for some $C > 0$. Now, observe that

$$\int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_{n,\tau_1}\cap B_{n,\tau_2}} dxd\tau_1 d\tau_2 \leq \int_{\mathcal{T}} d\tau_2 \cdot \left(\int_{\mathcal{T}}\int_{B_{n,\tau_1}} dxd\tau_1\right) \leq CQ(B_n),$$

because $\mathcal{T}$ is a bounded set. Thus the proof of Step 2 is completed. ∎

The next lemma shows the joint asymptotic normality of a Poissonized version of a normalized test statistic and a Poisson random variable. Using this result, we can apply the de-Poissonization lemma in Lemma C3. To define a Poissonized version of a normalized test statistic, we introduce some notation.

Let $\mathcal{C} \subset \mathbf{R}^d$ be a compact set such that $\mathcal{C}$ does not depend on $P \in \mathcal{P}$ and $\alpha_P \equiv P\{X \in \mathbf{R}^d \backslash \mathcal{C}\}$ satisfies that $0 < \inf_{P\in\mathcal{P}} \alpha_P \leq \sup_{P\in\mathcal{P}} \alpha_P < 1$. Existence of such $\mathcal{C}$ is assumed in Assumption A6(ii). For $c_n \to \infty$, we let $B_{n,A}(c_n;\mathcal{C}) \equiv B_{n,A}(c_n) \cap (\mathcal{C} \times \mathcal{T})$, where we recall the definition of $B_{n,A}(c_n) = B_{n,A}(c_n, c_n)$. (See the definition of $B_{n,A}(c_{n,1}, c_{n,2})$ before Lemma 1.) Define

$$\begin{aligned} \zeta_{n,A} &\equiv \int_{B_{n,A}(c_n;\mathcal{C})} \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{n,\tau}(x))dQ(x,\tau), \text{ and} \\ \zeta_{N,A} &\equiv \int_{B_{n,A}(c_n;\mathcal{C})} \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x))dQ(x,\tau). \end{aligned}$$

Let μ_A's be real numbers indexed by $A \in \mathcal{N}_J$, and define

$$\sigma_n^2(\mathcal{C}) \equiv \sum_{A\in\mathcal{N}_J}\sum_{A'\in\mathcal{N}_J} \mu_A\mu_{A'}\sigma_{n,A,A'}(B_{n,A}(c_n;\mathcal{C}), B_{n,A'}(c_n;\mathcal{C})),$$

where we recall the definition of $\sigma_{n,A,A'}(\cdot,\cdot)$ prior to Lemma C6. Define

$$S_n \equiv h^{-d/2} \sum_{A \in \mathcal{N}_J} \mu_A \left\{ \zeta_{N,A} - \mathbf{E}\zeta_{N,A} \right\}.$$

Also define

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

Let

$$H_n \equiv \left[\frac{S_n}{\sigma_n(\mathcal{C})}, \frac{U_n}{\sqrt{1-\alpha_P}} \right]^\top.$$

The following lemma establishes the joint convergence of H_n. In doing so, we need to be careful in dealing with uniformity in $P \in \mathcal{P}$, and potential degeneracy of the normalized test statistic S_n.

Lemma C8. *Suppose that the conditions of Lemma C6 hold and that $c_n \to \infty$ as $n \to \infty$.*
(i) If $\liminf_{n\to\infty} \inf_{P\in\mathcal{P}} \sigma_n^2(\mathcal{C}) > 0$, then

$$\sup_{P\in\mathcal{P}} \sup_{t\in\mathbf{R}^2} \left| P\left\{H_n \leq t\right\} - P\left\{\mathbb{Z} \leq t\right\} \right| \to 0,$$

where $\mathbb{Z} \sim N(0, I_2)$.
(ii) If $\limsup_{n\to\infty} \sigma_n^2(\mathcal{C}) = 0$, then for each $(t_1, t_2) \in \mathbf{R}^2$,

$$\left| P\left\{ S_n \leq t_1 \text{ and } \frac{U_n}{\sqrt{1-\alpha_P}} \leq t_2 \right\} - 1\{0 \leq t_1\} P\left\{\mathbb{Z}_1 \leq t_2\right\} \right| \to 0,$$

where $\mathbb{Z}_1 \sim N(0,1)$.

Remark 5. The joint convergence result is divided into two separate results. The first case is a situation where S_n is asymptotically nondegenerate uniformly in $P \in \mathcal{P}$. The second case deals with a situation where S_n is asymptotically degenerate for some $P \in \mathcal{P}$.

Proof of Lemma C8. (i) Define $\bar{\varepsilon} > 0$ and let

$$H_{n,\bar{\varepsilon}} \equiv \left[\frac{S_{n,\bar{\varepsilon}}}{\sigma_{n,\bar{\varepsilon}}(\mathcal{C})}, \frac{U_n}{\sqrt{1-\alpha_P}} \right]^\top,$$

where $S_{n,\bar{\varepsilon}}$ is equal to S_n, except that $\zeta_{N,A}$ is replaced by

$$\zeta_{N,A,\bar{\varepsilon}} \equiv \int_{B_{n,A}(c_n;\mathcal{C})} \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x;\eta_1)) dQ(x,\tau),$$

and $\mathbf{z}_{N,\tau}(x;\eta_1)$ is as defined prior to Lemma C6, and $\sigma_{n,\bar{\varepsilon}}(\mathcal{C})$ is $\sigma_n(\mathcal{C})$ except that $\tilde{\Sigma}_{n,\tau_1,\tau_2}(x,u)$ is replaced by $\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)$. Also let

$$C_n \equiv \mathbf{E}H_nH_n^\top \text{ and } C_{n,\bar{\varepsilon}} \equiv \mathbf{E}H_{n,\bar{\varepsilon}}H_{n,\bar{\varepsilon}}^\top.$$

First, we show the following statements.

Step 1: For some $C>0$, $\sup_{P\in\mathcal{P}}|Cov(S_{n,\bar{\varepsilon}}-S_n,U_n)| \leq C\sqrt{\bar{\varepsilon}}$, for each fixed $\bar{\varepsilon}>0$.

Step 2: $\sup_{P\in\mathcal{P}}|Cov(S_{n,\bar{\varepsilon}},U_n)| = o(h^{d/2})$, as $n\to\infty$.

Step 3: There exists $c>0$ such that from some large n on,

$$\inf_{P\in\mathcal{P}}\lambda_{\min}(C_n) > c.$$

Step 4: As $n\to\infty$,

$$\sup_{P\in\mathcal{P}}\sup_{t\in\mathbf{R}^2}\left|P\left\{C_n^{-1/2}H_n \leq t\right\} \to P\left\{\mathbb{Z}\leq t\right\}\right| \to 0.$$

From Steps 1-3, we find that $\sup_{P\in\mathcal{P}}\|C_n - I_2\| \to 0$, as $n\to\infty$ and as $\bar{\varepsilon}\to 0$. By Step 4, we obtain (i) of Lemma C8.

Proof of Step 1: Observe that from an inequality similar to (C.26) in the proof of Lemma C7,

$$|\zeta_{N,A,\bar{\varepsilon}} - \zeta_{N,A}| \leq C||\eta_1||\int_{B_{n,A}(c_n;\mathcal{C})}\left\|\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)\right\|^{p-1}dQ(x,\tau).$$

Using the fact that $\mathcal{S}$ is compact and does not depend on $P\in\mathcal{P}$, for some constants $C_1,C_2,C_3>0$ that do not depend on $P\in\mathcal{P}$,

$$\begin{aligned}\mathbf{E}\left|\zeta_{N,A,\bar{\varepsilon}} - \zeta_{N,A}\right|^2 &\leq C_1\mathbf{E}\left[||\eta_1||^2\right]\cdot\int_{B_{n,A}(c_n;\mathcal{C})}\mathbf{E}\left\|\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)\right\|^{2p-2}dQ(x,\tau)\\ &\leq C_2\bar{\varepsilon}\cdot\int_{B_{n,A}(c_n;\mathcal{C})}\mathbf{E}\left\|\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)\right\|^{2p-2}dQ(x,\tau) \leq C_3\bar{\varepsilon},\end{aligned}$$

by the independence between η_1 and $\{\mathbf{z}_{N,\tau}(x):(x,\tau)\in\mathcal{S}\}$, and by the second statement of Lemma C5. From the fact that

$$\sup_{P\in\mathcal{P}}\mathbf{E}U_n^2 \leq \sup_{P\in\mathcal{P}}(1-\alpha_P) \leq 1,$$

we obtain the desired result.

Proof of Step 2: Let $\Sigma_{2n,\tau,\bar{\varepsilon}}$ be the covariance matrix of $[(q_{n,\tau}(x)+\eta_1)^\top,\tilde{U}_n]^\top$, where

$\tilde{U}_n = U_n/\sqrt{P\{X \in \mathcal{C}\}}$. We can write $\Sigma_{2n,\tau,\bar{\varepsilon}}$ as

$$\begin{aligned}
&\left[\begin{array}{cc} \Sigma_{n,\tau,\tau}(x,0) + \bar{\varepsilon} I_J & \mathbf{E}[(q_{n,\tau}(x)+\eta_1)\tilde{U}_n] \\ \mathbf{E}[(q_{n,\tau}(x)+\eta_1)^\top \tilde{U}_n] & 1 \end{array}\right] \\
= &\left[\begin{array}{cc} \Sigma_{n,\tau,\tau}(x,0) & \sqrt{1-\bar{\varepsilon}}\mathbf{E}[q_{n,\tau}(x)\tilde{U}_n] \\ \sqrt{1-\bar{\varepsilon}}\mathbf{E}[q_{n,\tau}^\top(x)\tilde{U}_n] & 1-\bar{\varepsilon} \end{array}\right] + \left[\begin{array}{cc} \bar{\varepsilon} I_J & \mathbf{0} \\ \mathbf{0}^\top & \bar{\varepsilon} \end{array}\right] + A_{n,\tau}(x),
\end{aligned}$$

where

$$A_{n,\tau}(x) \equiv \left[\begin{array}{cc} \mathbf{0} & \left(1-\sqrt{1-\bar{\varepsilon}}\right)\mathbf{E}[q_{n,\tau}(x)\tilde{U}_n] \\ \left(1-\sqrt{1-\bar{\varepsilon}}\right)\mathbf{E}[q_{n,\tau}^\top(x)\tilde{U}_n] & \mathbf{0} \end{array}\right].$$

The first matrix on the right hand side is certainly positive semidefinite. Note that

$$\left(q_{n,\tau,j}(x), \tilde{U}_n\right) \stackrel{d}{=} \left(\frac{1}{\sqrt{n}}\sum_{k=1}^n q_{n,\tau,j}^{(k)}(x), \frac{1}{\sqrt{n}}\sum_{k=1}^n \tilde{U}_n^{(k)}\right),$$

where $(q_{n,\tau,j}^{(k)}(x), \tilde{U}_n^{(k)})$'s with $k = 1, \cdots, n$ are i.i.d. copies of $(q_{n,\tau,j}(x), \bar{U}_n)$, where

$$\bar{U}_n \equiv \frac{1}{\sqrt{P\{X \in \mathcal{C}\}}}\left\{\sum_{1 \le i \le N_1} 1\{X_i \in \mathcal{C}\} - P\left\{X_i \in \mathcal{C}\right\}\right\},$$

where N_1 is the Poisson random variable with mean 1 that is involved in the definition of $q_{n,\tau,j}(x)$. Hence as for $A_{n,\tau}(x)$, note that for $C_1, C_2 > 0$,

$$\begin{aligned}
\text{(C.32)} \quad \sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\left|\mathbf{E}\left[q_{n,\tau,j}(x)\tilde{U}_n\right]\right| &\le \sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\left|\mathbf{E}\left[q_{n,\tau,j}^{(k)}(x)\tilde{U}_n^{(k)}\right]\right| \\
&\le \sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\frac{\mathbf{E}\left[|q_{n,\tau,j}(x)|\right]}{\sqrt{P\{X_i \in \mathcal{C}\}}} \\
&\le \frac{C_1 h^d \sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}} k_{n,\tau,j,1}}{h^{d/2}\left(1-\alpha_P\right)} \le C_2 h^{d/2}.
\end{aligned}$$

We conclude that

$$\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}||A_{n,\tau}(x)|| = O(h^{d/2}).$$

Therefore, from some large n on,

$$\text{(C.33)} \quad \inf_{(x,\tau)\in\mathcal{S}}\inf_{P\in\mathcal{P}}\lambda_{\min}\left(\Sigma_{2n,\tau,\bar{\varepsilon}}\right) \ge \bar{\varepsilon}/2.$$

Let

$$W_{n,\tau}(x;\eta_1) \equiv \Sigma_{2n,\tau,\bar{\varepsilon}}^{-1/2}\left[\begin{array}{c} q_{n,\tau}(x)+\eta_1 \\ \tilde{U}_n \end{array}\right].$$

Similarly as in (C.22), we find that for some $C>0$, from some large n on,

$$\begin{aligned}&\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}\left\|W_{n,\tau}(x;\eta_1)\right\|^3\\ \le\ & C\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\lambda_{\max}^{3/2}\left(\Sigma_{2n,\tau,\bar{\varepsilon}}^{-1}\right)\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\left\{\mathbf{E}\left[||q_{n,\tau}(x)+\eta_1||^3\right]+\mathbf{E}\left[|\tilde{U}_n|^3\right]\right\}\\ \le\ & C\left(\frac{\bar{\varepsilon}}{2}\right)^{-3/2}\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\left\{\mathbf{E}\left[||q_{n,\tau}(x)+\eta_1||^3\right]+\mathbf{E}\left[|\tilde{U}_n|^3\right]\right\},\end{aligned}$$

where the last inequality uses (C.33). As for the last expectation, note that by Rosenthal's inequality, we have

$$\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}\left[|\tilde{U}_n|^3\right]\le C$$

for some $C>0$. We apply the first statement of Lemma C5 to conclude that

$$\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}\left\|W_{n,\tau}(x;\eta_1)\right\|^3\le C\bar{\varepsilon}^{-3/2}h^{-d/2},$$

for some $C>0$. For any vector $\mathbf{v}=[\mathbf{v}_1^\top,v_2]^\top\in\mathbf{R}^{J+1}$, we define

$$D_{n,\tau,p}(\mathbf{v})\equiv\Lambda_{A,p}\left(\left[\Sigma_{2n,\tau,\bar{\varepsilon}}^{1/2}\mathbf{v}\right]_1\right)\left[\Sigma_{2n,\tau,\bar{\varepsilon}}^{1/2}\mathbf{v}\right]_2,$$

where $[a]_1$ of a vector $a\in\mathbf{R}^{J+1}$ indicates the vector of the first J entries of a, and $[a]_2$ the last entry of a. By Theorem 1 of Sweeting (1977), we find that (with $\bar{\varepsilon}>0$ fixed)

$$\mathbf{E}\left[D_{n,\tau,p}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{n,\tau}^{(i)}(x;\eta_1)\right)\right]=\mathbf{E}\left[D_{n,\tau,p}\left(\mathbb{Z}_{J+1}\right)\right]+O(n^{-1/2}h^{-d/2}),$$

where $\mathbb{Z}_{J+1}\sim N(0,I_{J+1})$ and $W_{n,\tau}^{(i)}(x;\eta_1)$'s are i.i.d. copies of $W_{n,\tau}(x;\eta_1)$. Since $O(n^{-1/2}h^{-d/2})=o(h^{d/2})$ (by the condition that $n^{-1/2}h^{-d-\nu}\to 0$, as $n\to\infty$),

$$Cov\left(\Lambda_{A,p}\left(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x;\eta_1)\right),U_n\right)=\mathbf{E}\left[D_{n,\tau,p}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{n,\tau}^{(i)}(x;\eta_1)\right)\right]+o(h^{d/2}).$$

Noting that $\mathbf{E}[D_{n,\tau,p}\left(\mathbb{Z}_{J+1}\right)]=0$, we conclude that

$$\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\left|Cov\left(\Lambda_{A,p}\left(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x;\eta_1)\right),U_n\right)\right|=o(h^{d/2}).$$

By applying the Dominated Convergence Theorem, we obtain Step 2.

Proof of Step 3: First, we show that

(C.34) $$Var\left(S_n\right)=\sigma_n^2(\mathcal{C})+o(1),$$

where $o(1)$ is an asymptotically negligible term uniformly over $P \in \mathcal{P}$. Note that

$$Var\,(S_n) = \sum_{A \in \mathcal{N}_J} \sum_{A' \in \mathcal{N}_J} \mu_A \mu_{A'} Cov(\psi_{n,A}, \psi_{n,A'}),$$

where $\psi_{n,A} \equiv h^{-d/2}(\zeta_{N,A} - \mathbf{E}\zeta_{N,A})$. By Lemma C6, we find that for $A, A' \in \mathcal{N}_J$,

$$Cov(\psi_{n,A}, \psi_{n,A'}) = \sigma_{n,A,A'}(B_{n,A}(c_n;\mathcal{C}), B_{n,A'}(c_n;\mathcal{C})) + o(1),$$

uniformly in $P \in \mathcal{P}$, yielding the desired result.

Combining Steps 1 and 2, we deduce that

$$\sup_{P \in \mathcal{P}} |Cov(S_n, U_n)| \le C\sqrt{\bar{\varepsilon}} + o(h^{d/2}). \tag{C.35}$$

Let $\bar{\sigma}_1^2 \equiv \inf_{P \in \mathcal{P}} \sigma_n^2(\mathcal{C})$ and $\bar{\sigma}_2^2 \equiv \inf_{P \in \mathcal{P}}(1 - \alpha_P)$. Note that for some $C_1 > 0$,

$$\inf_{P \in \mathcal{P}} \bar{\sigma}_1^2 \bar{\sigma}_2^2 > C_1, \tag{C.36}$$

by the condition of the lemma. A simple calculation gives us

$$\begin{aligned} \lambda_{\min}(C_n) &= \frac{\bar{\sigma}_1^2 + \bar{\sigma}_2^2}{2} - \frac{1}{2}\left(\sqrt{\left(\bar{\sigma}_1^2 + \bar{\sigma}_2^2\right)^2 - 4\left\{\bar{\sigma}_1^2\bar{\sigma}_2^2 - Cov(S_n, U_n)^2\right\}}\right) \\ &\ge \frac{1}{2}\left\{\sqrt{\left(\bar{\sigma}_1^2 + \bar{\sigma}_2^2\right)^2} - \left(\sqrt{\left(\bar{\sigma}_1^2 + \bar{\sigma}_2^2\right)^2 - 4\bar{\sigma}_1^2\bar{\sigma}_2^2}\right)\right\} - |Cov(S_n, U_n)| \\ &\ge \bar{\sigma}_1^2\bar{\sigma}_2^2 - |Cov(S_n, U_n)| \ge C_1 - C\sqrt{\bar{\varepsilon}} + o(h^{d/2}), \end{aligned} \tag{C.37}$$

where the last inequality follows by (C.35) and (C.36). Taking $\bar{\varepsilon}$ small enough, we obtain the desired result.

Proof of Step 4: Suppose that $\liminf_{n\to\infty} \inf_{P \in \mathcal{P}} \sigma_n^2(\mathcal{C}) > 0$. Let κ be the diameter of the compact set $\mathcal{K}_0$ introduced in Assumption A2. Let $\mathcal{C}$ be given as in the lemma. Let $\mathbb{Z}^d$ be the set of d-tuples 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 $R_{n,\mathbf{i}} = [a_{n,\mathbf{i}_1}, b_{n,\mathbf{i}_1}] \times \cdots \times [a_{n,\mathbf{i}_d}, b_{n,\mathbf{i}_d}]$, where $\mathbf{i}_j$ is the j-th entry of $\mathbf{i}$, and $h\kappa \le b_{n,\mathbf{i}_j} - a_{n,\mathbf{i}_j} \le 2h\kappa$, for all $j = 1, \cdots, d$, 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 $\mathcal{X}$, 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

$$\begin{aligned} B_{n,A,x}(c_n) &\equiv \{\tau \in \mathcal{T} : (x, \tau) \in B_A(c_n)\}, \\ B_{n,\mathbf{i}} &\equiv R_{n,\mathbf{i}} \cap \mathcal{C}, \end{aligned}$$

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 $C_1 h^d$ and the cardinality of the set $\mathcal{I}_n$ is bounded by $C_2 h^{-d}$ for some positive constants C_1 and

C_2. Now let us define

$$\Delta_{n,A,\mathbf{i}} \equiv h^{-d/2} \int_{B_{n,\mathbf{i}}} \int_{B_{n,A,x}(c_n)} \left\{ \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)) - \mathbf{E}\left[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)) \right] \right\} d\tau dx.$$

And also define $B_{n,A,\mathbf{i}}(c_n) \equiv (B_{n,\mathbf{i}} \times \mathcal{T}) \cap B_{n,A}(c_n)$,

$$\begin{aligned} \alpha_{n,\mathbf{i}} &\equiv \frac{\sum_{A \in \mathcal{N}_J} \mu_A \Delta_{n,A,\mathbf{i}}}{\sigma_n(\mathcal{C})} \text{ and} \\ u_{n,\mathbf{i}} &\equiv \frac{1}{\sqrt{n}} \left\{ \sum_{i=1}^{N} 1\{X_i \in B_{n,\mathbf{i}}\} - nP\{X_i \in B_{n,\mathbf{i}}\} \right\}. \end{aligned}$$

Then, we can write

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

By the definition of $\mathcal{K}_0$ in Assumption A2, by the definition of $R_{n,\mathbf{i}}$ and by the properties of Poisson processes, one can see that the array $\{(\alpha_{n,\mathbf{i}}, u_{n,\mathbf{i}})\}_{\mathbf{i} \in \mathcal{I}_n}$ is an array of 1-dependent random field. (See Mason and Polonik (2009) for details.) For any $q_1, q_2 \in \mathbf{R}$, let $y_{n,\mathbf{i}} \equiv q_1 \alpha_{n,\mathbf{i}} + q_2 u_{n,\mathbf{i}}$. The focus is on the convergence in distribution of $\sum_{\mathbf{i} \in \mathcal{I}_n} y_{n,\mathbf{i}}$ uniform over $P \in \mathcal{P}$. Without loss of generality, we choose $q_1, q_2 \in \mathbf{R} \backslash \{0\}$. Define

$$Var_P\left(\sum_{\mathbf{i} \in \mathcal{I}_n} y_{n,\mathbf{i}} \right) = q_1^2 + q_2^2(1 - \alpha_P) + 2q_1 q_2 c_{n,P},$$

uniformly over $P \in \mathcal{P}$, where $c_{n,P} = Cov(S_n, U_n)$. On the other hand, using Lemma C4 and following the proof of Lemma A8 of Lee, Song, and Whang (2013), we deduce that

$$\sup_{P \in \mathcal{P}} \sum_{\mathbf{i} \in \mathcal{I}_n} \mathbf{E}|y_{n,\mathbf{i}}|^r = o(1) \tag{C.38}$$

as $n \to \infty$, for any $r \in (2, (2p+2)/p]$. By Theorem 1 of Shergin (1993), we have

$$\begin{aligned} &\sup_{P \in \mathcal{P}} \sup_{t \in \mathbf{R}} \left| P\left\{ \frac{1}{\sqrt{q_1^2 + q_2^2(1-\alpha_P) + 2q_1 q_2 c_{n,P}}} \sum_{\mathbf{i} \in \mathcal{I}_n} y_{n,\mathbf{i}} \le t \right\} - \Phi(t) \right| \\ \le\ &\sup_{P \in \mathcal{P}} \frac{C}{\{q_1^2 + q_2^2(1-\alpha_P) + 2q_1 q_2 c_{n,P}\}^{r/2}} \left\{ \sum_{\mathbf{i} \in \mathcal{I}_n} \mathbf{E}|y_{n,\mathbf{i}}|^r \right\}^{1/2} = o(1), \end{aligned}$$

for some $C > 0$, by (C.38). Therefore, by Lemma C2(i), we have for each $t \in \mathbf{R}$, and each $q \in \mathbf{R}^2 \backslash \{0\}$, as $n \to \infty$,

$$\sup_{P \in \mathcal{P}} \left| \mathbf{E}\left[\exp\left(it \frac{q^\top H_n}{\sqrt{q^\top C_n q}} \right) \right] - \exp\left(-\frac{t^2}{2} \right) \right| \to 0.$$

Thus by Lemma C2(ii), for each $t \in \mathbf{R}^2$, we have

$$\sup_{P \in \mathcal{P}} \left| P\left\{ C_n^{-1/2} H_n \le t \right\} - P\left\{ \mathbb{Z} \le t \right\} \right| \to 0.$$

Since the limit distribution of $C_n^{-1/2} H_n$ is continuous, the convergence above is uniform in $t \in \mathbf{R}^2$.

(ii) We fix $P \in \mathcal{P}$ such that $\text{limsup}_{n\to\infty} \sigma_n^2(\mathcal{C}) = 0$. Then by (C.34) above,

$$Var\left(S_n\right) = \sigma_n^2(\mathcal{C}) + o(1) = o(1).$$

Hence, we find that $S_n = o_P(1)$. The desired result follows by applying Theorem 1 of Shergin (1993) to the sum $U_n = \sum_{\mathbf{i} \in \mathcal{I}_n} u_{n,\mathbf{i}}$, and then applying Lemma C2(ii). ∎

Lemma C9. *Let $\mathcal{C}$ be the Borel set in Lemma C8.*

(i) Suppose that the conditions of Lemma C8(i) are satisfied. Then for each $t \in \mathbf{R}$, as $n \to \infty$,

$$\sup_{P \in \mathcal{P}} \sup_{t \in \mathbf{R}} \left| P\left\{ \frac{h^{-d/2} \sum_{A \in \mathcal{N}_J} \mu_A \left\{ \zeta_{n,A} - \mathbf{E}\zeta_{N,A} \right\}}{\sigma_n(\mathcal{C})} \le t \right\} - \Phi(t) \right| \to 0.$$

(ii) Suppose that the conditions of Lemma C8(ii) are satisfied. Then as $n \to \infty$,

$$h^{-d/2} \sum_{A \in \mathcal{N}_J} \mu_A \left\{ \zeta_{n,A} - \mathbf{E}\zeta_{N,A} \right\} \xrightarrow{p} 0.$$

Note that in both statements, the location normalization has $\mathbf{E}\zeta_{N,A}$ instead of $\mathbf{E}\zeta_{n,A}$.

Proof of Lemma C9. (i) The conditional distribution of $S_n/\sigma_n(\mathcal{C})$ given $N = n$ is equal to that of

$$\frac{\sum_{A \in \mathcal{N}_J} \mu_A \int_{B_{n,A}(c_n;\mathcal{C}) \cap \mathcal{C}} \left\{ \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{n,\tau}(x)) - \mathbf{E}\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x)) \right\} dQ(x,\tau)}{h^{d/2}\sigma_n(\mathcal{C})}.$$

Using Lemmas C3(i) and C8(i), we find that

$$\frac{h^{-d/2} \sum_{A \in \mathcal{N}_J} \mu_A \left\{ \zeta_{n,A} - \mathbf{E}\zeta_{N,A} \right\}}{\sigma_n(\mathcal{C})} \xrightarrow{d} N(0,1).$$

Since the limit distribution $N(0,1)$ is continuous and the convergence is uniform in $P \in \mathcal{P}$, we obtain the desired result.

(ii) Similarly as before, the result follows from Lemmas C3(ii), C2(ii), and C8(ii). ∎

Appendix D. Proofs of Auxiliary Results for Lemmas A2(ii), Lemma A4(ii), and Theorem 1

The auxiliary results in this section are mostly bootstrap versions of the results in Appendix C. To facilitate comparison, we name the first lemma to be Lemma D3, which is used to control the discrepancy between the sample version of the scale normalizer σ_n, and its population version. Then we proceed to prove Lemmas D4-D9 which run in parallel with Lemmas C4-C9 as their bootstrap counterparts. We finish this subsection with Lemmas D10-D12 which are crucial for dealing with the bootstrap test statistic's location normalization. More specifically, Lemmas D10 and D11 are auxiliary moment bound results that are used for proving Lemma D12. Lemma D12 essentially delivers the result of Lemma A1 in Appendix A. This lemma is used to deal with the discrepancy between the population location normalizer and the sample location normalizer. Controlling this discrepancy to the rate $o_P(h^{d/2})$ is crucial for our purpose, because our bootstrap test statistic does not involve the sample version of the location normalizer a_n for computational reasons. Lemmas D10 and D11 provide necessary moment bounds to achieve this convergence rate.

The random variables N and N_1 represent Poisson random variables with mean n and 1 respectively. These random variables are independent of $\left((Y_i^{*\top}, X_i^{*\top})_{i=1}^{\infty}, (Y_i^{\top}, X_i^{\top})_{i=1}^{\infty}\right)$. Let η_1 and η_2 be centered normal random vectors that are independent of each other and independent of

$$\left((Y_i^{*\top}, X_i^{*\top})_{i=1}^{\infty}, (Y_i^{\top}, X_i^{\top})_{i=1}^{\infty}, N, N_1\right).$$

We will specify their covariance matrices in the proofs below. Throughout the proofs, the bootstrap distribution P^* and expectations $\mathbf{E}^*$ are viewed as the distribution of

$$\left((Y_i^*, X_i^*)_{i=1}^{n}, N, N_1, \eta_1, \eta_2\right),$$

conditional on $(Y_i, X_i)_{i=1}^{n}$.

Define

$$\begin{aligned} \tilde{\rho}_{n,\tau_1,\tau_2,j,k}(x,u) &\equiv \frac{1}{h^d}\mathbf{E}^*\left[\beta_{n,x,\tau_1,j}\left(Y_{ij}^*, \frac{X_i^*-x}{h}\right)\beta_{n,x,\tau_2,k}\left(Y_{ik}^*, \frac{X_i^*-x}{h}+u\right)\right] \text{ and} \\ \tilde{k}_{n,\tau,j,m}(x) &\equiv \frac{1}{h^d}\mathbf{E}^*\left[\left|\beta_{n,x,\tau,j}\left(Y_{ij}^*, \frac{X_i^*-x}{h}\right)\right|^m\right]. \end{aligned}$$

Note that $\tilde{\rho}_{n,\tau_1,\tau_2,j,k}(x,u)$ and $\tilde{k}_{n,\tau,j,m}(x)$ are bootstrap versions of $\rho_{n,\tau_1,\tau_2,j,k}(x,u)$ and $\tilde{k}_{n,\tau,j,m}(x)$. The lemma below establishes that the bootstrap version $\tilde{\rho}_{n,\tau_1,\tau_2,j,k}(x,u)$ is consistent for $\rho_{n,\tau_1,\tau_2,j,k}(x,u)$.

Lemma D3. *Suppose that Assumption A6(i) holds and that $n^{-1/2}h^{-d/2} \to 0$, as $n \to \infty$. Then for each $\varepsilon \in (0, \varepsilon_1)$, with $\varepsilon_1 > 0$ as in Assumption A6(i), as $n \to \infty$,*

$$\sup_{(\tau_1,\tau_2)\in\mathcal{T}\times\mathcal{T}} \sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \mathbf{E}\left(\left|\tilde{\rho}_{n,\tau_1,\tau_2,j,k}(x,u) - \rho_{n,\tau_1,\tau_2,j,k}(x,u)\right|^2\right) \to 0.$$

Proof of Lemma D3. Define $\pi_{n,x,u,\tau_1,\tau_2}(y,z) = \beta_{n,x,\tau_1,j}(y_j,(z-x)/h)\beta_{n,x,\tau_2,k}(y_k,(z-x)/h+u)$ for $y = (y_1,\cdots,y_J)^\top \in \mathbf{R}^J$, and write

$$\tilde{\rho}_{n,\tau_1,\tau_2,j,k}(x,u) - \rho_{n,\tau_1,\tau_2,j,k}(x,u) = \frac{1}{nh^d}\sum_{i=1}^n \left\{\pi_{n,x,u,\tau_1,\tau_2}(Y_i,X_i) - \mathbf{E}\left[\pi_{n,x,u,\tau_1,\tau_2}(Y_i,X_i)\right]\right\}.$$

First, we note that

$$\mathbf{E}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \left\{\pi_{n,x,u,\tau_1,\tau_2}(Y_i,X_i) - \mathbf{E}\left[\pi_{n,x,u,\tau_1,\tau_2}(Y_i,X_i)\right]\right\}\right)^2 \leq \mathbf{E}\left[\pi^2_{n,x,u,\tau_1,\tau_2}(Y_i,X_i)\right].$$

By change of variables and Assumption A6(i), we have $\mathbf{E}\left[\pi^2_{n,x,u,\tau_1,\tau_2}(Y_i,X_i)\right] = O(h^d)$ uniformly over $(\tau_1,\tau_2) \in \mathcal{T}\times\mathcal{T}$, $(x,u) \in (\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}$ and over $P \in \mathcal{P}$. Hence

$$\mathbf{E}\left(\left|\tilde{\rho}_{n,\tau_1,\tau_2,j,k}(x,u) - \rho_{n,\tau_1,\tau_2,j,k}(x,u)\right|^2\right) = O\left(n^{-1}h^{-d}\right),$$

uniformly over $(\tau_1,\tau_2) \in \mathcal{T}\times\mathcal{T}$, $(x,u) \in (\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}$ and over $P \in \mathcal{P}$. Since we have assumed that $n^{-1/2}h^{-d/2} \to 0$ as $n \to \infty$, we obtain the desired result. ∎

Lemma D4. *Suppose that Assumption A6(i) holds and that for some $C > 0$,*

$$limsup_{n\to\infty} n^{-1/2}h^{-d/2} \leq C.$$

Then for all $m \in [2,M]$ and all $\varepsilon \in (0,\varepsilon_1)$, with $M > 0$ and $\varepsilon_1 > 0$ being the constants that appear in Assumption A6(i)), there exists $C_1 \in (0,\infty)$ that does not depend on n such that for each $j \in \mathbb{N}_J$,

$$\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\sup_{P\in\mathcal{P}}\mathbf{E}\left[\tilde{k}^2_{n,\tau,j,m}(x)\right] \leq C_1.$$

Proof of Lemma D4. Since $\mathbf{E}^*[|\beta_{n,x,\tau,j}(Y^*_{ij},(X^*_i-x)/h)|^m] = \frac{1}{n}\sum_{i=1}^n |\beta_{n,x,\tau,j}(Y_{ij},(X_i-x)/h)|^m$, we find that

$$\tilde{k}^2_{n,\tau,j,m}(x) \leq 2k^2_{n,\tau,j,m}(x) + 2e^2_{n,\tau,j,m}(x),$$

where

$$e_{n,\tau,j,m}(x) \equiv \left|\frac{1}{nh^d}\sum_{i=1}^n \left|\beta_{n,x,\tau,j}\left(Y_{ij},\frac{X_i-x}{h}\right)\right|^m - \frac{1}{h^d}\mathbf{E}\left(\left|\beta_{n,x,\tau,j}\left(Y_{ij},\frac{X_i-x}{h}\right)\right|^m\right)\right|.$$

Similarly as in the proof of Lemma D3, we note that

$$\begin{aligned}&\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\sup_{P\in\mathcal{P}}\mathbf{E}\left[\left|e^2_{n,\tau,j,m}(x)\right|\right]\\ \leq\ &\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\sup_{P\in\mathcal{P}}\frac{1}{nh^{2d}}\mathbf{E}\left[\left|\beta_{n,x,\tau,j}\left(Y_{ij},\frac{X_i-x}{h}\right)\right|^{2m}\right] = O(n^{-1}h^{-d}) = o(1),\ \text{as } n\to\infty.\end{aligned}$$

Hence the desired statement follows from Lemma C4. ∎

Let

$$\begin{aligned}\mathbf{z}^*_{n,\tau}(x) &\equiv \frac{1}{nh^d}\sum_{i=1}^{n}\beta_{n,x,\tau}\left(Y_i^*, \frac{X_i^*-x}{h}\right) - \frac{1}{h^d}\mathbf{E}^*\left[\beta_{n,x,\tau}\left(Y_i^*, \frac{X_i^*-x}{h}\right)\right], \text{ and}\\ \mathbf{z}^*_{N,\tau}(x) &\equiv \frac{1}{nh^d}\sum_{i=1}^{N}\beta_{n,x,\tau}\left(Y_i^*, \frac{X_i^*-x}{h}\right) - \frac{1}{h^d}\mathbf{E}^*\left[\beta_{n,x,\tau}\left(Y_i^*, \frac{X_i^*-x}{h}\right)\right].\end{aligned}$$

We also let

$$\begin{aligned}q^*_{n,\tau}(x) &\equiv \frac{1}{\sqrt{h^d}}\sum_{i\le N_1}\left\{\beta_{n,x,\tau}(Y_i^*,(X_i^*-x)/h) - \mathbf{E}^*\beta_{n,x,\tau}(Y_i^*,(X_i^*-x)/h)\right\} \text{ and}\\ \bar{q}^*_{n,\tau}(x) &\equiv \frac{1}{\sqrt{h^d}}\left\{\beta_{n,x,\tau}(Y_i^*,(X_i^*-x)/h) - \mathbf{E}^*\beta_{n,x,\tau}(Y_i^*,(X_i^*-x)/h)\right\}.\end{aligned}$$

Lemma D5. *Suppose that Assumption A6(i) holds and that for some $C > 0$,*

$$limsup_{n\to\infty} n^{-1/2}h^{-d/2} \le C.$$

Then for any $m \in [2, M]$ (with M being the constant M in Assumption A6(i)),

$$\text{(D.1)}\qquad \begin{aligned}\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\sqrt{\mathbf{E}\left[\left(\mathbf{E}^*\left[||q^*_{n,\tau}(x)||^m\right]\right)^2\right]} &\le \bar{C}_1 h^{d(1-(m/2))}, \text{ and}\\ \sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\sqrt{\mathbf{E}\left[\left(\mathbf{E}^*\left[||\bar{q}^*_{n,\tau}(x)||^m\right]\right)^2\right]} &\le \bar{C}_2 h^{d(1-(m/2))},\end{aligned}$$

where $\bar{C}_1, \bar{C}_2 > 0$ are constants that depend only on m. If furthermore,

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

for some constant $C > 0$, then

$$\text{(D.2)}\qquad \begin{aligned}\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}\left[\mathbf{E}^*\left[||\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x)||^m\right]\right] &\le \left(\frac{15m}{\log m}\right)^m \max\left\{\bar{C}_1, 2\bar{C}_1 C\right\}, \text{ and}\\ \sup_{(x,\tau)\in\mathcal{X}^{\varepsilon/2}\times\mathcal{T}}\sup_{P\in\mathcal{P}}\mathbf{E}\left[\mathbf{E}^*\left[||\sqrt{nh^d}\mathbf{z}^*_{n,\tau}(x)||^m\right]\right] &\le \left(\frac{15m}{\log m}\right)^m \max\left\{\bar{C}_2, 2\bar{C}_2 C\right\},\end{aligned}$$

where $\bar{C}_1, \bar{C}_2 > 0$ are the constants that appear in (D.1).

Proof of Lemma D5. Let $q^*_{n,\tau,j}(x)$ be the j-th entry of $q^*_{n,\tau}(x)$. For the first statement of the lemma, it suffices to observe that for each $\varepsilon \in (0, \varepsilon_1)$, there exist $C_1 > 0$ and $\bar{C}_1 > 0$ such that

$$\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\mathbf{E}\left[\left(\mathbf{E}^*\left[|q^*_{n,\tau,j}(x)|^m\right]\right)^2\right] \le \frac{C_1 h^{2d}\sum_{j=1}^{J}\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\sup_{P\in\mathcal{P}}\mathbf{E}\left[\tilde{k}^2_{n,\tau,j,m}(x)\right]}{h^{dm}} \le \bar{C}_1 h^{2d(1-(m/2))},$$

where the last inequality uses Lemma D4. The second inequality in (D.1) follows similarly.

Let us consider (D.2). Let $z^*_{N,\tau,j}(x)$ be the j-th entry of $\mathbf{z}^*_{N,\tau}(x)$. Then using Rosenthal's inequality (e.g. (2.3) of Giné, Mason, and Zaitsev (2003)), for some constant $C_1 > 0$,

$$\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\sup_{P\in\mathcal{P}}\mathbf{E}\left[\mathbf{E}^*[|\sqrt{nh^d}z^*_{N,\tau,j}(x)|^m]\right]$$
$$\leq \left(\frac{15m}{\log m}\right)^{2m}\sup_{\tau\in\mathcal{T},x\in\mathcal{S}_\tau(\varepsilon)}\sup_{P\in\mathcal{P}}\left\{\left(\mathbf{E}\left[\mathbf{E}^*\left(q^{*2}_{n,\tau,j}(x)\right)\right]\right)^{m/2}+\mathbf{E}\left[n^{-(m/2)+1}\mathbf{E}^*|q^*_{n,\tau,j}(x)|^m\right]\right\}.$$

The first expectation is bounded by $\bar{C}_1$ by (D.1).

The second expectation is bounded by $\bar{C}_1 n^{-(m/2)+1}h^{d(1-(m/2))}$. This gives the first bound in (D.2). The second bound in (D.2) can be obtained similarly. ∎

For any Borel sets $B, B' \subset \mathcal{S}$ and $A, A' \subset \mathbb{N}_J$, let

$$\tilde{\sigma}^R_{n,A,A'}(B,B') \equiv \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B'_{\tau_2}}\int_{B_{\tau_1}} C^*_{n,\tau_1,\tau_2,A,A'}(x,v)dxdvd\tau_1 d\tau_2,$$

where $B_\tau \equiv \{x \in \mathcal{X} : (x,\tau) \in B\}$,

$$C^*_{n,\tau_1,\tau_2,A,A'}(x,v) \equiv h^{-d}Cov^*\left(\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{N,\tau_1}(x)), \Lambda_{A',p}(\sqrt{nh^d}\mathbf{z}^*_{N,\tau_2}(v))\right), \tag{D.3}$$

and Cov^* represents covariance under P^*. We also define

$$\tilde{\sigma}^R_{n,A}(B) \equiv \tilde{\sigma}^R_{n,A,A}(B,B), \tag{D.4}$$

for brevity. Also, let $\Sigma^*_{n,\tau_1,\tau_2}(x,u)$ be a $J \times J$ matrix whose (j,k)-th entry is given by $\tilde{\rho}_{n,\tau_1,\tau_2,j,k}(x,u)$. Fix $\bar{\varepsilon} > 0$ and define

$$\tilde{\Sigma}^*_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) \equiv \begin{bmatrix} \Sigma^*_{n,\tau_1,\tau_1}(x,0)+\bar{\varepsilon}I_J & \Sigma^*_{n,\tau_1,\tau_2}(x,u) \\ \Sigma^*_{n,\tau_1,\tau_2}(x,u) & \Sigma^*_{n,\tau_2,\tau_2}(x,0)+\bar{\varepsilon}I_J \end{bmatrix}.$$

We also define

$$\xi^*_{N,\tau_1,\tau_2}(x,u;\eta_1,\eta_2) \equiv \sqrt{nh^d}\Sigma^{*-1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\begin{bmatrix}\mathbf{z}^*_{N,\tau_1}(x;\eta_1)\\ \mathbf{z}^*_{N,\tau_2}(x+uh;\eta_2)\end{bmatrix},$$

where $\eta_1 \in \mathbf{R}^J$ and $\eta_2 \in \mathbf{R}^J$ are random vectors that are independent, and independent of $((Y_i^*,X_i^*)_{i=1}^\infty,(Y_i,X_i)_{i=1}^\infty,N,N_1)$, each following $N(0,\bar{\varepsilon}I_J)$, and define $\mathbf{z}^*_{N,\tau}(x;\eta_1) \equiv \mathbf{z}^*_{N,\tau}(x) + \eta_1/\sqrt{nh^d}$.

Lemma D6. *Suppose that Assumption A6(i) holds and that $nh^d \to \infty$, and*

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

for some $C > 0$ and some $m \in [2(p+1), M]$.

Then for any sequences of Borel sets $B_n, B_n' \subset \mathcal{S}$ and for any $A, A' \subset \mathbb{N}_J$,

$$\sup_{P \in \mathcal{P}} \mathbf{E}\left(\left|\left(\tilde{\sigma}_{n,A,A'}^R(B_n, B_n')\right)^2 - \sigma_{n,A,A'}^2(B_n, B_n')\right|\right) \to 0,$$

where $\sigma_{n,A,A'}^2(B_n, B_n')$ is as defined in (C.15).

Proof of Lemma D6. The proof is very similar to that of Lemma C6. For brevity, we sketch the proof here. Define for $\bar{\varepsilon} > 0$,

$$\begin{aligned}
\tilde{\sigma}_{n,A,A',\bar{\varepsilon}}^R(B_n, B_n') &\equiv \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_{n,\tau_1}}\int_{\mathcal{U}} \tilde{g}_{1n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) w_{\tau_1,B_n}(x) w_{\tau_2,B_n'}(x+uh) du dx d\tau_1 d\tau_2, \\
\tilde{\tau}_{n,A,A',\bar{\varepsilon}}(B_n, B_n') &\equiv \int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_{n,\tau_1}}\int_{\mathcal{U}} \tilde{g}_{2n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) w_{\tau_1,B_n}(x) w_{\tau_2,B_n'}(x+uh) du dx d\tau_1 d\tau_2,
\end{aligned}$$

where

$$\begin{aligned}
\tilde{g}_{1n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) &\equiv h^{-d} Cov^*(\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_1}^*(x;\eta_1)), \Lambda_{A',p}(\sqrt{nh^d}\mathbf{z}_{N,\tau_2}^*(x+uh;\eta_2))), \text{ and} \\
\tilde{g}_{2n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) &\equiv Cov^*(\Lambda_{A,p}(\tilde{\mathbb{Z}}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x)), \Lambda_{A',p}(\tilde{\mathbb{Z}}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x+uh))),
\end{aligned}$$

and $[\tilde{\mathbb{Z}}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^\top(x), \tilde{\mathbb{Z}}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^\top(z)]^\top$ is a centered normal $\mathbf{R}^{2J}$-valued random vector with the same covariance matrix as the covariance matrix of $[\sqrt{nh^d}\mathbf{z}_{N,\tau_1}^{*\top}(x;\eta_1), \sqrt{nh^d}\mathbf{z}_{N,\tau_2}^{*\top}(z;\eta_2)]^\top$ under the product measure of the bootstrap distribution P^* and the distribution of $(\eta_1^\top, \eta_2^\top)^\top$. As in the proof of Lemma C6, it suffices for the lemma to show the following two statements.
(*Step 1*): As $n \to \infty$,

$$\begin{aligned}
\sup_{P \in \mathcal{P}} \mathbf{E}\left(\left|\tilde{\sigma}_{n,A,A',\bar{\varepsilon}}^R(B_n, B_n') - \tilde{\tau}_{n,A,A',\bar{\varepsilon}}(B_n, B_n')\right|\right) &\to 0, \text{ and} \\
\sup_{P \in \mathcal{P}} \mathbf{E}\left(\left|\tilde{\tau}_{n,A,A',\bar{\varepsilon}}(B_n, B_n') - \sigma_{n,A,A',\bar{\varepsilon}}(B_n, B_n')\right|\right) &\to 0.
\end{aligned}$$

(*Step 2*): For some $C > 0$ that does not depend on $\bar{\varepsilon}$ or n,

$$\sup_{P \in \mathcal{P}} |\tilde{\sigma}_{n,A,A',\bar{\varepsilon}}^R(B_n, B_n') - \tilde{\sigma}_{n,A,A'}^R(B_n, B_n')| \le C\sqrt{\bar{\varepsilon}}.$$

Then the desired result follows by sending $n \to \infty$ and $\bar{\varepsilon} \downarrow 0$, while chaining Steps 1 and 2 and the second convergence in Step 2 in the proof of Lemma C6.

We first focus on the first statement of (Step 1). For any vector $\mathbf{v} = [\mathbf{v}_1^\top, \mathbf{v}_2^\top]^\top \in \mathbf{R}^{2J}$, we define

$$\tilde{C}_{n,p}(\mathbf{v}) \equiv \Lambda_{A,p}\left(\left[\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^{*1/2}(x,u)\mathbf{v}\right]_1\right) \Lambda_{A',p}\left(\left[\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^{*1/2}(x,u)\mathbf{v}\right]_2\right), \tag{D.5}$$

where $[a]_1$ of a vector $a \in \mathbf{R}^{2J}$ indicates the vector of the first J entries of a, and $[a]_2$ the vector of the remaining J entries of a. Also, similarly as in (C.19),

$$\lambda_{\min}\left(\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^*(x,u)\right) \ge \bar{\varepsilon}. \tag{D.6}$$

Let $\bar{q}^*_{n,\tau}(x;\eta_1)$ be the column vector of entries $\bar{q}^*_{n,\tau,j}(x;\eta_{1j})$ with j running in the set $\mathbb{N}_J$, and with

$$\bar{q}^*_{n,\tau,j}(x;\eta_{1j}) \equiv p^*_{n,\tau,j}(x) + \eta_{1j},$$

where

$$p^*_{n,\tau,j}(x) = \frac{1}{\sqrt{h^d}} \sum_{1\leq i\leq N_1} \left\{ \beta_{n,x,\tau,j}\left(Y^*_{ij}, \frac{X^*_i - x}{h}\right) - \mathbf{E}\left[\beta_{n,x,\tau,j}\left(Y^*_{ij}, \frac{X^*_i - x}{h}\right)\right]\right\},$$

η_{1j} is the j-th entry of η_1, and N_1 is a Poisson random variable with mean 1 and $((\eta_{1j})_{j\in A}, N_1)$ is independent of $\{(Y_i^\top, X_i^\top, Y_i^{*\top}, X_i^{*\top})\}_{i=1}^n$. Let $[p^{*(i)}_{n,\tau_1}(x), p^{*(i)}_{n,\tau_2}(x+uh)]$ be the i.i.d. copies of $[p^*_{n,\tau_1}(x), p^*_{n,\tau_2}(x+uh)]$ conditional on the observations $\{(Y_i, X_i)\}_{i=1}^n$, and $\eta_1^{(i)}$ and $\eta_2^{(i)}$ be i.i.d. copies of η_1 and η_2. Define

$$q^{*(i)}_{n,\tau_1}(x;\eta_1^{(i)}) = p^{*(i)}_{n,\tau_1}(x) + \eta_1^{(i)} \text{ and } q^{*(i)}_{n,\tau_2}(x+uh;\eta_2^{(i)}) = p^{*(i)}_{n,\tau_2}(x+uh) + \eta_2^{(i)}.$$

Note that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\begin{array}{c} q^{*(i)}_{n,\tau_1}(x;\eta_1^{(i)}) \\ q^{*(i)}_{n,\tau_2}(x+uh;\eta_2^{(i)}) \end{array}\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\begin{array}{c} p^{*(i)}_{n,\tau_1}(x) \\ p^{*(i)}_{n,\tau_2}(x+uh) \end{array}\right] + \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\begin{array}{c} \eta_1^{(i)} \\ \eta_2^{(i)} \end{array}\right].$$

The last sum has the same distribution as $[\eta_1^\top, \eta_2^\top]^\top$ and the leading sum on the right-hand side has the same bootstrap distribution as that of $[\mathbf{z}^{*\top}_{N,\tau_1}(x), \mathbf{z}^{*\top}_{N,\tau_2}(x+uh)]^\top$, P-a.e. Therefore, we conclude that

$$\xi^*_{N,\tau_1,\tau_2}(x,u;\eta_1^{(i)},\eta_2^{(i)}) \overset{d^*}{=} \frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{W}^{(i)}_{n,\tau_1,\tau_2}(x,u;\eta_1^{(i)},\eta_2^{(i)}),$$

where $\overset{d^*}{=}$ indicates the distributional equivalence with respect to the product measure of the bootstrap distribution P^* and the joint distribution of $(\eta_1^{(i)},\eta_2^{(i)})$, P-a.e, and

$$\tilde{W}^{(i)}_{n,\tau_1,\tau_2}(x,u;\eta_1^{(i)},\eta_2^{(i)}) \equiv \tilde{\Sigma}^{*-1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u) \left[\begin{array}{c} q_n^{(i)}(x;\eta_1^{(i)}) \\ q_n^{(i)}(x+uh;\eta_2^{(i)}) \end{array}\right].$$

Following the arguments in the proof of Lemma C6, we find that for each $u \in \mathcal{U}$, and for $\varepsilon \in (0,\varepsilon_1)$ with ε_1 as in Assumption A6(i),

$$\begin{aligned} & \sup_{(x,u)\in(\mathcal{S}_{\tau_1}\cup\mathcal{S}_{\tau_2})\times\mathcal{U}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[\mathbf{E}^*||\tilde{W}^{(i)}_{n,\tau_1,\tau_2}(x,u;\eta_1^{(i)},\eta_2^{(i)})||^3\right] \\ \leq\ & C_1 \sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[\lambda^3_{\min}\left(\tilde{\Sigma}^{*-1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\right)\mathbf{E}^*||q^{*(i)}_{n,\tau_1}(x;\eta_1^{(i)})||^3\right] \\ & +C_1 \sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[\lambda^3_{\min}\left(\tilde{\Sigma}^{*-1/2}_{n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\right)\mathbf{E}^*||q^{*(i)}_{n,\tau_2}(x+uh;\eta_2^{(i)})||^3\right], \end{aligned}$$

for some $C_1 > 0$. As for the leading term,

$$\begin{aligned}
&\sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[\lambda_{\min}^3\left(\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^{*-1/2}(x,u)\right)\mathbf{E}^*||q_{n,\tau_1}^{*(i)}(x;\eta_1^{(i)})||^3\right] \\
\leq &\sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \sqrt{\mathbf{E}\left[\left(\mathbf{E}^*||q_{n,\tau_1}^{*(i)}(x;\eta_1^{(i)})||^3\right)^2\right]} \\
&\times \sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \sqrt{\mathbf{E}\left[\lambda_{\min}^6\left(\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^{*-1/2}(x,u)\right)\right]} \leq \frac{C_2\bar{\varepsilon}^{-3}}{\sqrt{h^d}},
\end{aligned}$$

by Lemma D5 and (D.6). Similarly, we observe that

$$\sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[\lambda_{\min}^3\left(\tilde{\Sigma}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^{*-1/2}(x,u)\right)\mathbf{E}^*||q_{n,\tau_2}^{*(i)}(x+uh;\eta_2^{(i)})||^3\right] \leq \frac{C_2\bar{\varepsilon}^{-3}}{\sqrt{h^d}}.$$

Define

$$c_{n,\tau_1,\tau_2}(x,u) = \tilde{C}_{n,p}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \tilde{W}_{n,\tau_1,\tau_2}^{(i)}(x,u;\eta_1^{(i)},\eta_2^{(i)})\right).$$

Let $\Phi_{n,\tau_1,\tau_2}(\cdot;x,u)$ be the joint CDF of the random vector $(\tilde{\mathbb{Z}}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^\top(x),\tilde{\mathbb{Z}}_{n,\tau_1,\tau_2,\bar{\varepsilon}}^\top(x+uh))^\top$. By Theorem 1 of Sweeting (1977),

$$\begin{aligned}
\text{(D.7)}\quad &\sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[\left|c_{n,\tau_1,\tau_2}(x,u) - \int \tilde{C}_{n,p}(\zeta)d\Phi_{n,\tau_1,\tau_2}(\zeta;x,u)\right|\right] \\
\leq &\frac{C_1}{\sqrt{n}}\sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \mathbf{E}\left[\mathbf{E}^*||\tilde{W}_{n,\tau_1,\tau_2}^{(i)}(x,u;\eta_1^{(i)},\eta_2^{(i)})||^3\right] \leq \frac{C_2\bar{\varepsilon}^{-3}}{\sqrt{nh^d}}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathbf{E}\left[\left|\int_{B_{\tau_1}}\int_{\mathcal{U}}\left\{\tilde{g}_{1n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)-\tilde{g}_{2n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\right\}w_{\tau_1,B}(x)w_{\tau_2,B'}(x+uh)dudx\right|\right] \\
\leq &\int_{B_{\tau_1}}\int_{\mathcal{U}}\mathbf{E}\left|\tilde{g}_{1n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)-\tilde{g}_{2n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\right|w_{\tau_1,B}(x)w_{\tau_2,B'}(x+uh)dudx \\
\leq &\int_{B_{\tau_1}} w_{\tau_1,B}(x)w_{\tau_2,B'}(x)dx \\
&\times \sup_{(x,u)\in(\mathcal{S}_{\tau_1}(\varepsilon)\cup\mathcal{S}_{\tau_2}(\varepsilon))\times\mathcal{U}} \sup_{P\in\mathcal{P}} \mathbf{E}\left|\tilde{g}_{1n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)-\tilde{g}_{2n,\tau_1,\tau_2,\bar{\varepsilon}}(x,u)\right| \\
\to &\ 0,
\end{aligned}$$

as $n\to\infty$. The last convergence is due to (D.7) and hence uniform over $(\tau_1,\tau_2)\in\mathcal{T}\times\mathcal{T}$. The proof of (Step 1) is thus complete.

We turn to the second statement of (Step 1). Similarly as in the proof of Step 1 in the proof of Lemma C6, the second statement of Step 1 follows by Lemma D4.

Now we turn to (Step 2). In view of the proof of Step 2 in the proof of Lemma C6, it suffices to show that with $s = (p+1)/(p-1)$ if $p > 1$ and $s = 2$ if $p = 1$,

$$\text{(D.8)} \qquad \sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} \sup_{P \in \mathcal{P}} \mathbf{E}\left[\mathbf{E}^* \left\|\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x)\right\|^{2s(p-1)}\right] < C \text{ and}$$

$$\sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} \sup_{P \in \mathcal{P}} \mathbf{E}\left[\mathbf{E}^* \left\|\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x;\eta_1)\right\|^{2s(p-1)}\right] < C,$$

for some $C > 0$. First note that for any $q > 0$,

$$\begin{aligned} &\sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} \sup_{P \in \mathcal{P}} \mathbf{E}\left[\mathbf{E}^* \left\|\sqrt{nh^d}\{\mathbf{z}^*_{N,\tau}(x) - \mathbf{z}^*_{N,\tau}(x;\eta_1)\}\right\|^{2q}\right] \\ &= \mathbf{E}\left\|\sqrt{\bar{\varepsilon}}\mathbb{Z}\right\|^{2q} = C\bar{\varepsilon}^q, \end{aligned}$$

where $\mathbb{Z} \in \mathbf{R}^J$ is a centered normal random vector with covariance matrix I_J. Also, we deduce that for some constants $C_1, C_2 > 0$,

$$\begin{aligned} &\sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} \sup_{P \in \mathcal{P}} \mathbf{E}\left[\mathbf{E}^* \left\|\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x)\right\|^{2s(p-1)}\right] \\ \leq &\sup_{\tau \in \mathcal{T}, x \in \mathcal{S}_\tau(\varepsilon)} \sup_{P \in \mathcal{P}} \mathbf{E}\left[\mathbf{E}^* \left\|\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x;\eta_1)\right\|^{2s(p-1)}\right] + C_1\bar{\varepsilon}^{s(p-1)} \leq C_1 + C_2\bar{\varepsilon}^{s(p-1)}, \end{aligned}$$

by the third statement of Lemma D5. This leads to the first and second statements of (D.8). Thus the proof of the lemma is complete. ∎

Lemma D7. *Suppose that for some small $\nu_1 > 0$, $n^{-1/2}h^{-d-\nu_1} \to \infty$, as $n \to \infty$ and the conditions of Lemma C6 hold. Then there exists $C > 0$ such that for any sequence of Borel sets $B_n \subset \mathcal{S}$, and $A \subset \mathbb{N}_J$, from some large n on,*

$$\begin{aligned} &\sup_{P \in \mathcal{P}} \mathbf{E}\left(\mathbf{E}^*\left[\left|h^{-d/2}\int_{B_n}\left\{\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{n,\tau}(x)) - \mathbf{E}^*\left[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x))\right]\right\}dQ(x,\tau)\right|\right]\right) \\ &\leq C\sqrt{Q(B_n)}. \end{aligned}$$

Proof of Lemma D7. We follow the proof of Lemma C7 and show that for some $C > 0$, we have the following:

Step 1: $\sup_{P \in \mathcal{P}} \mathbf{E}\left(\mathbf{E}^*\left[\left|h^{-d/2}\int_{B_n}\left\{\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{n,\tau}(x)) - \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x))\right\}dQ(x,\tau)\right|\right]\right) \leq CQ(B_n)$,

Step 2:

$$\begin{aligned} &\sup_{P \in \mathcal{P}} \mathbf{E}\left(\mathbf{E}^*\left[\left|h^{-d/2}\int_{B_n}\left\{\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{n,\tau}(x)) - \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x))\right\}dQ(x,\tau)\right|\right]\right) \\ \leq\ &C\sqrt{Q(B_n)}. \end{aligned}$$

Proof of Step 1: Similarly as in the proof of Step 1 in the proof of Lemma C7, we first write

$$\mathbf{z}_{n,\tau}^*(x) = \mathbf{z}_{N,\tau}^*(x) + \mathbf{v}_{n,\tau}^*(x) + \mathbf{s}_{n,\tau}^*(x),$$

where

$$\begin{aligned}\mathbf{v}_{n,\tau}^*(x) &\equiv \left(\frac{n-N}{n}\right) \cdot \frac{1}{h^d}\mathbf{E}^*\left[\beta_{n,x,\tau}\left(Y_i^*, \frac{X_i^*-x}{h}\right)\right] \text{ and} \\ \mathbf{s}_{n,\tau}^*(x) &\equiv \frac{1}{nh^d}\sum_{i=N+1}^{n}\left\{\beta_{n,x,\tau}\left(Y_i^*, \frac{X_i^*-x}{h}\right) - \mathbf{E}^*\left[\beta_{n,x,\tau}\left(Y_i^*, \frac{X_i^*-x}{h}\right)\right]\right\}.\end{aligned}$$

Similarly as in the proof of Lemma C7, we deduce that for some $C_1, C_2 > 0$,

$$\begin{aligned}&\left|\int_{B_n}\left\{\Lambda_{A,p}\left(\mathbf{z}_{n,\tau}^*(x)\right) - \Lambda_{A,p}\left(\mathbf{z}_{N,\tau}^*(x)\right)\right\}dQ(x,\tau)\right| \\ &\leq C_1\int_{B_n}\left\|\mathbf{v}_{n,\tau}^*(x)\right\|\left(\left\|\mathbf{z}_{n,\tau}^*(x)\right\|^{p-1} + \left\|\mathbf{z}_{N,\tau}^*(x)\right\|^{p-1}\right)dQ(x,\tau) \\ &\quad + C_2\int_{B_n}\left\|\mathbf{s}_{n,\tau}^*(x)\right\|\left(\left\|\mathbf{z}_{n,\tau}^*(x)\right\|^{p-1} + \left\|\mathbf{z}_{N,\tau}^*(x)\right\|^{p-1}\right)dQ(x,\tau) \\ &= D_{1n}^* + D_{2n}^*, \text{ say.}\end{aligned}$$

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

Claim 1: $\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}\left(\mathbf{E}^*[||\mathbf{v}_{n,\tau}^*(x)||^2]\right) = O(n^{-1})$.

Claim 2: $\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}\left(\mathbf{E}^*[||\mathbf{s}_{n,\tau}^*(x)||^2]\right) = O(n^{-3/2}h^{-d})$.

Proof of Claim 1: Similarly as in the proof of Lemma C7, we note that

$$\mathbf{E}\left(\mathbf{E}^*[||\mathbf{v}_{n,\tau}^*(x)||^2]\right) \leq \mathbf{E}\left|\left(\frac{n-N}{n}\right)\right|^2\mathbf{E}\left[\left\|\frac{1}{h^d}\mathbf{E}^*\left[\beta_{n,x,\tau}\left(Y_i^*, \frac{X_i^*-x}{h}\right)\right]\right\|^2\right].$$

By the first statement of Lemma D5, we have

$$\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}\left[\left\|\frac{1}{h^d}\mathbf{E}^*\left[\beta_{n,x,\tau}\left(Y_i^*, \frac{X_i^*-x}{h}\right)\right]\right\|^2\right] = O(1).$$

Since $\mathbf{E}\left|(n-N)/n\right|^2 = O(n^{-1})$, we obtain Claim 1.

Proof of Claim 2: Let

$$\mathbf{s}_{n,\tau}^*(x;\eta_1) = \mathbf{s}_{n,\tau}^*(x) + \frac{(N-n)\eta_1}{n^{3/2}h^{d/2}},$$

where η_1 is a random vector independent of $((Y_i^*, X_i^*)_{i=1}^n, (Y_i, X_i)_{i=1}^n, N)$ and follows $N(0, \bar{\varepsilon} I_J)$. Note that

$$\begin{aligned}
&\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left(\mathbf{E}^*\left\|\sqrt{nh^d}\mathbf{s}_{n,\tau}^*(x)\right\|^2\right) \\
\leq\ & 2\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left(\mathbf{E}^*\left\|\sqrt{nh^d}\mathbf{s}_{n,\tau}^*(x;\eta_1)\right\|^2\right) + \frac{2}{n}\mathbf{E}\left\|\frac{(N-n)\eta_1}{\sqrt{n}}\right\|^2 \\
\leq\ & 2\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left(\mathbf{E}^*\left\|\sqrt{nh^d}\mathbf{s}_{n,\tau}^*(x;\eta_1)\right\|^2\right) + \frac{C\bar{\varepsilon}^2}{n},
\end{aligned}$$

as in the proof of Lemma C7. As for the leading expectation on the right hand side of (C.27), we let $C_1 > 0$ be as in Lemma D4 and note that

$$\begin{aligned}
\mathbf{E}\left(\mathbf{E}^*\left\|\sqrt{nh^d}\mathbf{s}_{n,\tau}^*(x;\eta_1)\right\|^2\right) &= \sum_{j\in\mathbb{N}_J}\mathbf{E}\left(\mathbf{E}^*\left(\frac{1}{\sqrt{n}}\sum_{i=N+1}^n q_{n,\tau,j}^{*(i)}(x;\eta_{1j}^{(i)})\right)^2\right) \\
&= \frac{1}{n}\sum_{j\in\mathbb{N}_J}\mathbf{E}\left(\tilde{\sigma}_{n,\tau,j}^2(x)\mathbf{E}^*\left(\sum_{i=N+1}^n \frac{q_{n,\tau,j}^{*(i)}(x;\eta_{1j}^{(i)})}{\tilde{\sigma}_{n,\tau,j}(x)}\right)^2\right),
\end{aligned}$$

where $q_{n,\tau}^{*(i)}(x;\eta_1^{(i)})$'s $(i = 1, 2, \cdots)$ are as defined in the proof of Lemma D6 and $q_{n,\tau,j}^{*(i)}(x;\eta_{1j}^{(i)})$ is the j-th entry of $q_{n,\tau}^{*(i)}(x;\eta_1^{(i)})$ and $\tilde{\sigma}_{n,\tau,j}^2(x) = Var^*(q_{n,\tau,j}^{*(i)}(x;\eta_{1j}^{(i)})) > 0$ and Var^* denotes the variance with respect to the joint distribution of $((Y_i^*, X_i^*)_{i=1}^n, \eta_{1j}^{(i)})$ conditional on $(Y_i, X_i)_{i=1}^n$. We apply Lemma 1(i) of Horváth (1991) to deduce that

$$\text{(D.9)}\qquad \begin{aligned}
\mathbf{E}^*\left(\sum_{i=N+1}^n \frac{q_{n,\tau,j}^{*(i)}(x;\eta_{1j}^{(i)})}{\tilde{\sigma}_{n,\tau,j}(x)}\right)^2 &\leq C\sqrt{n} + C\mathbf{E}^*\left(\left|\frac{q_{n,\tau,j}^{*(i)}(x;\eta_{1j}^{(i)})}{\tilde{\sigma}_{n,\tau,j}(x)}\right|^3\right) \\
&\quad + C\mathbf{E}^*\left(\left|\frac{q_{n,\tau,j}^{*(i)}(x;\eta_{1j}^{(i)})}{\tilde{\sigma}_{n,\tau,j}(x)}\right|^4\right),
\end{aligned}$$

for some $C > 0$. Using this, Lemma D5, and following arguments similarly as in (C.29), (C.30) and (C.31), we conclude that

$$\begin{aligned}
\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\left(\mathbf{E}^*\left\|\sqrt{nh^d}\mathbf{s}_{n,\tau}^*(x)\right\|^2\right) &\leq O\left(n^{-1}h^{-\nu_1}\right) + O\left(n^{-1/2} + n^{-3/4}h^{-d/2-\nu_1} + n^{-1}h^{-d-\nu_1}\right) \\
&= O\left(n^{-1}h^{-\nu_1}\right) + O\left(n^{-1/2}\right),
\end{aligned}$$

since $n^{-1/2}h^{-d-\nu_1} \to 0$. This delivers Claim 2.

Using Claims 1 and 2, and following the arguments in the proof of Lemma C7, we obtain (Step 1).

Proof of Step 2: We can follow the proof of Lemma C6 to show that

$$\begin{aligned}
&\mathbf{E}\left[\mathbf{E}^*\left[h^{-d/2}\int_{B_n}\left(\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x))-\mathbf{E}^*\left[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x))\right]\right)dQ(x,\tau)\right]^2\right]\\
=\;&\mathbf{E}\left[\int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_{n,\tau_1}\cap B_{n,\tau_2}}\int_{\mathcal{U}}C^*_{n,\tau_1,\tau_2,A,A'}(x,u)dudxd\tau_1 d\tau_2\right]+o(1)\\
\leq\;& C\int_{\mathcal{T}}\int_{\mathcal{T}}\int_{B_{n,\tau_1}\cap B_{n,\tau_2}}dxd\tau_1 d\tau_2+o(1)\leq CQ(B_n),
\end{aligned}$$

where $C^*_{n,\tau_1,\tau_2,A,A'}(x,v)$ is as defined in (D.3). We obtain the desired result of Step 2. ∎

Let $\mathcal{C}\subset\mathbf{R}^d$, $\alpha_P\equiv P\{X\in\mathbf{R}^d\backslash\mathcal{C}\}$ and $B_{n,A}(c_n;\mathcal{C})$ be as introduced prior to Lemma C8. Define

$$\begin{aligned}
\zeta^*_{n,A}&\equiv\int_{B_{n,A}(c_n;\mathcal{C})}\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{n,\tau}(x))dQ(x,\tau),\text{ and}\\
\zeta^*_{N,A}&\equiv\int_{B_{n,A}(c_n;\mathcal{C})}\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x))dQ(x,\tau).
\end{aligned}$$

Let μ_A's be real numbers indexed by $A\subset\mathbb{N}_J$. We also define $B_{n,A}(c_n;\mathcal{C})$ as prior to Lemma C8 and let

$$\begin{aligned}
S^*_n&\equiv h^{-d/2}\sum_{A\in\mathcal{N}_J}\mu_A\left\{\zeta^*_{N,A}-\mathbf{E}^*\zeta^*_{N,A}\right\},\\
U^*_n&\equiv\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^N 1\{X^*_i\in\mathcal{C}\}-nP^*\left\{X^*_i\in\mathcal{C}\right\}\right\},\text{ and}\\
V^*_n&\equiv\frac{1}{\sqrt{n}}\left\{\sum_{i=1}^N 1\{X^*_i\in\mathbf{R}^d\backslash\mathcal{C}\}-nP^*\left\{X^*_i\in\mathbf{R}^d\backslash\mathcal{C}\right\}\right\}.
\end{aligned}$$

We let

$$H^*_n\equiv\left[\frac{S^*_n}{\sigma_n(\mathcal{C})},\frac{U^*_n}{\sqrt{1-\alpha_P}}\right].$$

The following lemma is a bootstrap counterpart of Lemma C8.

Lemma D8. *Suppose that the conditions of Lemma D6 hold and that $c_n\to\infty$, as $n\to\infty$.*
(i) If $\liminf_{n\to\infty}\inf_{P\in\mathcal{P}}\sigma_n^2(\mathcal{C})>0$, *then for all* $a>0$,

$$\sup_{P\in\mathcal{P}}P\left\{\sup_{t\in\mathbf{R}^2}\left|P^*\left\{H^*_n\leq t\right\}-P\left\{\mathbb{Z}\leq t\right\}\right|>a\right\}\to 0.$$

(ii) If $\limsup_{n\to\infty}\sigma_n^2(\mathcal{C})=0$, *then, for each* $(t_1,t_2)\in\mathbf{R}^2$ *and* $a>0$,

$$\sup_{P\in\mathcal{P}}P\left\{\left|P^*\left\{S^*_n\leq t_1\text{ and }\frac{U^*_n}{\sqrt{1-\alpha_P}}\leq t_2\right\}-1\left\{0\leq t_1\right\}P\left\{\mathbb{Z}_1\leq t_2\right\}\right|>a\right\}\to 0.$$

Proof of Lemma D8. Similarly as in the proof of Lemma D8, we fix $\bar{\varepsilon} > 0$ and let

$$H_{n,\bar{\varepsilon}}^* \equiv \left[\frac{S_{n,\bar{\varepsilon}}^*}{\sigma_{n,\bar{\varepsilon}}(\mathcal{C})}, \frac{U_n^*}{\sqrt{1-\alpha_P}}\right]^\top,$$

where $S_{n,\bar{\varepsilon}}^*$ is equal to S_n^*, except that $\zeta_{N,A}^*$ is replaced by

$$\zeta_{N,A,\bar{\varepsilon}}^* \equiv \int_{B_{n,A}(c_n;\mathcal{C})} \Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}^*(x;\eta_1))dQ(x,\tau),$$

and $\mathbf{z}_{N,\tau}^*(x;\eta_1)$ is as defined prior to Lemma D6. Also let

$$\tilde{C}_n \equiv \mathbf{E}^* H_n^* H_n^{*\top} \text{ and } \tilde{C}_{n,\bar{\varepsilon}} \equiv \mathbf{E}^* H_{n,\bar{\varepsilon}}^* H_{n,\bar{\varepsilon}}^{*\top}.$$

First, we show the following statements.

Step 1: $\sup_{P\in\mathcal{P}} P\left\{|Cov^*(S_{n,\bar{\varepsilon}}^* - S_n^*, U_n^*)| > M\sqrt{\bar{\varepsilon}}\right\} \to 0$, as $n \to \infty$ and $M \to \infty$.

Step 2: For any $a > 0$, $\sup_{P\in\mathcal{P}} P\left\{\left|Cov(S_{n,\bar{\varepsilon}}^*, U_n^*)\right| > ah^{d/2}\right\} \to 0$, as $n \to \infty$.

Step 3: There exists $c > 0$ such that from some large n on,

$$\inf_{P\in\mathcal{P}} \lambda_{\min}(\tilde{C}_n) > c.$$

Step 4: For any $a > 0$, as $n \to \infty$,

$$\sup_{P\in\mathcal{P}} P\left\{\sup_{t\in\mathbf{R}^2}\left|P^*\left\{\tilde{C}_n^{-1/2}H_n^* \le t\right\} \to P\left\{\mathbb{Z} \le t\right\}\right| > a\right\} \to 0.$$

Combining Steps 1-4, we obtain (i) of Lemma C8.

Proof of Step 1: Observe that

$$\left|\zeta_{N,A,\bar{\varepsilon}}^* - \zeta_{N,A}^*\right| \le C||\eta_1|| \int_{B_{n,A}(c_n;\mathcal{C})} \left\|\sqrt{nh^d}\mathbf{z}_{N,\tau}^*(x)\right\|^{p-1} dQ(x,\tau).$$

As in the proof of Step 1 in the proof of Lemma C8, we deduce that

$$\mathbf{E}^*\left[\left|\zeta_{N,A,\bar{\varepsilon}}^* - \zeta_{N,A}^*\right|^2\right] \le C\bar{\varepsilon}\int_{B_{n,A}(c_n;\mathcal{C})} \mathbf{E}^*\left\|\sqrt{nh^d}\mathbf{z}_{N,\tau}^*(x)\right\|^{2p-2} dQ(x,\tau).$$

Hence for some $C_1, C_2 > 0$,

$$\begin{aligned} &\mathbf{E}\left(\mathbf{E}^*\left[\left|\zeta_{N,A,\bar{\varepsilon}}^* - \zeta_{N,A}^*\right|^2\right]\right) \\ \le\ & C\bar{\varepsilon}\int_{B_{n,A}(c_n;\mathcal{C})} \mathbf{E}\left(\mathbf{E}^*\left\|\sqrt{nh^d}\mathbf{z}_{N,\tau}^*(x)\right\|^{2p-2}\right) dQ(x,\tau) \le C_2\bar{\varepsilon} \end{aligned} \tag{D.10}$$

by the second statement of Lemma D5.

On the other hand, observe that $\mathbf{E}^* U_n^{*2} \leq 1$. Hence

$$P\left\{|Cov^*(S_{n,\bar{\varepsilon}}^* - S_n^*, U_n^*)| > M\sqrt{\bar{\varepsilon}}\right\} \leq |\mathcal{N}_J| \cdot P\left\{\max_{A \in \mathcal{N}_J} \mathbf{E}^*\left[\left|\zeta_{N,A,\bar{\varepsilon}}^* - \zeta_{N,A}^*\right|^2\right] > M^2 \bar{\varepsilon}\right\}.$$

By Chebychev's inequality, the last probability is bounded by (for some $C > 0$ that does not depend on $P \in \mathcal{P}$)

$$M^{-2}\bar{\varepsilon}^{-1} \sum_{A \in \mathcal{N}_J} \mathbf{E}\left(\mathbf{E}^*\left[\left|\zeta_{N,A,\bar{\varepsilon}}^* - \zeta_{N,A}^*\right|^2\right]\right) \leq CM^{-2},$$

by (D.10). Hence we obtain the desired result.

Proof of Step 2: Let $\tilde{\Sigma}_{2n,\tau,\bar{\varepsilon}}^*$ be the covariance matrix of $[(q_{n,\tau}^*(x) + \eta_1)^\top, \tilde{U}_n^*]^\top$ under P^*, where $\tilde{U}_n^* = U_n^* / \sqrt{P\{X \in \mathcal{C}\}}$. Using Lemma D4 and following the same arguments in (C.32), we find that

$$\sup_{(x,\tau) \in \mathcal{S}} \sup_{P \in \mathcal{P}} \mathbf{E}\left[\mathbf{E}^*\left[q_{n,\tau,j}^*(x) \tilde{U}_n^*\right]\right] \leq C_2 h^{d/2},$$

for some $C_2 > 0$. Therefore, using this result and following the proof of Step 3 in the proof of Lemma C8, we deduce that (everywhere)

$$\lambda_{\min}\left(\tilde{\Sigma}_{2n,\tau,\bar{\varepsilon}}^*\right) \geq \bar{\varepsilon} - \left\|A_{n,\tau}^*(x)\right\|, \tag{D.11}$$

for some random matrix $A_{n,\tau}^*(x)$ such that

$$\sup_{(x,\tau) \in \mathcal{S}} \sup_{P \in \mathcal{P}} \mathbf{E}\left[\left\|A_{n,\tau}^*(x)\right\|\right] = O(h^{d/2}).$$

Hence by (D.11),

$$\begin{aligned} & \inf_{(x,\tau) \in \mathcal{S}} \inf_{P \in \mathcal{P}} P\left\{\lambda_{\min}\left(\tilde{\Sigma}_{2n,\tau,\bar{\varepsilon}}^*\right) \geq \bar{\varepsilon}/2\right\} \\ \geq \; & \inf_{(x,\tau) \in \mathcal{S}} \inf_{P \in \mathcal{P}} P\left\{\left\|A_{n,\tau}^*(x)\right\| \leq \bar{\varepsilon}/2\right\} \\ \geq \; & 1 - \frac{2}{\bar{\varepsilon}} \sup_{(x,\tau) \in \mathcal{S}} \sup_{P \in \mathcal{P}} \mathbf{E}\left[\left\|A_{n,\tau}^*(x)\right\|\right] \to 1, \end{aligned} \tag{D.12}$$

as $n \to \infty$.

Now note that

$$\left(q_{n,\tau,j}^*(x), \tilde{U}_n^*\right) \stackrel{d^*}{=} \left(\frac{1}{\sqrt{n}} \sum_{k=1}^n q_{n,\tau,j}^{(k)*}(x), \frac{1}{\sqrt{n}} \sum_{k=1}^n \tilde{U}_n^{(k)*}\right),$$

where $(q_{n,\tau,j}^{(k)*}(x), \tilde{U}_n^{(k)*})$'s with $k = 1, \cdots, n$ are i.i.d. copies of $(q_{n,\tau,j}^*(x), \bar{U}_n^*)$, and

$$\bar{U}_n^* \equiv \frac{1}{\sqrt{nP\{X \in \mathcal{C}\}}} \left\{\sum_{1 \leq i \leq N_1} 1\{X_i^* \in \mathcal{C}\} - P^*\{X_i^* \in \mathcal{C}\}\right\}.$$

Note also that by Rosenthal's inequality,

$$\text{limsup}_{n\to\infty} \sup_{P\in\mathcal{P}} P\left\{\mathbf{E}^*\left[|\tilde{U}_n^{(k)*}|^3\right] > M\right\} \to 0,$$

as $M \to \infty$. Define

$$W_{n,\tau}^*(x;\eta_1) \equiv \tilde{\Sigma}_{2n,\tau,\bar{\varepsilon}}^{*-1/2} \left[\begin{array}{c} q_{n,\tau}^*(x) + \eta_1 \\ \tilde{U}_n^* \end{array}\right].$$

Using (D.12) and Lemma D5, and following the same arguments in the proof of Step 2 in the proof of Lemma C8, we deduce that

$$\text{limsup}_{n\to\infty} \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} P\left\{\mathbf{E}^*\left\|W_{n,\tau}^*(x;\eta_1)\right\|^3 > M\bar{\varepsilon}^{-3/2}h^{-d/2}\right\} \to 0,$$

as $M \to \infty$. For any vector $\mathbf{v} = [\mathbf{v}_1^\top, v_2]^\top \in \mathbf{R}^{J+1}$, we define

$$\tilde{D}_{n,\tau,p}(\mathbf{v}) \equiv \Lambda_p\left(\left[\tilde{\Sigma}_{2n,\tau,\bar{\varepsilon}}^{*1/2}\mathbf{v}\right]_1\right)\left[\tilde{\Sigma}_{2n,\tau,\bar{\varepsilon}}^{*1/2}\mathbf{v}\right]_2,$$

where $[a]_1$ of a vector $a \in \mathbf{R}^{J+1}$ indicates the vector of the first J entries of a, and $[a]_2$ the last entry of a. By Theorem 1 of Sweeting (1977), we find that (with $\bar{\varepsilon} > 0$ fixed)

$$\mathbf{E}^*\left[\tilde{D}_{n,\tau,p}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{n,\tau}^{(i)*}(x;\eta_1)\right)\right] = \mathbf{E}\left[\tilde{D}_{n,\tau,p}\left(\mathbb{Z}_{J+1}\right)\right] + O_P(n^{-1/2}h^{-d/2}),\ \mathcal{P}\text{-uniformly},$$

where $\mathbb{Z}_{J+1} \sim N(0, I_{J+1})$ and $W_{n,\tau}^{(i)*}(x;\eta_1)$'s are i.i.d. copies of $W_{n,\tau}^*(x;\eta_1)$ under P^*. Since $O(n^{-1/2}h^{-d/2}) = o(h^{d/2})$,

$$Cov^*\left(\Lambda_{A,p}\left(\sqrt{nh^d}\mathbf{z}_{N,\tau}^*(x;\eta_1)\right), U_n^*\right) = \mathbf{E}^*\left[\tilde{D}_{n,\tau,p}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{n,\tau}^{(i)*}(x)\right)\right] + o_P(h^{d/2}),$$

uniformly in $P \in \mathcal{P}$, and that $\mathbf{E}^*[\tilde{D}_{n,\tau,p}\left(\mathbb{Z}_{J+1}\right)] = 0$, we conclude that

$$\text{(D.13)} \qquad \sup_{(x,\tau)\in\mathcal{S}}\left|Cov^*\left(\Lambda_{A,p}\left(\sqrt{nh^d}\mathbf{z}_{N,\tau}^*(x;\eta_1)\right), U_n^*\right)\right| = o_P(h^{d/2}),$$

uniformly in $P \in \mathcal{P}$.

Now for some $C > 0$,

$$P\left\{\left|Cov(S_{n,\bar{\varepsilon}}^*, U_n^*)\right| > ah^{d/2}\right\} \le P\left\{C\sup_{(x,\tau)\in\mathcal{S}}\left|Cov^*\left(\Lambda_{A,p}\left(\sqrt{nh^d}\mathbf{z}_{N,\tau}^*(x;\eta_1)\right), U_n^*\right)\right| > ah^{d/2}\right\}.$$

The last probability vanishes uniformly in $P \in \mathcal{P}$ by (D.13). By applying the Dominated Convergence Theorem, we obtain (Step 2).

Proof of Step 3: First, we show that

$$\text{(D.14)} \qquad Var^*\left(S_n^*\right) = \sigma_n^2(\mathcal{C}) + o_P(1),$$

where $o_P(1)$ is uniform over $P \in \mathcal{P}$. Note that

$$Var^*\left(S_n^*\right) = \sum_{A \in \mathcal{N}_J} \sum_{A' \in \mathcal{N}_J} \mu_A \mu_{A'} Cov^*(\psi_{n,A}^*, \psi_{n,A'}^*),$$

where $\psi_{n,A}^* \equiv h^{-d/2}(\zeta_{N,A}^* - \mathbf{E}^* \zeta_{N,A}^*)$. By Lemma D6, we find that for $A, A \in \mathcal{N}_J$,

$$Cov^*(\psi_{n,A}^*, \psi_{n,A'}^*) = \sigma_{n,A,A'}(B_{n,A}(c_n;\mathcal{C}), B_{n,A'}(c_n;\mathcal{C})) + o_P(1),$$

uniformly in $P \in \mathcal{P}$, yielding the desired result of (D.14).

Combining Steps 1 and 2, we deduce that for some $C > 0$,

$$\sup_{P \in \mathcal{P}} |Cov^*(S_n^*, U_n^*)| \le \sqrt{\bar{\varepsilon}} \cdot O_P(1) + o_P(h^{d/2}).$$

Let $\tilde{\sigma}_1^2 \equiv Var^*(S_n^*)$ and $\tilde{\sigma}_2^2 \equiv 1 - \tilde{\alpha}_P$, where $\tilde{\alpha}_P \equiv P^*\left\{X_i^* \in \mathbf{R}^d \backslash \mathcal{C}\right\}$. Observe that

$$\tilde{\sigma}_1^2 = \sigma_n(\mathcal{C}) + o_P(1) > C_1 + o_P(1),\ \mathcal{P}\text{-uniformly},$$

for some $C_1 > 0$ that does not depend on n or P by the assumption of the lemma. Also note that

$$\tilde{\alpha}_P = \alpha_P + o_P(1) < 1 - C_2 + o_P(1),\ \mathcal{P}\text{-uniformly},$$

for some $C_2 > 0$. Therefore, following the same arguments as in (C.37), we obtain the desired result.

Proof of Step 4: We take $\{R_{n,\mathbf{i}} : \mathbf{i} \in \mathbb{Z}^d\}$, and define

$$\begin{aligned} B_{A,x}(c_n) &\equiv \left\{\tau \in \mathcal{T} : (x,\tau) \in B_A(c_n)\right\}, \\ B_{n,\mathbf{i}} &\equiv R_{n,\mathbf{i}} \cap \mathcal{C}, \\ B_{n,A,\mathbf{i}}(c_n) &\equiv (B_{n,\mathbf{i}} \times \mathcal{T}) \cap B_A(c_n), \end{aligned}$$

and $\mathcal{I}_n \equiv \{\mathbf{i} \in \mathbb{Z}_n^d : B_{n,\mathbf{i}} \neq \varnothing\}$ as in the proof of Step 4 in the proof of Lemma C8. Also, define

$$\Delta_{n,A,\mathbf{i}}^* \equiv h^{-d/2} \int_{B_{n,\mathbf{i}}} \int_{B_{A,x}(c_n)} \left\{\Lambda_{A,p}(\mathbf{z}_{N,\tau}^*(x)) - \mathbf{E}^*\left[\Lambda_{A,p}(\mathbf{z}_{N,\tau}^*(x))\right]\right\} d\tau dx.$$

Also, define

$$\begin{aligned} \alpha_{n,\mathbf{i}}^* &\equiv \frac{\sum_{A \in \mathcal{N}_J} \mu_A \Delta_{n,A,\mathbf{i}}^*}{\sqrt{Var^*\left(S_n^*\right)}} \text{ and} \\ u_{n,\mathbf{i}}^* &\equiv \frac{1}{\sqrt{n}} \left\{\sum_{i=1}^N 1\left\{X_i^* \in B_{n,\mathbf{i}}\right\} - nP^*\{X_i^* \in B_{n,\mathbf{i}}\}\right\} \end{aligned}$$

and write

$$\frac{S_n^*}{\sqrt{Var^*\left(S_n^*\right)}} = \sum_{\mathbf{i} \in \mathcal{I}_n} \alpha_{n,\mathbf{i}}^* \text{ and } U_n^* = \sum_{\mathbf{i} \in \mathcal{I}_n} u_{n,\mathbf{i}}^*.$$

By the properties of Poisson processes, one can see that the array $\{(\alpha_{n,\mathbf{i}}^*, u_{n,\mathbf{i}}^*)\}_{\mathbf{i}\in\mathcal{I}_n}$ is an array of 1-dependent random field under P^*. For any $q = (q_1, q_2) \in \mathbf{R}^2\backslash\{0\}$, let $y_{n,\mathbf{i}}^* \equiv q_1\alpha_{n,\mathbf{i}}^* + q_2 u_{n,\mathbf{i}}^*$ and write

$$Var^*\left(\sum_{\mathbf{i}\in\mathcal{I}_n} y_{n,\mathbf{i}}^*\right) = q_1^2 + q_2^2(1-\tilde{\alpha}_P) + 2q_1q_2\tilde{c}_{n,P},$$

uniformly over $P \in \mathcal{P}$, where $\tilde{c}_{n,P} = Cov^*(S_n^*, U_n^*)$. On the other hand, following the proof of Lemma A8 of Lee, Song, and Whang (2013) using Lemma D4, we deduce that

$$\sum_{\mathbf{i}\in\mathcal{I}_n} \mathbf{E}^*|y_{n,\mathbf{i}}^*|^r = o_P(1),\ \mathcal{P}\text{-uniformly}, \tag{D.15}$$

as $n \to \infty$, for any $r \in (2, (2p+2)/p]$, uniformly over $P \in \mathcal{P}$. By Theorem 1 of Shergin (1993), we have

$$\begin{aligned}&\sup_{t\in\mathbf{R}}\left|P^*\left\{\frac{1}{\sqrt{q_1^2 + q_2^2(1-\tilde{\alpha}_P) + 2q_1q_2\tilde{c}_{n,P}}}\sum_{\mathbf{i}\in\mathcal{I}_n} y_{n,\mathbf{i}}^* \le t\right\} - \Phi^*(t)\right| \\ \le\ & \frac{C}{\{q_1^2 + q_2^2(1-\tilde{\alpha}_P) + 2q_1q_2\tilde{c}_{n,P}\}^{r/2}}\left\{\sum_{\mathbf{i}\in\mathcal{I}_n}\mathbf{E}^*|y_{n,\mathbf{i}}^*|^r\right\}^{1/2} = o_P(1),\end{aligned}$$

for some $C > 0$ uniformly in $P \in \mathcal{P}$, by (D.15). By Lemma C2(i), we have for each $t \in \mathbf{R}$ and $q \in \mathbf{R}^2\backslash\{\mathbf{0}\}$ as $n \to \infty$,

$$\left|\mathbf{E}^*\left[\exp\left(it\frac{q^\top H_n^*}{\sqrt{q^\top \tilde{C}_n q}}\right)\right] - \exp\left(-\frac{t^2}{2}\right)\right| = o_P(1),$$

uniformly in $P \in \mathcal{P}$. Thus by Lemma C2(ii), for each $t \in \mathbf{R}^2$, we have

$$\left|P^*\left\{\tilde{C}_n^{-1/2}H_n^* \le t\right\} - P\left\{\mathbb{Z} \le t\right\}\right| = o_P(1).$$

Since the limit distribution of $\tilde{C}_n^{-1/2}H_n^*$ is continuous, the convergence above is uniform in $t \in \mathbf{R}^2$.

(ii) We fix $P \in \mathcal{P}$ such that $\text{limsup}_{n\to\infty}\sigma_n^2(\mathcal{C}) = 0$. Then by (D.14) above and Lemma D6,

$$Var^*\left(S_n^*\right) = \sigma_n^2(\mathcal{C}) + o_P(1) = o_P(1).$$

Hence, we find that $S_n^* = o_{P^*}(1)$ in P. The desired result follows by applying Theorem 1 of Shergin (1993) to the sum $U_n^* = \sum_{\mathbf{i}\in\mathcal{I}_n} u_{n,\mathbf{i}}^*$, and then applying Lemma C2. ∎

Lemma D9. *Let $\mathcal{C}$ be the Borel set in Lemma D8.*

(i) Suppose that the conditions of Lemma D8(i) are satisfied. Then for each $a > 0$, as $n \to \infty$,

$$\sup_{P \in \mathcal{P}} P \left\{ \sup_{t \in \mathbf{R}} \left| P \left\{ \frac{h^{-d/2} \sum_{A \in \mathcal{N}_J} \mu_A \left\{ \zeta_{n,A}^* - \mathbf{E}^* \zeta_{N,A}^* \right\}}{\sigma_n(\mathcal{C})} \le t \right\} - \Phi(t) \right| > a \right\} \to 0.$$

(ii) Suppose that the conditions of Lemma D8(ii) are satisfied. Then for each $a > 0$, as $n \to \infty$,

$$\sup_{P \in \mathcal{P}} P \left\{ \left| h^{-d/2} \sum_{A \in \mathcal{N}_J} \mu_A \left\{ \zeta_{n,A}^* - \mathbf{E}^* \zeta_{N,A}^* \right\} \right| > a \right\} \to 0.$$

Proof of Lemma D9. The proofs are precisely the same as those of Lemma C9, except that we use Lemma D8 instead of Lemma C8 here. ∎

Lemma D10. *Suppose that the conditions of Lemma C5 hold. Then for any small $\nu > 0$, there exists a positive sequence $\varepsilon_n = o(h^d)$ such that for all $r \in [2, M/2]$ (with $M > 0$ being as in Assumption A6(i)),*

$$\sup_{(x,\tau) \in \mathcal{S}} \sup_{P \in \mathcal{P}} \mathbf{E} ||\Sigma_{n,\tau,\varepsilon_n}^{-1/2}(x) q_{n,\tau}(x; \eta_n)||^r = O\left(h^{-(r-2)\left(\frac{M-1}{M-2}\right)d - \nu} \right),$$

where $\eta_n \in \mathbf{R}^J$ is distributed as $N(0, \varepsilon_n I_J)$ and independent of $((Y_i^\top, X_i^\top)_{i=1}^\infty, N)$ in the definition of $q_{n,\tau}(x)$, and

$$\Sigma_{n,\tau,\varepsilon_n}(x) \equiv \Sigma_{n,\tau,\tau}(x, 0) + \varepsilon_n I_J \text{ and } q_{n,\tau}(x; \eta_n) \equiv q_{n,\tau}(x) + \eta_n. \tag{D.16}$$

Suppose furthermore that $\lambda_{\min}(\Sigma_{n,\tau,\tau}(x, 0)) > c > 0$ for some $c > 0$ that does not depend on n or $P \in \mathcal{P}$. Then

$$\sup_{(x,\tau) \in \mathcal{S}} \sup_{P \in \mathcal{P}} \mathbf{E} ||\Sigma_{n,\tau,\varepsilon_n}^{-1/2}(x) q_{n,\tau}(x; \eta_n)||^r = O\left(h^{-(r-2)d/2} \right).$$

Proof of Lemma D10. We first establish the following fact.

Fact: Suppose that W is a random vector such that $\mathbf{E}||W||^2 \le c_W$ for some constant $c_W > 0$. Then, for any $r \ge 2$ and a positive integer $m \ge 1$,

$$\mathbf{E}\left[||W||^r\right] \le C_m \left(\mathbf{E}\left[||W||^{a_m(r)}\right] \right)^{1/(2^m)},$$

where $a_m(r) = 2^m(r-2) + 2$, and $C_m > 0$ is a constant that depends only on m and c_W.

Proof of Fact: The result follows by repeated application of Cauchy-Schwarz inequality:

$$\mathbf{E}||W||^r \le \left(\mathbf{E}||W||^{2(r-1)} \right)^{1/2} \left(\mathbf{E}||W||^2 \right)^{1/2} \le c_W^{1/2} \left(\mathbf{E}||W||^{2(r-1)} \right)^{1/2},$$

where we replace r on the left hand side by $2(r-1)$, and repeat the procedure to obtain Fact.

Let us consider the first statement of the lemma. Using Fact, we take a small $\nu_1 > 0$ and

$\varepsilon_n = h^{d+\nu_1}$, and choose a largest integer $m \geq 1$ such that $a_m(r) \leq M$. Such an m exists because $2 \leq r \leq M/2$. We bound

$$\mathbf{E}||\Sigma_{n,\tau,\varepsilon_n}^{-1/2}(x)q_{n,\tau}(x;\eta_n)||^r \leq C_m \left(\mathbf{E}||\Sigma_{n,\tau,\varepsilon_n}^{-1/2}(x)q_{n,\tau}(x;\eta_n)||^{a_m(r)}\right)^{1/(2^m)}.$$

By Lemma C5, we find that

$$\begin{aligned} &\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\mathbf{E}||\Sigma_{n,\tau,\varepsilon_n}^{-1/2}(x)q_{n,\tau}(x;\eta_n)||^{a_m(r)} \\ \leq\ &\sup_{(x,\tau)\in\mathcal{S}}\sup_{P\in\mathcal{P}}\lambda_{\max}^{a_m(r)/2}\left(\Sigma_{n,\tau,\varepsilon_n}^{-1}(x)\right)\mathbf{E}||q_{n,\tau}(x;\eta_n)||^{a_m(r)} \\ \leq\ &\lambda_{\min}^{-a_m(r)/2}\left(\varepsilon_n I_J\right)h^{(1-(a_m(r)/2))d}. \end{aligned} \tag{D.17}$$

By the definition of $\varepsilon_n = h^{d+\nu_1}$,

$$\varepsilon_n^{-a_m(r)/2}h^{(1-(a_m(r)/2))d} = h^{(1-a_m(r))d-a_m(r)\nu_1/2}.$$

We conclude that

$$\begin{aligned} \mathbf{E}||\Sigma_{n,\tau,\varepsilon_n}^{-1/2}(x)q_{n,\tau}(x;\eta_n)||^r &\leq C_m\left(h^{(1-a_m(r))d-a_m(r)\nu_1/2}\right)^{1/2^m} \\ &= C_m\left(h^{(-1-2^m(r-2))d-(2^m(r-2)+2)\nu_1/2}\right)^{1/2^m} \\ &= C_m h^{(-2^{-m}-(r-2))d-((r-2)+2^{-m+1})\nu_1/2}. \end{aligned}$$

Since $a_m(r) \leq M$, or $2^{-m} \geq (r-2)/(M-2)$, the last term is bounded by

$$C_m h^{-(r-2)\left(\frac{M-1}{M-2}\right)d-\left((r-2)+\frac{2(r-2)}{M-2}\right)\nu_1/2}.$$

By taking ν_1 small enough, we obtain the desired result.

Now, let us turn to the second statement of the lemma. Since, under the additional condition,

$$\lambda_{\max}^{a_m(r)/2}\left(\Sigma_{n,\tau,\varepsilon_n}^{-1}(x)\right) < c^{-a_m(r)/2},$$

the last bound in (D.17) turns out to be

$$c^{-a_m(r)/2}h^{(1-(a_m(r)/2))d}.$$

Therefore, we conclude that

$$\begin{aligned} \mathbf{E}||\Sigma_{n,\tau,\varepsilon_n}^{-1/2}(x)q_{n,\tau}(x;\eta_n)||^r &\leq C_m\left(c^{-a_m(r)/2}h^{(1-(a_m(r)/2))d}\right)^{1/2^m} \\ &= C_m c^{-\{(r-2)+2^{1-m}\}/2}h^{(2^{-m}-\{(r-2)+2^{1-m}\}/2)d} \\ &= C_m c^{-\{(r-2)+2^{1-m}\}/2}h^{-(r-2)d/2}. \end{aligned}$$

Again, using the inequality $2^{-m} \geq (r-2)/(M-2)$, we obtain the desired result. ∎

Lemma D11. *Suppose that the conditions of Lemma D5 hold. Then for any small $\nu > 0$, there exists a positive sequence $\varepsilon_n = o(h^d)$ such that for all $r \in [2, M/2]$ (with $M > 0$ being as in Assumption A6(i)),*

$$\sup_{(x,\tau)\in\mathcal{S}} \mathbf{E}^* ||\tilde{\Sigma}_{n,\tau,\varepsilon_n}^{-1/2}(x) q_{n,\tau}^*(x;\eta_n)||^r = O_P\left(h^{-(r-2)\left(\frac{M-1}{M-2}\right)d-\nu}\right), \text{ uniformly in } P \in \mathcal{P},$$

where $\eta_n \in \mathbf{R}^J$ is distributed as $N(0, \varepsilon_n I_J)$ and independent of $((Y_i^{*\top}, X_i^{*\top})_{i=1}^n, (Y_i^\top, X_i^\top)_{i=1}^n, N)$ in the definition of $q_{n,\tau}^*(x)$, and

$$\tilde{\Sigma}_{n,\tau,\varepsilon_n}(x) \equiv \tilde{\Sigma}_{n,\tau,\tau}(x,0) + \varepsilon_n I_J.$$

Suppose furthermore that

$$\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} P\left\{\lambda_{\min}(\tilde{\Sigma}_{n,\tau,\tau}(x,0)) > c\right\} \to 0,$$

for some $c > 0$ that does not depend on n or $P \in \mathcal{P}$. Then

$$\sup_{(x,\tau)\in\mathcal{S}} \mathbf{E}^* ||\tilde{\Sigma}_{n,\tau,\varepsilon_n}^{-1/2}(x) q_{n,\tau}^*(x;\eta_n)||^r = O_P\left(h^{-(r-2)d/2}\right), \text{ uniformly in } P \in \mathcal{P}.$$

Proof of Lemma D11. The proof is precisely the same as that of Lemma D10, where we use Lemma D5 instead of Lemma C5. ∎

We let for a sequence of Borel sets B_n in $\mathcal{S}$ and $\lambda \in \{0, d/4, d/2\}$, $A \subset \mathbb{N}_J$, and a fixed bounded function δ on $\mathcal{S}$,

$$\begin{aligned}
a_n^R(B_n) &\equiv \int_{B_n} \mathbf{E}\left[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x) + h^\lambda \delta(x,\tau))\right] dQ(x,\tau) \\
a_n^{R*}(B_n) &\equiv \int_{B_n} \mathbf{E}^*\left[\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}^*(x) + h^\lambda \delta(x,\tau))\right] dQ(x,\tau), \text{ and} \\
a_n(B_n) &\equiv \int_{B_n} \mathbf{E}\left[\Lambda_{A,p}(\mathbb{W}_{n,\tau,\tau}^{(1)}(x,0) + h^\lambda \delta(x,\tau))\right] dQ(x,\tau),
\end{aligned}$$

where $\mathbf{z}_{N,\tau}^*(x)$ is a random vector whose j-th entry is given by

$$z_{N,\tau,j}^*(x) \equiv \frac{1}{nh^d}\sum_{i=1}^N \beta_{n,x,\tau,j}(Y_{ij}^*, (X_i^* - x)/h) - \frac{1}{h^d}\mathbf{E}^*\left[\beta_{n,x,\tau,j}(Y_{ij}^*, (X_i^* - x)/h)\right].$$

Lemma D12. *Suppose that the conditions of Lemmas D10 and D11 hold and that*

$$n^{-1/2} h^{-\left(\frac{3M-4}{2M-4}\right)d-\nu} \to 0,$$

as $n \to \infty$, for some small $\nu > 0$. Then for any sequence of Borel sets B_n in $\mathcal{S}$,

$$\begin{aligned}
\sup_{P\in\mathcal{P}} \left|a_n^R(B_n) - a_n(B_n)\right| &= o(h^{d/2}) \text{ and} \\
\sup_{P\in\mathcal{P}} P\left\{\left|a_n^{R*}(B_n) - a_n(B_n)\right| > ah^{d/2}\right\} &= o(1).
\end{aligned}$$

Proof of Lemma D12. For the statement, it suffices to show that uniformly in $P \in \mathcal{P}$,

$$\sup_{(x,\tau)\in\mathcal{S}} \left| \begin{array}{c} \mathbf{E}\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x) + h^\lambda\delta(x,\tau)) \\ -\mathbf{E}\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0) + h^\lambda\delta(x,\tau)) \end{array} \right| = o(h^{d/2}), \tag{D.18}$$

$$\sup_{(x,\tau)\in\mathcal{S}} \left| \begin{array}{c} \mathbf{E}^*\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}^*_{N,\tau}(x) + h^\lambda\delta(x,\tau)) \\ -\mathbf{E}\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau,\tau}(x,0) + h^\lambda\delta(x,\tau)) \end{array} \right| = o_P(h^{d/2}),$$

uniformly in $P \in \mathcal{P}$. We prove the first statement of (D.18). The proof of the second statement of (D.18) can be done in a similar way.

Take small $\nu > 0$. We apply Lemma D10 by choosing a positive sequence $\varepsilon_n = o(h^d)$ such that for any $r \in [2, M/2]$,

$$\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}||\Sigma^{-1/2}_{n,\tau,\varepsilon_n}(x) q_{n,\tau}(x;\eta_n)||^r = O\left(h^{-(r-2)\left(\frac{M-1}{M-2}\right)d-\nu}\right), \tag{D.19}$$

where $q_{n,\tau}(x;\eta_n)$ and $\Sigma_{n,\tau,\varepsilon_n}(x)$ are as in Lemma D10. We follow the arguments in the proof of Step 2 in Lemma C6 to bound the left-hand side in the first supremum in (D.18) by

$$\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \left| \mathbf{E}\Lambda_{A,p}(\sqrt{nh^d}\mathbf{z}_{N,\tau}(x;\eta_n) + h^\lambda\delta(x,\tau)) - \mathbf{E}\Lambda_{A,p}(\mathbb{W}^{(1)}_{n,\tau,\tau,\varepsilon_n}(x,0) + h^\lambda\delta(x,\tau)) \right| + C\sqrt{\varepsilon_n},$$

for some $C > 0$, where

$$\mathbf{z}_{N,\tau}(x;\eta_n) \equiv \mathbf{z}_{N,\tau}(x) + \eta_n/\sqrt{nh^d},$$

and $\mathbb{W}^{(1)}_{n,\tau,\tau,\varepsilon_n}(x,0)$ is as defined in (C.17). Let

$$\begin{aligned} \xi_{N,\tau}(x;\eta_n) &\equiv \sqrt{nh^d}\Sigma^{-1/2}_{n,\tau,\varepsilon_n}(x) \cdot \mathbf{z}_{N,\tau}(x;\eta_n) \text{ and} \\ \mathbb{Z}^{(1)}_{n,\tau,\tau,\varepsilon_n}(x,0) &\equiv \Sigma^{-1/2}_{n,\tau,\varepsilon_n}(x) \cdot \mathbb{W}^{(1)}_{n,\tau,\tau,\varepsilon_n}(x,0). \end{aligned}$$

We rewrite the previous absolute value as

$$\sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \left| \mathbf{E}\Lambda^\Sigma_{A,n,p}(\sqrt{nh^d}\xi_{N,\tau}(x;\eta_n)) - \mathbf{E}\Lambda^\Sigma_{n,p}(\mathbb{Z}^{(1)}_{n,\tau,\tau,\varepsilon_n}(x,0)) \right|, \tag{D.20}$$

where $\Lambda^\Sigma_{A,n,p}(\mathbf{v}) \equiv \Lambda_{A,p}(\Sigma^{1/2}_{n,\tau,\varepsilon_n}(x)\mathbf{v} + h^\lambda\delta(x,\tau))$. Note that the condition for M in Assumption A6(i) that $M \geq 2(p+2)$, we can choose $r = \max\{p, 3\}$. Then $r \in [2, M/2]$ as required. Using Theorem 1 of Sweeting (1977), we bound the above supremum by (with $r = \max\{p, 3\}$)

$$\begin{aligned} &\frac{C_1}{\sqrt{n}} \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}||\Sigma^{-1/2}_{n,\tau,\varepsilon_n}(x) q_{n,\tau}(x;\eta_n)||^3 \\ &+ \frac{C_2}{\sqrt{n^{r-2}}} \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}||\Sigma^{-1/2}_{n,\tau,\varepsilon_n}(x) q_{n,\tau}(x;\eta_n)||^r \\ &+ C_3 \sup_{(x,\tau)\in\mathcal{S}} \sup_{P\in\mathcal{P}} \mathbf{E}\omega_{n,p}\left(\mathbb{Z}^{(1)}_{n,\tau,\tau,\varepsilon_n}(x,0); \frac{C_4}{\sqrt{n}} \mathbf{E}||\Sigma^{-1/2}_{n,\tau,\varepsilon_n}(x) q_{n,\tau}(x;\eta_n)||^3\right), \end{aligned}$$

for some positive constants C_1, C_2, C_3, and C_4, where

$$\omega_{n,p}(\mathbf{v}; c) \equiv \sup\left\{|\Lambda^{\Sigma}_{A,n,p}(\mathbf{v}) - \Lambda^{\Sigma}_{A,n,p}(\mathbf{y})| : \mathbf{y} \in \mathbf{R}^{|A|}, ||\mathbf{v} - \mathbf{y}|| \leq c\right\}.$$

The proof is complete by (D.19) and by the condition $n^{-1/2}h^{-\left(\frac{3M-4}{2M-4}\right)d-\nu} \to 0$. ∎

References

Abadir, Karim, M., and J. R. Magnus (2005): *Matrix Algebra.* Cambridge University Press, New York, NY.

Acemoglu, D., and D. Autor (2011): "Skills, Tasks and Technologies: Implications for Employment and Earnings," in *Handbook of Labor Economics*, ed. by D. Card, and O. Ashenfelter, vol. 4, part B, pp. 1043–1171. Elsevier.

Anderson, G., O. Linton, and Y.-J. Whang (2012): "Nonparametric estimation and inference about the overlap of two distributions," *Journal of Econometrics*, 171(1), 1–23.

Andrews, D. W. K. (1994): "Empirical Process Method in Econometrics," in *The Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, pp. 2247–2294. North-Holland.

Andrews, D. W. K., and P. J. Barwick (2012): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *Econometrica*, 80(6), 2805–2826.

Andrews, D. W. K., S. T. Berry, and P. Jia (2004): "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," working paper, Cowles Foundation.

Andrews, D. W. K., and P. Guggenberger (2009): "Validity of Subsampling and Plug-In Asymptotic Inference for Parameters Defined by Moment Inequalities," *Econometric Theory*, 25, 669–709.

Andrews, D. W. K., and X. Shi (2009): "Inference Based on Conditional Moment Inequalities," Working Paper, Yale Unversity.

——— (2013): "Inference Based on Conditional Moment Inequalities," *Econometrica*, 81(2), 609–666.

——— (2014): "Nonparametric inference based on conditional moment inequalities," *Journal of Econometrics*, 179(1), 31–45.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78(1), 119–157.

Aradillas-López, A., A. Gandhi, and D. Quint (2013a): "Identification and Inference in Ascending Auctions With Correlated Private Values," *Econometrica*, 81(2), 489–534.

——— (2013b): "A Simple Test for Moment Inequality Models with an Application to English Auctions," Working Paper, University of Wisconsin, Madison and Penn State University.

Aradillas-López, A., and A. Rosen (2013): "Inference in ordered response games with complete information," CeMMAP Working Papers, CWP33/1.

Aradillas-López, A., and E. Tamer (2008): "The Identification Power of Equilibrium in Simple Games," *Journal of Business & Economic Statistics*, 26(3), 261–283.

Armstrong, T. B. (2011a): "Asymptotically Exact Inference in Conditional Moment Inequality Models," working paper, Stanford University.

——— (2011b): "Weighted KS Statistics for Inference on Conditional Moment Inequalities," Working Paper, Stanford University.

Armstrong, T. B., and H. P. Chan (2013): "Multiscale Adaptive Inference on Conditional Moment Inequalities," Discussion Paper 1885, Cowles Foundation.

Baraud, Y., S. Huet, and B. Laurent (2005): "Testing convex hypotheses on the mean of a Gaussian vector. Application to testing qualitative hypotheses on a regression function," *Annals of Statistics*, 33(1), 214–257.

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.

Beresteanu, A., I. Molchanov, and F. Molinari (2011): "Sharp Identification Regions in Models With Convex Moment Predictions," *Econometrica*, 79(6), 1785–1821.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

Berry, S. T., and E. Tamer (2007): "Identification in Models of Oligopoly Entry," in *Advances in Econometrics, Ninth World Congress*, ed. by R. Blundell, W. Newey, and T. Persson, vol. 2, pp. 46–85. Cambridge University Press.

Biau, G., B. Cadre, D. Mason, and B. Pelletier (2009): "Asymptotic normality in density support estimation," *Electronic Journal of Probability*, 14(91), 2617–2635.

Bickel, P. J., Y. Ritov, and T. M. Stoker (2006): "Tailor-made tests for goodness of fit to semiparametric hypotheses," *Annals of Statistics*, 34(2), 721–741.

Blundell, R., M. Browning, and I. Crawford (2008): "Best Nonparametric Bounds on Demand Responses," *Econometrica*, 76(6), 1227–1262.

Blundell, R., D. Kristensen, and R. Matzkin (2014): "Bounding quantile demand functions using revealed preference inequalities," *Journal of Econometrics*, forthcoming.

Bugni, F. (2010): "Bootstrap Inference for Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78(2), 735–753.

Canay, I. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156(2), 408–425.

Chaudhuri, P. (1991a): "Global nonparametric estimation of conditional quantile functions and their derivatives," *Journal of Multivariate Analysis*, 39(2), 246 – 269.

——— (1991b): "Nonparametric Estimates of Regression Quantiles and Their Local Bahadur Representation," *Annals of Statistics*, 19(2), 760–777.

Chen, X., O. Linton, and I. Van Keilegom (2003): "Estimation of semiparametric models when the criterion function is not smooth," *Econometrica*, 71(5), 1591–1608.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Chernozhukov, V., S. Lee, and A. Rosen (2013): "Intersection Bounds: Estimation and Inference," *Econometrica*, 81(2), 667–737.

Chesher, A., and A. Rosen (2012): "Simultaneous Equations Models for Discrete Outcomes: Coherence, Completeness, and Identification," CeMMAP working paper CWP21/12.

——— (2014): "Generalized Instrumental Variable Models," CeMMAP working paper CWP04/14.

Chetverikov, D. (2011): "Adaptive rate optimal test of conditional moment inequalities," working paper, MIT.

Chetverikov, D. (2012): "Testing Regression Monotonicity in Econometric Models," Working Paper, MIT.

Ciliberto, F., and E. Tamer (2009): "Market Structure and Multiple Equilibria in Airline Markets," *Econometrica*, 77(6), 1791–1828.

de Paula, A. (2013): "Econometric Analysis of Games with Multiple Equilibria," *Annual Review of Economics*, 5(1), 107–131.

Delgado, M. A., and J. C. Escanciano (2012): "Distribution-free tests of stochastic monotonicity," *Journal of Econometrics*, 170(1), 68–75.

Dümbgen, L., and V. G. Spokoiny (2001): "Multiscale testing of qualitative hypotheses," *Annals of Statistics*, 29(1), 124–152.

Durrett, R. (2010): *Probability: Theory and Examples, Fourth Edition.* Cambridge University Press, New York, NY.

Fan, J., and I. Gijbels (1996): *Local Polynomial Modelling and Its Applications.* Chapman & Hall, London, UK.

Fan, Y., and S. Park (2014): "Nonparametric inference for counterfactual means: Bias-correction, confidence sets, and weak IV," *Journal of Econometrics*, 178(Part 1), 45–56.

Feller, W. (1966): *An Introduction to Probability Theory and Its Applications, Vol. 2.* John Wiley & Sons, New York, NY.

Galichon, A., and M. Henry (2009): "A Test of Non-identifying Restrictions and Confidence Regions for Partially Identified Parameters," *Journal of Econometrics*, 159(2), 186–196.

——— (2011): "Set Identification in Models with Multiple Equilibria," *Review of Economic Studies*, 78(4), 12641298.

Ghosal, S., A. Sen, and A. W. van der Vaart (2000): "Testing Monotonicity of Regression," *Annals of Statistics*, 28, 1054–1082.

Gimenes, N., and E. Guerre (2013): "Augmented quantile regression methods for first-price auction," Working paper, Queen Mary University of London.

Giné, E. (1997): "Lectures on some aspects of the bootstrap," in *Lectures on Probability Theory and Statistics*, ed. by P. Bernard, vol. 1665 of *Lecture Notes in Mathematics*, pp. 37–151. Springer Berlin Heidelberg.

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

Giné, E., and J. Zinn (1990): "Bootstrapping General Empirical Measures," *Annals of Probability*, 18(2), 851–869.

Guerre, E., I. Perrigne, and Q. Vuong (2009): "Nonparametric Identification of Risk Aversion in First-Price Auctions Under Exclusion Restrictions," *Econometrica*, 77(4), 1193–1227.

Guerre, E., and C. Sabbah (2012): "Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function," *Econometric Theory*, 28, 87–129.

Haile, P., H. Hong, and M. Shum (2003): "Nonparametric Tests for Common Values at First- Price Sealed-Bid Auctions," Working Paper, Yale University.

Haile, P. A., and E. Tamer (2003): "Inference with an Incomplete Model of English Auctions," *Journal of Political Economy*, 111(1), 1–51.

Härdle, W., and E. Mammen (1993): "Comparing Nonparametric Versus Parametric Regression Fits," *Annals of Statistics*, 21(4), 1926–1947.

Hoderlein, S., and J. Stoye (2013): "Revealed Preferences in a Heterogeneous Population," *Review of Economics and Statistics*, forthcoming.

Horváth, L. (1991): "On L_p-Norms of Multivariate Density Estimators," *Annals of Statistics*, 19(4), 1933–1949.

Imbens, G., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72(6), 1845–1857.

Jun, S. J., Y. Lee, and Y. Shin (2011): "Testing for Distributional Treatment Effects: A Set Identification Approach," working paper.

Khan, S., M. Ponomareva, and E. Tamer (2013): "Identification of Panel Data Models with Endogenous Censoring," Working Paper, Duke University, Northern Illinois University, and Northwestern University.

Khan, S., and E. Tamer (2009): "Inference on endogenously censored regression models using conditional moment inequalities," *Journal of Econometrics*, 152(2), 104 – 119.

Kim, K. i. (2009): "Set Estimation and Inference with Models Characterized by Conditional Moment Inequalities," working paper, University of Minnesota.

Kitamura, Y., and J. Stoye (2013): "Nonparametric Analysis of Random Utility Models: Testing," Discussion Paper 1902, Cowles Foundation.

Krasnokutskaya, E., K. Song, and X. Tang (2013): "The Role of Quality in Service Markets Organized as Multi-attribute Auctions," working paper, University of Pennsylvania.

Lee, S., O. Linton, and Y.-J. Whang (2009): "Testing for Stochastic Monotonicity," *Econometrica*, 77(2), 585–602.

Lee, S., K. Song, and Y.-J. Whang (2013): "Testing Functional Inequalities," *Journal of Econometrics*, 172(1), 14–32.

Lee, S., and Y.-J. Whang (2009): "Nonparametric tests of conditional treatment effects," CeMMAP working paper CWP36/09.

Li, T., and T. Oka (2013): "Set identification of the Censored Quantile Regression Model for Short Panels with Fixed Effects," Working paper, Vanderbilt University and National University of Singapore.

Linton, O., K. Song, and Y.-J. Whang (2010): "An improved bootstrap test of stochastic dominance," *Journal of Econometrics*, 154(2), 186–202.

Lu, J., and I. Perrigne (2008): "Estimating risk aversion from ascending and sealed-bid auctions: the case of timber auction data," *Journal of Applied Econometrics*, 23(7), 871–896.

Magnus, J. R., and H. Neudecker (2001): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York, NY.

Mammen, E., I. Van Keilegom, and K. Yu (2013): "Expansion for moments of regression quantiles with application to nonparametric testing," Working paper, arXiv:1306.6179.

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

——— (2007): *Identification for Prediction and Decision*. Harvard University Press, New York.

Manski, C. F., and J. V. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68(4), 997–1010.

Marmer, V., A. Shneyerov, and P. Xu (2013): "What model for entry in first-price auctions? A nonparametric approach," *Journal of Econometrics*, 176(1), 46 – 58.

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

Massart, P. (2007): *Concentration Inequalities and Model Selection*. Springer-Verlag, Berlin Heidelberg.

Menzel, K. (2009): "Estimation and Inference with Many Weak Moment Inequalities," working paper, MIT.

Nitikin, Y. (1995): *Asymptotic Efficiency of Nonparametric Tests.* Cambridge University Press, Cambridge.

Ponomareva, M. (2010): "Inference in Models Defined by Conditional Moment Inequalities with Continuous Covariates," working paper, University of Western Ontario.

Proschan, M. A., and B. Presnell (1998): "Expect the Unexpected from Conditional Expectation," *American Statistician*, 52(3), 248–252.

Rio, E. (1994): "Local Invariance Principles and Their Application to Density Estimation," *Probability Theory and Related Fields*, 98, 21–45.

Romano, J. P., and A. M. Shaikh (2008): "Inference for Identifiable Parameters in Partially Identified Econometric Models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78(1), 169–211.

Romano, J. P., A. M. Shaikh, and M. Wolf (2013): "A Practical Two-Step Method for Testing Moment Inequalities," working paper, Stanford University, University of Chicago, and University of Zurich.

Rosen, A. M. (2008): "Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities," *Journal of Econometrics*, 146, 107–117.

——— (2012): "Set identification via quantile restrictions in short panels," *Journal of Econometrics*, 166, 127–137.

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

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

Tamer, E. (2003): "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies*, 70(1), 147–165.

——— (2010): "Partial Identification in Econometrics," *Annual Review of Economics*, 2, 167–195.

van der Vaart, A. W. (1998): *Asymptotic Statistics.* Cambridge University Press, New York, NY.