Robust Inference for Convex Pairwise Difference Estimators*

Matias D. Cattaneo† Michael Jansson‡ Kenichi Nagasawa§

August 26, 2025

Abstract

This paper develops distribution theory and bootstrap-based inference methods for a broad class of convex pairwise difference estimators. These estimators minimize a kernel-weighted convex-in-parameter function over observation pairs that are similar in terms of certain covariates, where the similarity is governed by a localization (bandwidth) parameter. While classical results establish asymptotic normality under restrictive bandwidth conditions, we show that valid Gaussian and bootstrap-based inference remains possible under substantially weaker assumptions. First, we extend the theory of small bandwidth asymptotics to convex pairwise estimation settings, deriving robust Gaussian approximations even when a smaller than standard bandwidth is used. Second, we employ a debiasing procedure based on generalized jackknifing to enable inference with larger bandwidths, while preserving convexity of the objective function. Third, we construct a novel bootstrap method that adjusts for bandwidth-induced variance distortions, yielding valid inference across a wide range of bandwidth choices. Our proposed inference method enjoys demonstrable more robustness, while retaining the practical appeal of convex pairwise difference estimators.

Keywords: small bandwidth asymptotics, generalized jackknife, U-process, pairwise comparisons, robust distribution theory.

*This paper was prepared for the Econometric Theory Lecture delivered at the 2025 International Symposium on Econometric Theory and Applications (SETA), University of Macau (China), June 1–3, 2025. Cattaneo gratefully acknowledges financial support from the National Science Foundation through grants SES-1947805, DMS-2210561, and SES-2241575. Jansson gratefully acknowledges financial support from the National Science Foundation through grant SES-1947662 and from the Aarhus Center for Econometrics (ACE) funded by the Danish National Research Foundation grant number DNRF186. Nagasawa gratefully acknowledges financial support from the British Academy through grant SRG24\241614.

†Department of Operations Research and Financial Engineering, Princeton University.

‡Department of Economics, UC Berkeley and ACE.

§Department of Economics, University of Warwick.

Contents

1 Introduction

Suppose $\mathbf{z}_1, \ldots, \mathbf{z}_n$ is a random sample from the distribution of a random vector $\mathbf{z}$. This paper studies the large-sample properties of the following *convex* pairwise difference estimator:

$$\widehat{\boldsymbol{\theta}}_n \in \operatorname*{arg\,min}_{\boldsymbol{\theta}\in\Theta} \sum_{i<j} m(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) K_{h_n}(\mathbf{w}_i - \mathbf{w}_j), \qquad K_h(\mathbf{u}) = \frac{1}{h^d} K\left(\frac{\mathbf{u}}{h}\right), \tag{1.1}$$

where $\Theta \subseteq \mathbb{R}^k$ is a parameter space, $\boldsymbol{\theta} \mapsto m(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta})$ is a *convex* function, K is a symmetric, non-negative kernel, h_n is a positive bandwidth (or localization) parameter sequence, $\mathbf{w}$ is a continuously distributed d-dimensional subvector of $\mathbf{z}$, and where $\sum_{i<j}$ denotes $\sum_{j=2}^{n}\sum_{i=1}^{j-1}$. Pairwise difference estimation, which relies on local comparisons between observation pairs, has been used to address heterogeneity in nonlinear models. See Powell (1994), Honoré and Powell (2005), and Aradillas-Lopez, Honoré, and Powell (2007) for overviews, and Section 2 for three motivating examples.

In contrast to classical extremum estimators, $\widehat{\boldsymbol{\theta}}_n$ is a local M-estimator that employs observation pairs (i, j) for which $\mathbf{w}_i$ and $\mathbf{w}_j$ are similar. The bandwidth h_n governs the degree of similarity: When $h_n \to 0$ (as $n \to \infty$), the estimator increasingly focuses on nearly identical-in-$\mathbf{w}$ pairs. In turn, focusing on such pairs is natural in settings where identification can be based on the condition $\mathbf{w}_i \approx \mathbf{w}_j$ (combined with smoothness assumptions). The localization introduces a familiar trade-off for estimation and inference: A smaller h_n reduces bias from dissimilarity between $\mathbf{w}_i$ and $\mathbf{w}_j$, but increases variance due to fewer available usable pairs. As a consequence, the large-sample behavior of $\widehat{\boldsymbol{\theta}}_n$ depends critically on a delicate bias-variance trade-off determined by h_n. This paper develops novel inference methods for convex pairwise difference estimators that are demonstrably more robust to bandwidth choice than existing methods.

Under regularity conditions and assuming that

$$nh_n^d \to \infty \qquad \text{and} \qquad nh_n^4 \to 0,$$

the pairwise difference estimator is asymptotically linear:

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{\xi}_0(\mathbf{z}_i) + o_{\mathbb{P}}(1) \rightsquigarrow \mathsf{N}(\mathbf{0}, \mathbb{E}[\boldsymbol{\xi}_0(\mathbf{z})\boldsymbol{\xi}_0(\mathbf{z})']), \tag{1.2}$$

where $\boldsymbol{\theta}_0$ is the estimand and $\boldsymbol{\xi}_0(\cdot)$ is the influence function. (The exact form of $\boldsymbol{\xi}_0$ is given below.) The condition $nh_n^d \to \infty$ lower bounds the level of localization h_n allowed for, while the condition $nh_n^4 \to 0$ upper bounds the level of localization. The purpose of the latter condition is to control a smoothing bias term. The bias condition $nh_n^4 \to 0$ could be replaced by the weaker condition $nh_n^{2L} \to 0$ if a (higher-order) kernel of order $L > 2$ were used, but a higher-order kernel annihilates the convexity of the objective function because higher-order kernels take negative values.

The main results of this paper are obtained by combining three ideas:

1. *Small Bandwidth Asymptotics.* Utilizing the framework introduced by Cattaneo, Crump, and

Jansson (2014a), we establish a more robust Gaussian distributional approximation for the pairwise difference estimator that allows for higher levels of localization by remaining valid even when the condition $nh_n^d \to \infty$ is violated. This generalized distributional approximation shows that, while the localization restriction $nh_n^d \to \infty$ is necessary for establishing asymptotic linearity, a Gaussian approximation can hold under the substantially weaker condition $n^2h_n^d \to \infty$, albeit with a convergence rate and large sample variance that depends explicitly on the level of localization used.

2. *Debiasing.* Following Honoré and Powell (2005) we debias the pairwise difference estimator using the method of *generalized jackknifing* introduced by Schucany and Sommers (1977). Doing so allows for (larger) bandwidths that violate the bias condition $nh_n^4 \to \infty$. This debiasing approach retains the convexity of the objective function, which is attractive for both theoretical (weaker regularity conditions) and practical (faster computation) reasons. The debiasing procedure combines linearly a collection of convex pairwise difference estimators constructed using different levels of localization. The resulting ensembling-based pairwise difference estimator admits a small bandwidth Gaussian approximation with an associated bias condition of the form $nh_n^{2L} \to 0$, where $L \geq 2$ denotes the order of a certain (equivalent) kernel induced by the debiasing procedure.

3. *Bootstrapping.* Building on insights in Cattaneo, Crump, and Jansson (2014b), we develop a valid bootstrap-based distributional approximation for the debiased pairwise difference estimator rescaling the localization parameter. The nonparametric bootstrap distributional approximation exhibits a mismatch in its asymptotic variance under small bandwidth asymptotics. The mismatch is characterized by a known multiplicative factor involving the localization parameter h_n. As a result, bootstrapping the (debiased) pairwise difference estimator with a different localization parameter (namely, $3^{1/d}h_n$ rather than h_n) leads to a valid bootstrap-based inference procedure also under small bandwidth asymptotics.

In combination, these three ideas enable us to offer a novel resampling-based inference method for (convex) pairwise difference estimators that are demonstrably more robust to a wider set of choices of the localization parameter h_n.

Our theoretical work is carefully developed to retain and leverage convexity of the objective function defining the pairwise difference estimator. This feature not only allows for fast implementation of the estimator and resampling-based methods, but also enables us to proceed under relative weak conditions when obtaining theoretical results. When developing our theoretical results, we rely heavily on the foundational work of Hjort and Pollard (1993) and Pollard (1991), which we apply to the case of U-processes.

This paper is connected to several strands of the literature. Contributions to the pairwise difference estimation literature include Ahn and Powell (1993), Powell, Ichimura, Powell, and Ruud (2018), Aradillas-Lopez (2012), Blundell and Powell (2004), Hong and Shum (2010), Honoré (1992), Honoré, Kyriazidou, and Udry (1997), Honoré and Powell (1994), Jochmans (2013), and Kyriazidou

(1997). The theoretical and practical features of small bandwidth asymptotics, and their connection with resampling methods for inference, are discussed in Cattaneo, Crump, and Jansson (2010), Cattaneo et al. (2014b), Cattaneo, Jansson, and Newey (2018), Cattaneo and Jansson (2018), Matsushita and Otsu (2021), Cattaneo and Jansson (2022), Cattaneo, Farrell, Jansson, and Masini (2025a), and references therein. The generalized jackknife has been successfully used for debiasing in density weighted average derivative estimation (Powell, Stock, and Stoker, 1989), asymptotically linear pairwise difference estimation (Honoré and Powell, 2005), nonlinear semiparametric estimation (Cattaneo, Crump, and Jansson, 2013), monotone estimation (Cattaneo, Jansson, and Nagasawa, 2024), and random forest estimation (Cattaneo, Klusowski, and Underwood, 2025b), among other settings. Shao and Tu (2012) give a textbook introduction to jackknifing, bootstrapping, and other resampling methods.

The rest of the paper proceeds as follows. Section 2 introduces the three motivating examples that are used throughout the paper to motivate our work and to illustrate the verification of the high-level assumptions imposed. Section 3 present our main theoretical distributional and bootstrap results for robust inference employing convex pairwise difference estimators. The proofs of these results are given in Section 4. Section 5 showcases how the high-level sufficient conditions imposed in our theoretical developments are verified for the three motivating examples. Section 6 gives final remarks.

2 Motivating Examples

We use three examples to motivate and illustrate our work. The first example involves an estimator that can be written in closed form (because it has a quadratic-in-$\boldsymbol{\theta}$ function $\boldsymbol{\theta} \mapsto m(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta})$), while the other two examples do not. The second example has a smooth-in-$\boldsymbol{\theta}$ function $m(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta})$, while the third example does not. All three examples have convex-in-$\boldsymbol{\theta}$ functions $m(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta})$ and employ the following notation: $\mathbf{z}_i = (y_i, \mathbf{x}_i', \mathbf{w}_i')'$ with y_i a scalar outcome variable, $\mathbf{x}_i$ a k-dimensional covariate, and $\mathbf{w}_i$ a d-dimensional covariate. For more details on the examples, see Powell (1994), Honoré and Powell (2005), and Aradillas-Lopez et al. (2007).

2.1 Partially Linear Regression Model

The partially linear regression model studied here is of the form

$$y_i = \mathbf{x}_i'\boldsymbol{\theta}_0 + \eta_0(\mathbf{w}_i) + \varepsilon_i, \qquad \mathbb{E}[\varepsilon_i|\mathbf{x}_i, \mathbf{w}_i] = 0,$$

where $\boldsymbol{\theta}_0$ is the parameter of interest, while $\eta_0(\cdot)$ is an unknown (nuisance) function. A pairwise difference estimator of $\boldsymbol{\theta}_0$ can be based on

$$m(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) = m_{\mathtt{PLR}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) = \frac{1}{2}\left(\dot{y}_{i,j} - \dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta}\right)^2,$$

where $\dot{y}_{i,j} = y_i - y_j$ and $\dot{\mathbf{x}}_{i,j} = \mathbf{x}_i - \mathbf{x}_j$. Setting $\Theta = \mathbb{R}^k$, the pairwise difference estimator in (1.1) admits a closed form solution (provided that a non-negative kernel function is used):

$$\widehat{\boldsymbol{\theta}}_n = \left(\sum_{i<j} \dot{\mathbf{x}}_{i,j} \dot{\mathbf{x}}_{i,j}' K_{h_n}(\mathbf{w}_i - \mathbf{w}_j) \right)^{-1} \sum_{i<j} \dot{\mathbf{x}}_{i,j} \dot{y}_{i,j} K_{h_n}(\mathbf{w}_i - \mathbf{w}_j).$$

2.2 Partially Linear Logit Model

The partially linear logit model studies here is of the form

$$y_i = \mathbb{1}\{\mathbf{x}_i'\boldsymbol{\theta}_0 + \eta_0(\mathbf{w}_i) + \varepsilon_i \geq 0\},$$

where $\boldsymbol{\theta}_0$ is the parameter of interest, $\eta_0(\cdot)$ is an unknown nuisance function, and where

$$\mathbb{P}\left[\varepsilon_i \leq u | \mathbf{x}_i, \mathbf{w}_i\right] = \Lambda(u), \qquad \Lambda(u) = \frac{\exp(u)}{1 + \exp(u)}.$$

The parameter $\boldsymbol{\theta}_0$ can be estimated using a pairwise difference estimator with $\Theta = \mathbb{R}^k$ and

$$m(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) = m_{\texttt{PLL}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) = -\mathbb{1}\{\dot{y}_{i,j} \neq 0\} \left(y_i \ln \Lambda(\dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta}) + y_j \ln \Lambda(-\dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta}) \right).$$

The estimator does not admit a closed form solution, but $u \mapsto -\ln \Lambda(u)$ is convex, rendering the minimization problem convex (provided that a non-negative kernel function is used).

2.3 Partially Linear Tobit Model

The partially linear censored regression model studied here is of the form

$$y_i = \max\{\mathbf{x}_i'\boldsymbol{\theta}_0 + \eta_0(\mathbf{w}_i) + \varepsilon_i, 0\},$$

where $\boldsymbol{\theta}_0$ is the parameter of interest, $\eta_0(\cdot)$ is an unknown nuisance function, $\mathbf{x}_i \perp\!\!\!\perp \varepsilon_i | \mathbf{w}_i$, and the conditional distribution $\varepsilon_i | \mathbf{w}_i$ admits a Lebesgue density. A pairwise difference estimator of $\boldsymbol{\theta}_0$ can be obtained by setting $\Theta = \mathbb{R}^k$ and employing

$$m(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) = m_{\texttt{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) = \tilde{m}_{\texttt{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) - \tilde{m}_{\texttt{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \mathbf{0}),$$

where

$$\tilde{m}_{\texttt{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) = \begin{cases} |y_i| - \left(\dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta} + y_j\right) \operatorname{sgn}(y_i) & \text{if } \dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta} \leq -y_j \\ \left|\dot{y}_{i,j} - \dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta}\right| & \text{if } -y_j < \dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta} < y_i \\ |y_j| + \left(\dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta} - y_i\right) \operatorname{sgn}(y_j) & \text{if } y_i \leq \dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta} \end{cases}.$$

Because $\tilde{m}_{\texttt{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \mathbf{0})$ does not depend on $\boldsymbol{\theta}$, the presence of $\tilde{m}_{\texttt{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \mathbf{0})$ in $m_{\texttt{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta})$ does not affect the minimization problem defining the estimator. Nevertheless, it is theoretically attractive to work with $m_{\texttt{PLT}}$ rather than $\tilde{m}_{\texttt{PLT}}$, as doing so allows for weaker regularity conditions for the existence of the expectation of the objective function.

For future reference, we note that $m_{\texttt{PLT}}$ admits the alternative representation

$$m_{\texttt{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) = \begin{cases} |\dot{y}_{i,j} - \dot{\mathbf{x}}_{i,j}' \boldsymbol{\theta}| - |\dot{y}_{i,j}| & \text{if } y_i > 0, y_j > 0 \\ \max\{y_i - \dot{\mathbf{x}}_{i,j}' \boldsymbol{\theta}, 0\} - y_i & \text{if } y_i > 0, y_j = 0 \\ \max\{y_j + \dot{\mathbf{x}}_{i,j}' \boldsymbol{\theta}, 0\} - y_j & \text{if } y_i = 0, y_j > 0 \\ 0 & \text{if } y_i = 0, y_j = 0 \end{cases}.$$

The function $\boldsymbol{\theta} \mapsto m_{\texttt{PLT}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})$ is convex and so is the minimization problem defining the estimator (provided that a non-negative kernel function is used).

3 Distributional Approximation and Bootstrap Inference

As is standard in the literature, we generalize (1.1) slightly and define our estimator $\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n(h_n)$ to be any approximate minimizer of $\widehat{M}_n(\boldsymbol{\theta}; h_n)$, where

$$\widehat{M}_n(\boldsymbol{\theta}; h) = \binom{n}{2}^{-1} \sum_{i<j} m(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) K_h(\mathbf{w}_i - \mathbf{w}_j).$$

To be specific, we require

$$\widehat{M}_n(\widehat{\boldsymbol{\theta}}_n(h); h) \leq \inf_{\boldsymbol{\theta} \in \Theta} \widehat{M}_n(\boldsymbol{\theta}; h) + o_{\mathbb{P}}\left(n^{-1}\right).$$

The objective function $\widehat{M}_n$ is a sample counterpart of the function M given by

$$M(\boldsymbol{\theta}; h) = \mathbb{E}\Big[\widehat{M}_n(\boldsymbol{\theta}; h)\Big] = \mathbb{E}\big[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) K_h(\mathbf{w}_1 - \mathbf{w}_2)\big].$$

Under regularity conditions, this function approximates, as $h \downarrow 0$, a function M_0, which (does not depend on K and) admits a unique minimizer, namely the parameter of interest $\boldsymbol{\theta}_0$.

For the purposes of analyzing $\widehat{\boldsymbol{\theta}}_n$ it is convenient to define $\boldsymbol{\theta}_n = \boldsymbol{\theta}(h_n)$, where

$$\boldsymbol{\theta}(h) \in \operatorname*{arg\,min}_{\boldsymbol{\theta} \in \Theta} M(\boldsymbol{\theta}; h)$$

is interpretable as a (fixed-h) "pseudo" parameter. With the help of $\boldsymbol{\theta}_n$ we can decompose the estimation error $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0$ into a (non-stochastic) "bias" component $\boldsymbol{\theta}_n - \boldsymbol{\theta}_0$ and a "centered" (stochastic) component $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n$. Each component can be analyzed separately and in both cases the analysis will leverage convexity.

3.1 Regularity Conditions

The purpose of the following assumption is to enable us to analyze the bias component $\boldsymbol{\theta}_n - \boldsymbol{\theta}_0$. Among other things, the assumption guarantees that the bias component vanishes asymptotically.

Assumption 1. (i) The kernel function K is a symmetric, bounded probability density.

(ii) $\Theta \subseteq \mathbb{R}^k$ is convex and $\boldsymbol{\theta} \mapsto m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})$ is convex with probability one.

(iii) The distribution of $\mathbf{w}$ admits a Lebesgue density $f_{\mathbf{w}}$, which is bounded and continuous on its support $\mathcal{W}$.

(iv) On Θ, the function M_0 given by

$$M_0(\boldsymbol{\theta}) = \int_{\mathcal{W}} \mathbb{E}[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) | \mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w}] f_{\mathbf{w}}(\mathbf{w})^2 d\mathbf{w}$$

is uniquely minimized at an interior point $\boldsymbol{\theta}_0$.

(v) For each $\boldsymbol{\theta} \in \Theta$,

$$\mathbb{E}[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})] + \mathbb{E}\left[\sup_{\mathbf{w}_2 \in \mathcal{W}} \mathbb{E}[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) | \mathbf{w}_1, \mathbf{w}_2] f_{\mathbf{w}}(\mathbf{w}_2)\right] < \infty$$

and (with probability one)

$$\lim_{\mathbf{u} \to \mathbf{0}} \mathbb{E}[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) | \mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w} + \mathbf{u}] = \mathbb{E}[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) | \mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}].$$

The purpose of the next assumption is to enable us to analyze the asymptotic properties of $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n$. To accommodate examples (such as the partially linear Tobit model) where $\boldsymbol{\theta} \mapsto m(\mathbf{z}_i, \mathbf{z}_j, \boldsymbol{\theta})$ is not fully differentiable, we assume the existence of (stochastic derivative-like) functions $\mathbf{s}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) \in \mathbb{R}^k$ and $\mathbf{H}(\mathbf{w}_i, \mathbf{w}_j; \boldsymbol{\theta}, \mathbf{t}) \in \mathbb{R}^{k \times k}$ such that, for any (direction) $\mathbf{t} \in \mathbb{R}^k$, the approximation errors

$$e_{\mathbf{s}}(\boldsymbol{\theta}, \mathbf{t}, \tau) = \frac{1}{\tau}\left(m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta} + \mathbf{t}\tau) - m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})\right) - \mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})'\mathbf{t}$$

and

$$e_{\mathbf{H}}(\boldsymbol{\theta}, \mathbf{t}, \tau) = \frac{1}{\tau^2}\left(\mathbb{E}[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta} + \mathbf{t}\tau) - m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) - \mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})'\mathbf{t}\tau | \mathbf{w}_1, \mathbf{w}_2]\right) - \frac{1}{2}\mathbf{t}'\mathbf{H}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t})\mathbf{t}$$

are suitably small for $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$ and $\tau > 0$ near zero. As further discussed below, functions $\mathbf{s}$ and $\mathbf{H}$ satisfying the following assumption exist (and are relatively easy to find) in each of our motivating examples.

Assumption 2. (i) There exists a function $b(\cdot)$ with

$$\mathbb{E}[b(\mathbf{z})^4] + \sup_{\mathbf{w} \in \mathcal{W}} \mathbb{E}[b(\mathbf{z})^4 | \mathbf{w}] f_{\mathbf{w}}(\mathbf{w}) < \infty$$

such that $\|\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})\| \leq b(\mathbf{z}_1)b(\mathbf{z}_2)$ for $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$.

(ii) There exist functions $\boldsymbol{\xi}_0(\cdot)$ and $\boldsymbol{\Xi}_0(\cdot)$ such that (with probability one)

$$\lim_{(\boldsymbol{\theta},\mathbf{u})\to(\boldsymbol{\theta}_0,\mathbf{0})} 2\mathbb{E}[\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})|\mathbf{z}_1 = \mathbf{z}, \mathbf{w}_2 = \mathbf{w} + \mathbf{u}]f_{\mathbf{w}}(\mathbf{w}) = \boldsymbol{\xi}_0(\mathbf{z})$$

and

$$\lim_{(\boldsymbol{\theta},\boldsymbol{\vartheta},\mathbf{u})\to(\boldsymbol{\theta}_0,\boldsymbol{\theta}_0,\mathbf{0})} \mathbb{E}[\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\vartheta})'|\mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w} + \mathbf{u}]f_{\mathbf{w}}(\mathbf{w}) = \boldsymbol{\Xi}_0(\mathbf{w}).$$

(iii) For each $\mathbf{t} \in \mathbb{R}^k$, there is some $\delta > 0$ such that

$$\mathbb{E}\left[\sup_{\tau\in(0,\delta),\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta,\mathbf{w}_2\in\mathcal{W}} |\mathbb{E}\left[e_{\mathbf{s}}(\boldsymbol{\theta}, \mathbf{t}, \tau)|\mathbf{z}_1, \mathbf{w}_2\right]| \, f_{\mathbf{w}}(\mathbf{w}_2)^2\right] < \infty,$$
$$\mathbb{E}\left[\sup_{\tau\in(0,\delta),\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta,\mathbf{w}_2\in\mathcal{W}} \mathbb{E}\left[e_{\mathbf{s}}(\boldsymbol{\theta}, \mathbf{t}, \tau)^2|\mathbf{w}_1, \mathbf{w}_2\right] f_{\mathbf{w}}(\mathbf{w}_2)\right] < \infty,$$
$$\mathbb{E}\left[\sup_{\tau\in(0,\delta),\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta,\mathbf{w}_2\in\mathcal{W}} |e_{\mathbf{H}}(\boldsymbol{\theta}, \mathbf{t}, \tau)| f_{\mathbf{w}}(\mathbf{w}_2)\right] < \infty,$$
$$\mathbb{E}\left[\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|<\delta,\mathbf{w}_2\in\mathcal{W}} \|\mathbf{H}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t})\| \, f_{\mathbf{w}}(\mathbf{w}_2)\right] < \infty.$$

(iv) There exists a function $\mathbf{G}_0(\cdot)$ such that, for each $\mathbf{t} \in \mathbb{R}^k$ (and with probability one),

$$\lim_{\tau\downarrow 0,(\boldsymbol{\theta},\mathbf{u})\to(\boldsymbol{\theta}_0,\mathbf{0})} \mathbb{E}\left[e_{\mathbf{s}}(\boldsymbol{\theta}, \mathbf{t}, \tau)|\mathbf{z}_1 = \mathbf{z}, \mathbf{w}_2 = \mathbf{w} + \mathbf{u}\right] = 0,$$
$$\lim_{\tau\downarrow 0,(\boldsymbol{\theta},\mathbf{u})\to(\boldsymbol{\theta}_0,\mathbf{0})} \mathbb{E}\left[e_{\mathbf{s}}(\boldsymbol{\theta}, \mathbf{t}, \tau)^2|\mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w} + \mathbf{u}\right] = 0,$$
$$\lim_{\tau\downarrow 0,(\boldsymbol{\theta},\mathbf{u})\to(\boldsymbol{\theta}_0,\mathbf{0})} \mathbb{E}\left[e_{\mathbf{H}}(\boldsymbol{\theta}, \mathbf{t}, \tau)|\mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w} + \mathbf{u}\right] = 0,$$
$$\lim_{(\boldsymbol{\theta},\mathbf{u})\to(\boldsymbol{\theta}_0,\mathbf{0})} \|\mathbf{H}(\mathbf{w}, \mathbf{w} + \mathbf{u}; \boldsymbol{\theta}, \mathbf{t})f_{\mathbf{w}}(\mathbf{w}) - \mathbf{G}_0(\mathbf{w})\| = 0.$$

(v) $\boldsymbol{\Gamma}_0 = \mathbb{E}[\mathbf{G}_0(\mathbf{w})]$, $\boldsymbol{\Sigma}_0 = \mathbb{E}[\boldsymbol{\xi}_0(\mathbf{z})\boldsymbol{\xi}_0(\mathbf{z})']$, and $\mathbb{E}[\boldsymbol{\Xi}_0(\mathbf{w})]$ are positive definite.

3.2 Small Bandwidth Asymptotics

Defining

$$\mathbf{V}_n = \boldsymbol{\Gamma}_0^{-1}\left[n^{-1}\boldsymbol{\Sigma}_0 + \binom{n}{2}^{-1} h_n^{-d}\boldsymbol{\Delta}_0(K)\right]\boldsymbol{\Gamma}_0^{-1}, \qquad \boldsymbol{\Delta}_0(K) = \mathbb{E}\left[\boldsymbol{\Xi}_0(\mathbf{w})\right]\int_{\mathbb{R}^d} K^2(\mathbf{u})d\mathbf{u},$$

and letting Φ_k denote the distribution function of a k-dimensional standard Gaussian random vector, we have the following result.

Theorem 1. *Suppose Assumptions 1 and 2 hold. If $n^2 h_n^d \to \infty$ and if $h_n \to 0$, then*

$$\sup_{\mathbf{t}\in\mathbb{R}^k} \left| \mathbb{P}\left[\mathbf{V}_n^{-1/2}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n) \le \mathbf{t} \right] - \Phi_k(\mathbf{t}) \right| \to 0.$$

The convergence rate of $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n$ equals $\sqrt{n \min(1, nh^d)}$, the magnitude of $\mathbf{V}_n^{-1/2}$. Provided that the bias is "small" in the sense that $\mathbf{V}_n^{-1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = o(1)$, Theorem 1 therefore encompasses the following three distinct large-sample regimes:

- *Asymptotic Linearity*: If $nh_n^d \to \infty$, then $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ satisfies (1.2). In particular, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ convergences in law to a mean-zero Gaussian distribution with asymptotic variance

$$\lim_{n\to\infty} n\mathbf{V}_n = \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Sigma}_0\boldsymbol{\Gamma}_0^{-1}.$$

- *Root-n Consistency without Asymptotic Linearity*: If $nh_n^d \to c \in (0, \infty)$, then $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ is not asymptotically linear, but convergences in law to a mean-zero Gaussian distribution with asymptotic variance

$$\lim_{n\to\infty} n\mathbf{V}_n = \boldsymbol{\Gamma}_0^{-1}\left[\boldsymbol{\Sigma}_0 + \frac{2}{c}\boldsymbol{\Delta}_0(K)\right]\boldsymbol{\Gamma}_0^{-1}.$$

- *Slower than Root-n Consistency*: If $nh_n^d \to 0$ (but $n^2 h_n^d \to \infty$), then $\sqrt{n^2 h^d}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ converges weekly to a mean-zero Gaussian distribution with asymptotic variance

$$\lim_{n\to\infty} n^2 h^d \mathbf{V}_n = 2\boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Delta}_0(K)\boldsymbol{\Gamma}_0^{-1}.$$

The small bandwidth component (i.e., the term involving $\boldsymbol{\Delta}_0(K)$) in $\mathbf{V}_n$ captures the additional uncertainty generated from increasing the localization of the observations pairs. Incorporating this component in the approximate variance is key to enabling us to replace the condition $nh_n^d \to \infty$ by the weaker condition $n^2 h_n^d \to \infty$ when obtaining a Gaussian approximation. As demonstrated by Cattaneo et al. (2025a), incorporating the small bandwidth component can furthermore lead to a higher-order corrected distributional approximation even under asymptotic linearity.

3.3 Debiasing

In Theorem 1, we centered the estimator $\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n(h_n)$ at $\boldsymbol{\theta}_n = \boldsymbol{\theta}(h_n)$ to circumvent bias issues. This section focuses on the bias term $\boldsymbol{\theta}_n - \boldsymbol{\theta}_0$ and introduces an automatic debiasing approach under high-level conditions. Assumption 1 implies that $\lim_{h\downarrow 0} \boldsymbol{\theta}(h) = \boldsymbol{\theta}_0$ and therefore $\boldsymbol{\theta}_n - \boldsymbol{\theta}_0 = o(1)$. Under mild additional smoothness conditions, it is possible to develop an expansion of $\boldsymbol{\theta}_n - \boldsymbol{\theta}_0$ in powers of h_n. As a proof of concept, the following result gives conditions under which $\boldsymbol{\theta}_n - \boldsymbol{\theta}_0 = O(h_n^2)$. When stating the result, we employ standard multi-index notation: for $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)' \in \mathbb{Z}_+^d$,

$\mathbf{v} = (v_1, \ldots, v_d)' \in \mathbb{R}^d$, and a sufficiently smooth-in-$\mathbf{v}$ function $f(\mathbf{w}, \mathbf{v})$,

$$\partial_{\mathbf{v}}^{\boldsymbol{\alpha}} f(\mathbf{w}, \mathbf{v}) = \frac{\partial^{|\boldsymbol{\alpha}|}}{\partial v_1^{\alpha_1} \cdots \partial v_d^{\alpha_d}} f(\mathbf{w}, \mathbf{v}), \qquad |\boldsymbol{\alpha}| = \sum_{j=1}^{d} \alpha_j.$$

Proposition 1. *Suppose Assumptions 1 and 2 hold, and that*

(i) $\int_{\mathbb{R}^d} \|\mathbf{u}\|^2 K(\mathbf{u}) d\mathbf{u} < \infty$, *and*

(ii) With probability one, $\boldsymbol{\psi}(\mathbf{w}_1, \mathbf{w}_2) = \mathbb{E}[\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)|\mathbf{w}_1, \mathbf{w}_2] f_{\mathbf{w}}(\mathbf{w}_2)$ *is twice continuously differentiable in* $\mathbf{w}_2$ *with* $\mathbb{E}[\sup_{\mathbf{v} \in \mathcal{W}} \|\partial_{\mathbf{v}}^{\boldsymbol{\alpha}} \boldsymbol{\psi}(\mathbf{w}, \mathbf{v})\|] < \infty$ *for all* $|\boldsymbol{\alpha}| \leq 2$.

Then $\boldsymbol{\theta}(\cdot)$ *admits a* $\mathbf{b}_2 \in \mathbb{R}^k$ *such that*

$$\boldsymbol{\theta}(h) - \boldsymbol{\theta}_0 = \mathbf{b}_2 h^2 + o(h^2) \qquad \textit{as } h \downarrow 0.$$

The proof of Proposition 1 leverages convexity and may therefore be of independent interest. Under appropriate smoothness conditions and with some additional work, it is possible to leverage convexity also to obtain a polynomial expansion of $\boldsymbol{\theta}(h)$ with a remainder of order $o(h^L)$ with $L > 2$. As a proof of concept, Section 4.5 illustrates how to do so when $L = 4$, giving conditions under which $\boldsymbol{\theta}(\cdot)$ admits $\mathbf{b}_2, \mathbf{b}_4 \in \mathbb{R}^k$ such that

$$\boldsymbol{\theta}(h) - \boldsymbol{\theta}_0 = \mathbf{b}_2 h^2 + \mathbf{b}_4 h^4 + o(h^4) \qquad \text{as } h \downarrow 0.$$

Going beyond $L = 4$ seems feasible, but tedious, so going forward we follow Honoré and Powell (2005, Section 3.3) and discuss debiasing under a high-level condition, namely

Assumption 3. For some even $L \geq 0$, $\boldsymbol{\theta}(\cdot)$ admits $\mathbf{b}_{2l} \in \mathbb{R}^k$ (for $l = 1, \ldots, L/2$) such that

$$\boldsymbol{\theta}(h) - \boldsymbol{\theta}_0 = \sum_{l=1}^{L/2} \mathbf{b}_{2l} h^{2l} + o(h^L) \qquad \text{as } h \downarrow 0.$$

Assumptions 1 and 2 imply that Assumption 3 holds with $L = 0$, while Proposition 1 gives conditions under which the assumption holds with $L = 2$. More generally, as indicated by the discussion in Section 4.5, mild restrictions on the kernel can be combined with smoothness conditions under which an expansion of the form given in the assumption is valid. In particular, the fact that the coefficients on odd powers of h are zero is a consequence of the symmetry assumption on K imposed in Assumption 2(i).

To describe the debiasing procedure based on generalized jackknifing, define $c_0 = 1$ and let

$\mathbf{c} = (c_0, \ldots, c_{L/2})'$ be a vector of (distinct) positive constants such that the vector

$$\begin{pmatrix} \lambda_0(\mathbf{c}) \\ \lambda_1(\mathbf{c}) \\ \vdots \\ \lambda_{L/2}(\mathbf{c}) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & c_1^2 & \cdots & c_{L/2}^2 \\ \vdots & & \ddots & \\ 1 & c_1^L & \cdots & c_{L/2}^L \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

is well defined. The debiased estimator is

$$\widetilde{\boldsymbol{\theta}}_n = \widetilde{\boldsymbol{\theta}}_n(\mathbf{c}) = \sum_{l=0}^{L/2} \lambda_l(\mathbf{c}) \widehat{\boldsymbol{\theta}}_n(c_l h_n),$$

the construction of which involves solving $L/2 + 1$ convex optimization problems. As defined, the debiased estimator is a generalization of the original pairwise difference estimator because if $L = 0$, then $\mathbf{c} = 1 = \lambda_0$ and therefore $\widetilde{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n$.

The next theorem generalizes Theorem 1 by establishing the small bandwidth Gaussian approximation for $\widetilde{\boldsymbol{\theta}}_n$. To state the theorem, let

$$\bar{\boldsymbol{\theta}}_n = \bar{\boldsymbol{\theta}}_n(\mathbf{c}) = \sum_{l=0}^{L/2} \lambda_l(\mathbf{c}) \boldsymbol{\theta}(c_l h_n)$$

and

$$\bar{\mathbf{V}}_n = \bar{\mathbf{V}}_n(\mathbf{c}) = \boldsymbol{\Gamma}_0^{-1} \left[n^{-1} \boldsymbol{\Sigma}_0 + \binom{n}{2}^{-1} h_n^{-d} \boldsymbol{\Delta}_0(\bar{K}) \right] \boldsymbol{\Gamma}_0^{-1}, \qquad \bar{K}(\mathbf{u}) = \bar{K}(\mathbf{u}; \mathbf{c}) = \sum_{l=0}^{L/2} \lambda_l(\mathbf{c}) K_{c_l}(\mathbf{u}).$$

As they should, the expression have the feture that if $L = 0$, then $\bar{\boldsymbol{\theta}}_n = \boldsymbol{\theta}_n$ and $\bar{\mathbf{V}}_n = \mathbf{V}_n$. Another noteworthy feature of the expressions is that debiasing via generalized jackknifing affects the variance $\bar{\mathbf{V}}_n$ only through the kernel shape entering its small bandwidth component.

Theorem 2. *Suppose Assumptions 1 and 2 hold. If $n^2 h_n^d \to \infty$ and if $h_n \to 0$, then*

$$\sup_{\mathbf{t} \in \mathbb{R}^k} \left| \mathbb{P}\left[\bar{\mathbf{V}}_n^{-1/2} (\widetilde{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}_n) \leq \mathbf{t} \right] - \Phi_k(\mathbf{t}) \right| \to 0.$$

As a consequence, if also Assumption 3 holds and if $n h_n^{2L} \to 0$, then

$$\sup_{\mathbf{t} \in \mathbb{R}^k} \left| \mathbb{P}\left[\bar{\mathbf{V}}_n^{-1/2} (\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \leq \mathbf{t} \right] - \Phi_k(\mathbf{t}) \right| \to 0$$

The magnitude of $\bar{\mathbf{V}}_n^{-1/2}$ is the same as that of $\mathbf{V}_n^{-1/2}$. As a consequence, with obvious modifications the discussion of $\widehat{\boldsymbol{\theta}}_n$ following Theorem 1 applies to $\widetilde{\boldsymbol{\theta}}_n$, the only noteworthy difference being that (by design), the relevant "small bias" condition is different (and typically milder) in the case of $\widetilde{\boldsymbol{\theta}}_n$.

It is interesting to note that the equivalent kernel $\bar{K}$ is of higher order, even though the debiased estimator $\widetilde{\boldsymbol{\theta}}_n$ only employs estimators constructed using second-order kernels, hereby retaining the desired convexity for implementation. To be specific, if $\int_{\mathbb{R}^d} \|\mathbf{u}\|^{L+2} K(\mathbf{u}) d\mathbf{u} < \infty$, then

$$\int_{\mathbb{R}^d} \bar{K}(\mathbf{u}) d\mathbf{u} = \sum_{l=0}^{L/2} \lambda_l(\mathbf{c}) \int_{\mathbb{R}^d} K_{c_l}(\mathbf{u}) d\mathbf{u} = \sum_{l=0}^{L/2} \lambda_l(\mathbf{c}) = 1$$

and, for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)' \in \mathbb{Z}_+^d$ with $0 < |\boldsymbol{\alpha}| \leq L+1$,

$$\int_{\mathbb{R}^d} \mathbf{u}^{\boldsymbol{\alpha}} \bar{K}(\mathbf{u}) d\mathbf{u} = \sum_{l=0}^{L/2} \lambda_l(\mathbf{c}) \int_{\mathbb{R}^d} \mathbf{u}^{\boldsymbol{\alpha}} K_{c_l}(\mathbf{u}) d\mathbf{u} = \sum_{l=0}^{L/2} \lambda_l(\mathbf{c}) c_l^{|\boldsymbol{\alpha}|} \int_{\mathbb{R}^d} \mathbf{v}^{\boldsymbol{\alpha}} K(\mathbf{v}) d\mathbf{v} = 0$$

where the last equality uses the defining property of $\boldsymbol{\lambda}$ and the symmetry of K, and where $\mathbf{u}^{\boldsymbol{\alpha}}$ denotes $\prod_{j=1}^d u_j^{\alpha_j}$ for $\mathbf{u} = (u_1, \ldots, u_d)' \in \mathbb{R}^d$. In other words, $\bar{K}$ is of order $L+2$.

3.4 Bootstrapping

To develop feasible inference procedures that do not require (explicit) estimation of $\bar{\mathbf{V}}_n$, we consider nonparametric bootstrap-based approximations to the distribution of $\widetilde{\boldsymbol{\theta}}_n$. Since $\widetilde{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n$ when $L = 0$, results for $\widehat{\boldsymbol{\theta}}_n$ can be extracted by setting $L = 0$ in what follows.

Letting $\{\mathbf{z}_{i,n}^* : i = 1, \ldots, n\}$ be a bootstrap i.i.d. sample drawn from the empirical CDF computed from the original observations $\{\mathbf{z}_i : i = 1, \ldots, n\}$, the defining property of $\widehat{\boldsymbol{\theta}}_n^*(h)$, the nonparametric bootstrap analogue of $\widehat{\boldsymbol{\theta}}_n(h)$, is the following:

$$\widehat{M}_n^*(\widehat{\boldsymbol{\theta}}_n^*(h); h) \leq \inf_{\boldsymbol{\theta} \in \Theta} \widehat{M}_n^*(\boldsymbol{\theta}; h) + o_{\mathbb{P}}(n^{-1}),$$

where

$$\widehat{M}_n^*(\boldsymbol{\theta}; h) = \binom{n}{2}^{-1} \sum_{i<j} m(\mathbf{z}_i^*, \mathbf{z}_j^*; \boldsymbol{\theta}) K_h(\mathbf{w}_i^* - \mathbf{w}_j^*).$$

Similarly, the nonparametric bootstrap analogue of $\widetilde{\boldsymbol{\theta}}_n$ is

$$\widetilde{\boldsymbol{\theta}}_n^* = \widetilde{\boldsymbol{\theta}}_n^*(\mathbf{c}) = \sum_{l=0}^{L/2} \lambda_l(\mathbf{c}) \widehat{\boldsymbol{\theta}}_n^*(c_l h_n).$$

The following theorem characterizes the large sample properties of $\widetilde{\boldsymbol{\theta}}_n^* - \widehat{\boldsymbol{\theta}}_n^*$, the bootstrap counterpart of $\widehat{\boldsymbol{\theta}}_n^* - \boldsymbol{\theta}_n$. In perfect analogy with the results in Cattaneo et al. (2014b), we find that the bootstrap distribution estimator is consistent only when $nh_n^d \to \infty$, but otherwise exhibits a variance inflation making the distributional approximation inconsistent. To state the result, let

$\mathbb{P}_n^*[\cdot] = \mathbb{P}[\cdot|\mathbf{z}_1, \ldots, \mathbf{z}_n]$, let $\to_\mathbb{P}$ denote convergence in probability, and define

$$\bar{\mathbf{V}}_n^* = \bar{\mathbf{V}}_n^*(\mathbf{c}) = \mathbf{\Gamma}_0^{-1}\left[n^{-1}\mathbf{\Sigma}_0 + 3\binom{n}{2}^{-1} h_n^{-d}\mathbf{\Delta}_0(\bar{K})\right]\mathbf{\Gamma}_0^{-1}.$$

Theorem 3. *Suppose Assumptions 1 and 2 hold and that, for $\boldsymbol{\theta}$ near $\boldsymbol{\theta}_0$, $m(\mathbf{z}, \mathbf{z}; \boldsymbol{\theta}) = 0$ and $\mathbf{s}(\mathbf{z}, \mathbf{z}; \boldsymbol{\theta}) = \mathbf{0}$ (with probability one). If $n^2 h_n^d \to \infty$ and if $h_n \to 0$, then*

$$\sup_{\mathbf{t}\in\mathbb{R}^k}\left|\mathbb{P}^*\left[\bar{\mathbf{V}}_n^{*-1/2}(\widetilde{\boldsymbol{\theta}}_n^* - \widetilde{\boldsymbol{\theta}}_n) \le \mathbf{t}\right] - \Phi_k(\mathbf{t})\right| \to_\mathbb{P} 0.$$

Because $\bar{\mathbf{V}}_n^{-1}\bar{\mathbf{V}}_n^* \to \mathbf{I}_k$ if and only if $nh_n^d \to \infty$ (where $\mathbf{I}_k$ denotes the k-dimensional identity matrix), under the assumptions of Theorem 3

$$\sup_{\mathbf{t}\in\mathbb{R}^k}\left|\mathbb{P}_n^*\left[\widetilde{\boldsymbol{\theta}}_n^* - \widetilde{\boldsymbol{\theta}}_n \le \mathbf{t}\right] - \mathbb{P}\left[\widetilde{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}_n \le \mathbf{t}\right]\right| \to_\mathbb{P} 0$$

if and only if $nh_n^d \to \infty$. In particular, if $\liminf_{n\to\infty} nh_n^d < \infty$, then the nonparametric bootstrap is inconsistent, albeit conservative in the sense that the (approximate) variance under the bootstrap distribution is larger than the (approximate) variance of the asymptotic distribution: $\bar{\mathbf{V}}_n^* > \bar{\mathbf{V}}_n$ in a positive definite sense.

The variance inflation problem associated with the nonparametric bootstrap under the small bandwidth regime can be easily fixed by appropriately rescaling the bandwidth used for the bootstrap implementation of the pairwise estimator: employing

$$\breve{\boldsymbol{\theta}}_n^* = \breve{\boldsymbol{\theta}}_n^*(\mathbf{c}) = \sum_{l=0}^{L/2}\lambda_l(\mathbf{c})\widehat{\boldsymbol{\theta}}_n^*(3^{1/d}c_l h_n).$$

and centering its distribution at

$$\breve{\boldsymbol{\theta}}_n = \breve{\boldsymbol{\theta}}_n(\mathbf{c}) = \sum_{l=0}^{L/2}\lambda_l(\mathbf{c})\widehat{\boldsymbol{\theta}}_n(3^{1/d}c_l h_n).$$

automatically adjusts the bootstrap variance, leading to a consistent distributional approximation.

Corollary 1. *If the assumptions of Theorem 3 hold, then*

$$\sup_{\mathbf{t}\in\mathbb{R}^k}\left|\mathbb{P}_n^*\left[\breve{\boldsymbol{\theta}}_n^* - \breve{\boldsymbol{\theta}}_n \le \mathbf{t}\right] - \mathbb{P}\left[\widetilde{\boldsymbol{\theta}}_n - \bar{\boldsymbol{\theta}}_n \le \mathbf{t}\right]\right| \to_\mathbb{P} 0.$$

As a consequence, if also Assumption 3 holds and if $nh_n^{2L} \to 0$, then

$$\sup_{\mathbf{t}\in\mathbb{R}^k}\left|\mathbb{P}_n^*\left[\breve{\boldsymbol{\theta}}_n^* - \breve{\boldsymbol{\theta}}_n \le \mathbf{t}\right] - \mathbb{P}\left[\widetilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \le \mathbf{t}\right]\right| \to_\mathbb{P} 0.$$

The statement of Corollary 1 emphasizes the rate-adaptive nature of the consistency property enjoyed by the bootstrap distributional approximation. The result has immediate implications for robust inference. For example, letting $\alpha \in (0,1)$, $\mathbf{a} \in \mathbb{R}^k$ be a fixed vector, and using the "percentile method" (in the terminology of van der Vaart, 1998), the (nominal) level $1-\alpha$ bootstrap confidence interval for $\mathbf{a}'\boldsymbol{\theta}_0$ is

$$\check{\mathsf{CI}}^*_{1-\alpha,n} = \left[\mathbf{a}'\widetilde{\boldsymbol{\theta}}_n - \check{q}^*_{1-\alpha/2,n}\ ,\ \mathbf{a}'\widetilde{\boldsymbol{\theta}}_n - \check{q}^*_{\alpha/2,n}\right], \qquad \check{q}^*_{t,n} = \inf\left\{q \in \mathbb{R} : \mathbb{P}^*_n[\mathbf{a}'\check{\boldsymbol{\theta}}^*_n - \mathbf{a}'\check{\boldsymbol{\theta}}_n \leq q] \geq t\right\}.$$

If Assumptions 1–3 hold and if $n^2h_n^d \to \infty$ and $nh_n^{2L} \to 0$, then

$$\lim_{n\to\infty} \mathbb{P}[\mathbf{a}'\boldsymbol{\theta}_0 \in \check{\mathsf{CI}}^*_{1-\alpha,n}] = 1-\alpha.$$

4 Proofs and Other Technical Results

In the sequel, we use C to denote a positive constant that does not depend on the sample size. In different places, C may refer to different constants. Recall

$$r_n = \left(n^{-1/2} + (n^2h_n^d)^{-1/2}\right)^{-1} = \sqrt{n}\frac{\sqrt{nh_n^d}}{1+\sqrt{nh_n^d}} = O\Big(\min\Big\{\sqrt{n}, \sqrt{n^2h_n^d}\Big\}\Big).$$

Theorem 1 follows from Theorem 2 by setting $\mathbf{c} = 1$, and Corollary 1 follows from Theorems 2 and 3. Thus, we only give proofs for Theorems 2 and 3.

4.1 Proof of Theorem 2

Let $M_{n,l}(\boldsymbol{\theta}) = M(\boldsymbol{\theta}; c_l h_n)$, $\widehat{M}_{n,l}(\boldsymbol{\theta}) = \widehat{M}_n(\boldsymbol{\theta}; c_l h_n)$, $\boldsymbol{\theta}_{n,l} = \boldsymbol{\theta}(c_l h_n)$,

$$\begin{aligned} &\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2) = \mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_{n,l})K_{c_lh_n}(\mathbf{w}_1-\mathbf{w}_2), \qquad \text{and} \\ &e_{n,l}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{t}) = e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_{n,l},\mathbf{t},r_n^{-1})K_{c_lh_n}(\mathbf{w}_1-\mathbf{w}_2). \end{aligned}$$

For $\mathbf{t} \in \mathbb{R}^k$,

$$\begin{aligned} &\widehat{M}_{n,l}(\boldsymbol{\theta}_{n,l}+\mathbf{t}r_n^{-1}) - \widehat{M}_{n,l}(\boldsymbol{\theta}_{n,l}) \\ &= M_{n,l}(\boldsymbol{\theta}_{n,l}+\mathbf{t}r_n^{-1}) - M_{n,l}(\boldsymbol{\theta}_{n,l}) + \binom{n}{2}^{-1}\sum_{i<j}\big\{\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j) - \mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)]\big\}'\mathbf{t}r_n^{-1} \\ &\quad + r_n^{-1}\binom{n}{2}^{-1}\sum_{i<j}\big\{e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t}) - \mathbb{E}[e_{n,l}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{t})]\big\}. \end{aligned}$$

By Lemma 3 below, $r_n^2[M_{n,l}(\boldsymbol{\theta}_{n,l}+\mathbf{t}r_n^{-1}) - M_{n,l}(\boldsymbol{\theta}_{n,l})] = \frac{1}{2}\mathbf{t}'\mathbf{H}_0\mathbf{t} + o(1)$. By Hoeffding decomposition and Lemma 4,

$$\binom{n}{2}^{-1}\sum_{i<j} e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t}) - \mathbb{E}[e_{n,l}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{t})] = o_{\mathbb{P}}\Big(n^{-1/2} + n^{-1}h_n^{-d/2}\Big).$$

Writing $\widehat{\mathbf{U}}_{n,l} = \binom{n}{2}^{-1} \sum_{i<j} \mathbf{s}_{n,l}(\mathbf{z}_i, \mathbf{z}_j) - \mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)]$, we have

$$r_n^2 \big[\widehat{M}_{n,l}(\boldsymbol{\theta}_n + \mathbf{t} r_n^{-1}) - \widehat{M}_{n,l}(\boldsymbol{\theta}_n)\big] = \frac{1}{2}\mathbf{t}'\mathbf{H}_0\mathbf{t} + r_n \widehat{\mathbf{U}}_{n,l}'\mathbf{t} + o_{\mathbb{P}}(1).$$

Since $\mathbf{t} \mapsto \widehat{M}_{n,l}(\boldsymbol{\theta}_{n,l} + \mathbf{t} r_n^{-1})$ is convex almost surely, $\mathbf{H}_0$ is positive definite, and $r_n \widehat{\mathbf{U}}_{n,l} = O_{\mathbb{P}}(1)$ (which we prove below), the corollary following Lemma 2 of Hjort and Pollard (1993) implies that

$$r_n\big(\widehat{\boldsymbol{\theta}}_n(c_l h_n) - \boldsymbol{\theta}_{n,l}\big) - \big(-\mathbf{H}_0^{-1} r_n \widehat{\mathbf{U}}_{n,l}\big) = o_{\mathbb{P}}(1).$$

Since the above holds for each $l = 0, \ldots, L/2$, we have

$$r_n\big(\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}} - \boldsymbol{\theta}_{n,\mathbf{c}}\big) = -\mathbf{H}_0^{-1} \sum_{l=0}^{L/2} \lambda_l r_n \widehat{\mathbf{U}}_{n,l} + o_{\mathbb{P}}(1). \tag{4.1}$$

Under Assumption 3 and $nh_n^{2L} \to 0$,

$$r_n\big(\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}} - \boldsymbol{\theta}_0\big) = r_n\big(\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}} - \boldsymbol{\theta}_{n,\mathbf{c}}\big) + r_n\big(\boldsymbol{\theta}_{n,\mathbf{c}} - \boldsymbol{\theta}_0\big) = r_n\big(\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}} - \boldsymbol{\theta}_{n,\mathbf{c}}\big) + o(1).$$

Thus, it suffices to analyze asymptotic behavior of $\sum_{l=0}^{L/2} \lambda_l r_n \widehat{\mathbf{U}}_{n,l}$. By Hoeffding decomposition,

$$\sum_{l=0}^{L/2} \lambda_l \widehat{\mathbf{U}}_{n,l} = \frac{1}{n}\sum_{i=1}^{n}\sum_{l=0}^{L/2} 2\lambda_l \boldsymbol{\ell}_{n,l}(\mathbf{z}_i) + \binom{n}{2}^{-1} \sum_{i<j}\sum_{l=0}^{L/2} \lambda_l \boldsymbol{\omega}_{n,l}(\mathbf{z}_i, \mathbf{z}_j) =: \widehat{\mathbf{L}}_n + \widehat{\mathbf{W}}_n$$

where

$$\boldsymbol{\ell}_{n,l}(\mathbf{z}_1) = \mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)|\mathbf{z}_1] - \mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)]$$

and

$$\boldsymbol{\omega}_{n,l}(\mathbf{z}_1, \mathbf{z}_2) = \mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2) - \boldsymbol{\ell}_{n,l}(\mathbf{z}_1) - \boldsymbol{\ell}_{n,l}(\mathbf{z}_2) - \mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)].$$

Below, we prove

$$\begin{pmatrix} \sqrt{n}\widehat{\mathbf{L}}_n \\ \sqrt{n^2 h_n^d}\widehat{\mathbf{W}}_n \end{pmatrix} \rightsquigarrow \text{Normal}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_0 & \mathbf{O} \\ \mathbf{O} & 2\boldsymbol{\Delta}_0(K_{\mathbf{c}}) \end{bmatrix}\right) \tag{4.2}$$

where $\mathbf{O}$ is the $k \times k$ zero matrix. By (4.1) and Hoeffding decomposition, we have

$$\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}} - \boldsymbol{\theta}_{n,\mathbf{c}} = -\mathbf{H}_0^{-1}\left[\frac{1}{\sqrt{n}}\sqrt{n}\widehat{\mathbf{L}}_n + \frac{1}{\sqrt{n^2 h_n^d}}\sqrt{n^2 h_n^d}\widehat{\mathbf{W}}_n\right] + o_{\mathbb{P}}(r_n^{-1}),$$

and we can prove the desired result by invoking an almost sure representation theorem. Let $\mathbf{L}$ and $\mathbf{W}$ be mean-zero joint normal random vectors with the covariance matrix in (4.2), and with some abuse of notation,

$$\mathbb{P}\big[\mathbf{V}_{n,\mathbf{c}}^{-1/2}\big(\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}-\boldsymbol{\theta}_{n,\mathbf{c}}\big)\leq\mathbf{t}\big]=\mathbb{P}\left[-\mathbf{V}_{n,\mathbf{c}}^{-1/2}\mathbf{H}_0^{-1}\left(n^{-1/2}\mathbf{L}+(n^2h_n^d)^{-1/2}\mathbf{W}\right)+\mathbf{a}_n\leq\mathbf{t}\right]$$

where $\mathbf{a}_n=o(1)$ almost surely. Since the variance of $\mathbf{H}_0^{-1}(n^{-1/2}\mathbf{L}+(n^2h_n^d)^{-1/2}\mathbf{W})$ equals $\mathbf{V}_{n,\mathbf{c}}$, the desired result holds. □

Proving (4.2) For $\boldsymbol{\mu}_1,\boldsymbol{\mu}_2\in\mathbb{R}^k$, letting $\boldsymbol{\mu}=(\boldsymbol{\mu}_1'\ \boldsymbol{\mu}_2')'$, define

$$g_{in}(\boldsymbol{\mu})=2\left(n^{-1/2}\boldsymbol{\mu}_1'\boldsymbol{\ell}_n(\mathbf{z}_i)+(n-1)^{-1}\sqrt{h_n^d}\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)$$

where $\boldsymbol{\ell}_n(\mathbf{z}_i)=\sum_{l=0}^{L/2}\lambda_l\boldsymbol{\ell}_{n,l}(\mathbf{z}_i)$ and $\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)=\sum_{l=0}^{L/2}\lambda_l\boldsymbol{\omega}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)$. Note $(\sqrt{n}\widehat{\mathbf{L}}_n'\ \sqrt{n^2h_n^d}\widehat{\mathbf{W}}_n')\boldsymbol{\mu}=\sum_{i=1}^n g_{in}(\boldsymbol{\mu})$. Since $\mathbb{E}[g_{in}(\boldsymbol{\mu})|\mathbf{z}_1,\ldots,\mathbf{z}_j]=\mathbf{0}$ for $j\in\{1,\ldots,i-1\}$, $\{g_{in}(\boldsymbol{\mu}),\mathcal{F}_i\}_{i=1}^n$ is a martingale difference sequence where $\mathcal{F}_i$ is the sigma field generated by $\{\mathbf{z}_1,\ldots,\mathbf{z}_i\}$. Using this martingale structure, we apply the following result of Heyde and Brown (1970).

Lemma 1 (Heyde and Brown, 1970)**.** *Let $\{X_n,\mathcal{F}_n\}$ be a martingale with $X_0=0$ a.s., $X_n=\sum_{i=1}^n Y_i$ for $n\geq 1$, and $\mathcal{F}_n$ be the sigma field generated by $X_0,X_1,\ldots,X_n$. Define $\sigma_n^2=\mathbb{E}[Y_n^2|\mathcal{F}_{n-1}]$ and $\varsigma_n^2=\sum_{i=1}^n\mathbb{E}\sigma_i^2$. Suppose for some $\delta\in(0,1]$, $\mathbb{E}|Y_n|^{2+2\delta}<\infty$ for all n. Then, there exists a finite constant K depending only on δ such that*

$$\sup_{t\in\mathbb{R}}|\mathbb{P}(X_n\leq\varsigma_n t)-\Phi(t)|\leq K\left\{\varsigma_n^{-2(1+\delta)}\left(\sum_{i=1}^n\mathbb{E}|Y_i|^{2(1+\delta)}+\mathbb{E}\left|\sum_{i=1}^n\sigma_i^2-\varsigma_n^2\right|^{1+\delta}\right)\right\}^{\frac{1}{3+2\delta}}$$

where $\Phi(\cdot)$ denotes the cdf of a standard normal random variable.

Using the above central limit theorem and the Cramer-Wold device, to prove (4.2), it suffices to show

$$\varsigma_n^2=\sum_{i=1}^n\mathbb{E}[g_{in}(\boldsymbol{\mu})^2]\to\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_0\boldsymbol{\mu}_1+2\boldsymbol{\mu}_2'\boldsymbol{\Delta}_0(K)\boldsymbol{\mu}_2,\tag{4.3}$$

$$\frac{1}{\varsigma_n^4}\sum_{i=1}^n\mathbb{E}|g_{in}(\boldsymbol{\mu})|^4\to 0,\tag{4.4}$$

and

$$\mathbb{E}\left|\frac{1}{\varsigma_n^2}\sum_{i=1}^n\sigma_{in}^2-1\right|^2\to 0,\qquad\sigma_{in}^2=\mathbb{E}[g_{in}(\boldsymbol{\mu})^2|\mathbf{z}_1,\ldots,\mathbf{z}_{i-1}].\tag{4.5}$$

Verifying (4.3) By $\mathbb{E}[\boldsymbol{\ell}_n(\mathbf{z}_i)\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)']=\mathbf{O}$ for $i>j$ and $\mathbb{E}[\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_p)']=\mathbf{O}$ for $j\neq p$,

$$\mathbb{E}[g_{in}(\boldsymbol{\mu})^2]=\frac{4\mathbb{E}[(\boldsymbol{\mu}_1'\boldsymbol{\ell}_n(\mathbf{z}_i))^2]}{n}+\frac{4h_n^d}{(n-1)^2}\sum_{j=1}^{i-1}\mathbb{E}\left[\left(\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\right]$$

and

$$\varsigma_n^2 = 4\boldsymbol{\mu}_1'\mathbb{E}\left[\boldsymbol{\ell}_n(\mathbf{z}_1)\boldsymbol{\ell}_n(\mathbf{z}_1)'\right]\boldsymbol{\mu}_1 + 2\frac{n}{n-1}\boldsymbol{\mu}_2' h_n^d \mathbb{E}\left[\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_2)\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_2)'\right]\boldsymbol{\mu}_2'.$$

For the first term on the right-hand side above,

$$\begin{aligned}\mathbb{E}[\boldsymbol{\ell}_n(\mathbf{z}_1)\boldsymbol{\ell}_n(\mathbf{z}_1)'] &= \mathbb{E}\left[\sum_{l,\tilde{l}=0}^{L/2}\lambda_l\lambda_{\tilde{l}}\mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]\mathbb{E}[\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]'\right] \\ &\quad - \mathbb{E}\left[\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)\right]\mathbb{E}\left[\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)\right]'\end{aligned}$$

where the second term after the equality is zero (Lemma 3). By the dominated convergence theorem, for each $l = 0,\ldots,L/2$, $\mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1] \to_{a.s} \boldsymbol{\xi}(\mathbf{z}_1)$ and, another application of the dominated convergence theorem implies

$$\mathbb{E}\left[\sum_{l,\tilde{l}=0}^{L/2}\lambda_l\lambda_{\tilde{l}}\mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]\mathbb{E}[\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]'\right] \to \mathbb{E}\left[\sum_{l,\tilde{l}=0}^{L/2}\lambda_l\lambda_{\tilde{l}}\boldsymbol{\xi}_i(\mathbf{z})\boldsymbol{\xi}_i(\mathbf{z})'\right] = \frac{1}{4}\boldsymbol{\Sigma}_0$$

because $\sum_{l=0}^{L/2}\lambda_l = 1$. For the other term in the decomposition of ς_n^2,

$$\mathbb{E}[\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_2)\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_2)'] = \mathbb{E}\left[\sum_{l,\tilde{l}=0}^{L/2}\lambda_l\lambda_{\tilde{l}}\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)'\right] + O(1)$$

where $O(1)$ comes from $\mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)\mathbb{E}[\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]'] = \mathbb{E}[\mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]\mathbb{E}[\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]'] = O(1)$ and $\mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)] = \mathbf{0}$. The dominated convergence theorem implies

$$h_n^d\mathbb{E}\big[\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1,\mathbf{z}_2)'\big] \to \int \boldsymbol{\Xi}(\mathbf{w}_1)f_{\mathbf{w}}(\mathbf{w}_1)d\mathbf{w}_1\int K_{c_l}(\mathbf{u})K_{c_{\tilde{l}}}(\mathbf{u})d\mathbf{u}.$$

Since

$$\int K_{\mathbf{c}}^2(\mathbf{u})d\mathbf{u} = \int\left(\sum_{l=0}^{L/2}\lambda_l K_{c_l}(\mathbf{u})\right)^2 d\mathbf{u} = \sum_{l,\tilde{l}=0}^{L/2}\lambda_l\lambda_{\tilde{l}}\int K_{c_l}(\mathbf{u})K_{c_{\tilde{l}}}(\mathbf{u})d\mathbf{u},$$

we have

$$h_n^d\mathbb{E}[\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_2)\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_2)'] \to \boldsymbol{\Delta}_0(K_{\mathbf{c}}).$$

Thus, the desired result holds.

Verifying (4.4) Given ς_n^2 converges to a positive number, it suffices to show $\sum_{i=1}^n \mathbb{E}|g_{in}(\boldsymbol{\mu})|^4 = o(1)$. By $(a+b)^s \le 2^{s-1}(a^s+b^s)$,

$$\sum_{i=1}^n \mathbb{E}|g_{in}(\boldsymbol{\mu})/2|^4 \le 8\|\boldsymbol{\mu}_1\|^4 \frac{\mathbb{E}\|\boldsymbol{\ell}_n(\mathbf{z}_1)\|^4}{n} + \frac{8h_n^{2d}}{(n-1)^4}\sum_{i=1}^n \mathbb{E}\left|\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right|^4.$$

Let $\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2) = \sum_{l=0}^{L/2}\lambda_l \mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)$. Then, $\mathbb{E}\|\boldsymbol{\ell}_n(\mathbf{z}_1)\|^4$ is bounded by a constant multiple of $\mathbb{E}[\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]\|^4]$, which is $o(n)$ by Lemma 5. For the other term,

$$\begin{aligned}\mathbb{E}\left|\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right|^4 &= \sum_{j=1}^{i-1}\mathbb{E}\left[\left(\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^4\right] + \sum_{j=1}^{i-1}\sum_{p=1,p\neq j}^{i-1}\mathbb{E}\left[\left(\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\left(\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_p)\right)^2\right] \\ &\le n4^4\|\boldsymbol{\mu}_2\|^4\mathbb{E}\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\|^4 + n^2 4^4\|\boldsymbol{\mu}_2\|^4\mathbb{E}\big[\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\|^2|\mathbf{z}_1]^2\big]\end{aligned}$$

where the inequality uses Lemma 6. Then,

$$\frac{h_n^{2d}}{4^4\|\boldsymbol{\mu}_2\|^4 n^4}\sum_{i=1}^n \mathbb{E}\left|\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right|^4 \le \frac{h_n^{2d}}{n^2}\mathbb{E}\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\|^4 + \frac{h_n^{2d}}{n}\mathbb{E}\left[\mathbb{E}\left[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\|^2|\mathbf{z}_1\right]^2\right]$$

which is $o(1)$ by Lemma 5.

Verifying (4.5) Adding and subtracting $\frac{4h_n^d}{n-1}\mathbb{E}|\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_2)|^2 = \frac{4h_n^d}{(n-1)^2}\mathbb{E}|\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)|^2$,

$$\begin{aligned}\sum_{i=1}^n \sigma_{in}^2 = \varsigma_n^2 &+ \frac{4h_n^d}{(n-1)^2}\sum_{i=1}^n\left(\mathbb{E}\left[\left(\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\middle|\mathcal{F}_i\right] - \mathbb{E}\left(\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\right) \\ &+ \frac{8\sqrt{h_n^d}}{(n-1)\sqrt{n}}\sum_{i=1}^n\mathbb{E}\left[\sum_{j=1}^{i-1}\boldsymbol{\mu}_1'\boldsymbol{\ell}_n(\mathbf{z}_i)\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\middle|\mathcal{F}_i\right].\end{aligned}$$

Then, to show (4.5), it suffices to verify

$$\frac{h_n^{2d}}{n^4}\mathbb{E}\left|\sum_{i=1}^n\left(\mathbb{E}\left[\left(\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\middle|\mathcal{F}_i\right] - \mathbb{E}\left(\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\right)\right|^2 = o(1) \tag{4.6}$$

and

$$\frac{h_n^d}{n^3}\mathbb{E}\left|\sum_{i=1}^n\mathbb{E}\left[\sum_{j=1}^{i-1}\boldsymbol{\mu}_1'\boldsymbol{\ell}_n(\mathbf{z}_i)\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\middle|\mathcal{F}_i\right]\right|^2 = o(1). \tag{4.7}$$

For (4.6), letting $\bar{\omega}_n = \mathbb{E}[|\boldsymbol{\mu}_2'\boldsymbol{\omega}(\mathbf{z}_1,\mathbf{z}_2)|^2]$,

$$\begin{aligned}
&\mathbb{E}\left|\sum_{i=1}^{n}\left(\mathbb{E}\left[\left(\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\Big|\mathcal{F}_i\right]-\mathbb{E}\left(\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\right)\right|^2\\
&\leq 2\mathbb{E}\left|\sum_{i=1}^{n}\sum_{j=1}^{i-1}\left(\mathbb{E}\left[\left(\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\big|\mathbf{z}_j\right]-\mathbb{E}\left[\left(\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\right]\right)\right|^2\\
&\quad+4\mathbb{E}\left|\sum_{i=1}^{n}\sum_{j=1}^{i-1}\sum_{p=1}^{j-1}\mathbb{E}\left[\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_p)\big|\mathbf{z}_j,\mathbf{z}_p\right]\right|^2\\
&=2\sum_{j=1}^{n-1}\sum_{p=1}^{n-1}(n-j)(n-p)\mathbb{E}\left(\mathbb{E}\left[\left(\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z},\mathbf{z}_j)\right)^2\big|\mathbf{z}_j\right]-\bar{\omega}_n\right)\left(\mathbb{E}\left[\left(\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z},\mathbf{z}_p)\right)^2\big|\mathbf{z}_p\right]-\bar{\omega}_n\right)\\
&\quad+4\sum_{p_1=1}^{n-2}\sum_{j_1=p_1+1}^{n-1}\sum_{p_2=1}^{n-2}\sum_{j_2=p_2+1}^{n-1}(n-j_1)(n-j_2)\\
&\qquad\times\mathbb{E}\Big[\mathbb{E}\left[\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z},\mathbf{z}_{j_1})\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z},\mathbf{z}_{p_1})\big|\mathbf{z}_{j_1},\mathbf{z}_{p_1}\right]\mathbb{E}\left[\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z},\mathbf{z}_{j_2})\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z},\mathbf{z}_{p_2})|\mathbf{z}_{j_2},\mathbf{z}_{p_2}\right]\Big]\\
&\leq 2n^3\mathbb{E}\left|\mathbb{E}\left[\left(\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_2)\right)^2\big|\mathbf{z}_1\right]-\bar{\omega}_n\right|^2+4n^4\mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_2)\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_3)\big|\mathbf{z}_2,\mathbf{z}_3\right]^2\right]
\end{aligned}$$

where the last inequality uses $\mathbb{E}[\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_2)\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_1,\mathbf{z}_3)\big|\mathbf{z}_2]=0$. Then, (4.6) follows from

$$\begin{aligned}
&\frac{h_n^{2d}}{4^4n^4}\mathbb{E}\left|\sum_{i=1}^{n}\left(\mathbb{E}\left[\left(\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\Big|\mathcal{F}_i\right]-\mathbb{E}\left(\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2\right)\right|^2\\
&\leq 2\|\boldsymbol{\mu}_2\|^4n^{-1}h_n^{2d}\mathbb{E}\left[\mathbb{E}\left[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\|^2\big|\mathbf{z}_1\right]^2\right]+\|\boldsymbol{\mu}_2\|^4h_n^{2d}\mathbb{E}\left[\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]^2\right]
\end{aligned}$$

and Lemma 5.

For (4.7),

$$\begin{aligned}
&\mathbb{E}\left|\sum_{i=1}^{n}\mathbb{E}\left[\sum_{j=1}^{i-1}\boldsymbol{\mu}_1'\boldsymbol{\ell}_n(\mathbf{z}_i)\boldsymbol{\mu}_2'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)\Big|\mathcal{F}_i\right]\right|^2\\
&=\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{p=1}^{\min\{i,j\}-1}\mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\mu}_1'\boldsymbol{\ell}_n(\mathbf{z}_i)\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_p)'\boldsymbol{\mu}_2|\mathbf{z}_p\right]\mathbb{E}\left[\boldsymbol{\mu}_1'\boldsymbol{\ell}_n(\mathbf{z}_j)\boldsymbol{\omega}_n(\mathbf{z}_j,\mathbf{z}_p)'\boldsymbol{\mu}_2|\mathbf{z}_p\right]\right]\\
&\leq n^3\left(\mathbb{E}\left[\mathbb{E}\left[\boldsymbol{\mu}_1'\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)|\mathbf{z}_1]\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)'\boldsymbol{\mu}_2|\mathbf{z}_2\right]^2\right]+O(1)\right)
\end{aligned}$$

and Lemma 5 implies $h_n^d\mathbb{E}[\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2]^2]=o(1)$. □

4.2 Proof of Theorem 3

Define $\widehat{M}^*_{n,l}(\boldsymbol{\theta}) = \widehat{M}^*_n(\boldsymbol{\theta}; c_l h_n)$ and

$$\widehat{\mathbf{U}}^*_{n,l} = \binom{n}{2}^{-1} \sum_{i<j} \mathbf{s}_{n,l}(\mathbf{z}^*_i, \mathbf{z}^*_j) - \bar{\mathbf{s}}_{n,l}, \qquad \bar{\mathbf{s}}_{n,l} = \frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbf{s}_{n,l}(\mathbf{z}_i, \mathbf{z}_j).$$

For sufficiently large n,

$$\begin{aligned}
\widehat{M}^*_{n,l}(\boldsymbol{\theta}_{n,l} + \mathbf{t} r_n^{-1}) - \widehat{M}^*_{n,l}(\boldsymbol{\theta}_{n,l}) &= r_n^{-1} \binom{n}{2}^{-1} \sum_{i<j} e_{n,l}(\mathbf{z}^*_i, \mathbf{z}^*_j; \mathbf{t}) + r_n^{-1} \mathbf{t}' \binom{n}{2}^{-1} \sum_{i<j} \mathbf{s}_{n,l}(\mathbf{z}^*_i, \mathbf{z}^*_j) \\
&= r_n^{-1} \left(\binom{n}{2}^{-1} \sum_{i<j} e_{n,l}(\mathbf{z}^*_i, \mathbf{z}^*_j; \mathbf{t}) - \frac{1}{n^2} \sum_{i,j} e_{n,l}(\mathbf{z}_i, \mathbf{z}_j; \mathbf{t}) \right) \\
&\quad + (1 - n^{-1}) \big[\widehat{M}_{n,l}(\boldsymbol{\theta}_{n,l} + \mathbf{t} r_n^{-1}) - \widehat{M}_{n,l}(\boldsymbol{\theta}_{n,l}) \big] + r_n^{-1} \mathbf{t}' \widehat{\mathbf{U}}^*_{n,l}
\end{aligned}$$

where the second equality uses $m(\mathbf{z}, \mathbf{z}; \boldsymbol{\theta}) = 0$ and $\mathbf{s}(\mathbf{z}, \mathbf{z}) = \mathbf{0}$. By identical arguments to the proof of Theorem 1, $r_n^2 [\widehat{M}_{n,l}(\boldsymbol{\theta}_{n,l} + \mathbf{t} r_n^{-1}) - \widehat{M}_{n,l}(\boldsymbol{\theta}_{n,l})] = r_n \mathbf{t}' \widehat{\mathbf{U}}_{n,l} + \frac{1}{2} \mathbf{t}' \mathbf{H}_0 \mathbf{t} + o_{\mathbb{P}}(1)$. Combined with Lemma 7, we have

$$r_n^2 \big[\widehat{M}^*_{n,l}(\boldsymbol{\theta}_{n,l} + \mathbf{t} r_n^{-1}) - \widehat{M}^*_{n,l}(\boldsymbol{\theta}_{n,l}) \big] = r_n \mathbf{t}' \big(\widehat{\mathbf{U}}^*_{n,l} + \widehat{\mathbf{U}}_{n,l} \big) + \frac{1}{2} \mathbf{t}' \mathbf{H}_0 \mathbf{t} + o_{\mathbb{P}}(1).$$

Using $r_n \widehat{\mathbf{U}}^*_n = O_{\mathbb{P}}(1)$ and the corollary following Lemma 2 of Hjort and Pollard (1993), we have

$$r_n \big(\widehat{\boldsymbol{\theta}}^*_n(c_l h_n) - \boldsymbol{\theta}_{n,l} \big) - \Big(-\mathbf{H}_0^{-1} r_n \big(\widehat{\mathbf{U}}^*_{n,l} + \widehat{\mathbf{U}}_{n,l} \big) \Big) = o_{\mathbb{P}}(1)$$

and using $r_n (\widehat{\boldsymbol{\theta}}_n(c_l h_n) - \boldsymbol{\theta}_{n,l}) + \mathbf{H}_0^{-1} r_n \widehat{\mathbf{U}}_{n,l} = o_{\mathbb{P}}(1)$,

$$r_n \big(\widehat{\boldsymbol{\theta}}^*_n(c_l h_n) - \widehat{\boldsymbol{\theta}}_n(c_l h_n) \big) - \Big(-\mathbf{H}_0^{-1} r_n \widehat{\mathbf{U}}^*_{n,l} \Big) = o_{\mathbb{P}}(1).$$

The above display and Hoeffding decomposition imply

$$\widehat{\boldsymbol{\theta}}^*_{n,\mathbf{c}} - \widehat{\boldsymbol{\theta}}_{n,\mathbf{c}} = -\mathbf{H}_0^{-1} \left[\frac{1}{r_n} r_n \widehat{\mathbf{L}}^*_n + \frac{1}{\sqrt{n^2 h_n^d}} \sqrt{n^2 h_n^d} \widehat{\mathbf{W}}^*_n \right] + o_{\mathbb{P}}(r_n^{-1})$$

where

$$\widehat{\mathbf{L}}^*_n = \frac{1}{n} \sum_{i=1}^{n} \sum_{l=0}^{L/2} \lambda_l 2 \widehat{\boldsymbol{\ell}}_{n,l}(\mathbf{z}^*_i), \qquad \widehat{\boldsymbol{\ell}}_{n,l}(\mathbf{z}^*) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{s}_{n,l}(\mathbf{z}^*, \mathbf{z}_j) - \bar{\mathbf{s}}_{n,l}, \qquad \bar{\mathbf{s}}_{n,l} = \frac{1}{n^2} \sum_{i,j} \mathbf{s}_{n,l}(\mathbf{z}_i, \mathbf{z}_j),$$

and

$$\widehat{\mathbf{W}}_n^* = \binom{n}{2}^{-1} \sum_{i<j} \sum_{l=0}^{L/2} \lambda_l \widehat{\boldsymbol{\omega}}_{n,l}(\mathbf{z}_i^*, \mathbf{z}_j^*), \qquad \widehat{\boldsymbol{\omega}}_{n,l}(\mathbf{z}_1^*, \mathbf{z}_2^*) = \mathbf{s}_{n,l}(\mathbf{z}_1^*, \mathbf{z}_2^*) - \widehat{\boldsymbol{\ell}}_{n,l}(\mathbf{z}_1^*) - \widehat{\boldsymbol{\ell}}_{n,l}(\mathbf{z}_2^*) - \bar{\mathbf{s}}_{n,l}.$$

Let $\mathcal{F}_i^*$ be the sigma field generated by $\{\mathbf{z}_1^*, \dots, \mathbf{z}_i^*\}$, and for $\boldsymbol{\mu} = (\boldsymbol{\mu}_1' \ \boldsymbol{\mu}_2')' \in \mathbb{R}^{2k}$,

$$g_{in}^*(\boldsymbol{\mu}) = 2r_n n^{-1} \boldsymbol{\mu}_1' \sum_{l=0}^{L/2} \lambda_l \widehat{\boldsymbol{\ell}}_{n,l}(\mathbf{z}_i^*) + 2h_n^{d/2}(n-1)^{-1} \sum_{j=1}^{i-1} \boldsymbol{\mu}_2' \sum_{l=0}^{L/2} \lambda_l \widehat{\boldsymbol{\omega}}_{n,l}(\mathbf{z}_i^*, \mathbf{z}_j^*).$$

Note $(r_n \widehat{L}_n^{*\prime} \ nh_n^{d/2} \widehat{\mathbf{W}}_n^{*\prime})\boldsymbol{\mu} = \sum_{i=1}^n g_{in}^*(\boldsymbol{\mu})$ and $\{g_{in}^*(\boldsymbol{\mu}), \mathcal{F}_i^*\}_{i=1}^n$ is a martingale difference sequence with respect to the bootstrap measure. As in the non-bootstrap asymptotic distribution, we use the result of Heyde and Brown (1970). Below we show that

$$\widehat{\varsigma}_n^2 = \sum_{i=1}^n \mathbb{E}^*\left[g_{in}^*(\boldsymbol{\mu})^2\right] = \boldsymbol{\mu}_1' \left[\pi_n^2 \boldsymbol{\Sigma}_0 + (1-\pi_n)^2 4\boldsymbol{\Delta}_0(K_{\mathbf{c}})\right] \boldsymbol{\mu}_1 + 2\boldsymbol{\mu}_2' \boldsymbol{\Delta}_0(K_{\mathbf{c}}) \boldsymbol{\mu}_2 + o_{\mathbb{P}}(1) \tag{4.8}$$

where $\pi_n = \frac{\sqrt{nh_n^d}}{1+\sqrt{nh_n^d}}$,

$$\frac{1}{\widehat{\varsigma}_n^4} \sum_{i=1}^n \mathbb{E}^* |g_{in}^*(\boldsymbol{\mu})|^4 \to_{\mathbb{P}} 0, \tag{4.9}$$

and

$$\mathbb{E}^* \left| \frac{1}{\widehat{\varsigma}_n^2} \sum_{i=1}^n \sigma_{in}^{*2} - 1 \right|^2 \to_{\mathbb{P}} 0, \qquad \sigma_{in}^{*2} = \mathbb{E}^*[g_{in}^*(\boldsymbol{\mu})^2 | \mathbf{z}_1^*, \dots, \mathbf{z}_{i-1}^*]. \tag{4.10}$$

First consider the case in which the limit $\lim_n nh_n^d$ exists in the extended real. Below $\frac{1}{\infty}$ and $\frac{1}{0}$ are understood as 0 and ∞, respectively. Writing $\pi_0 = \lim_n \frac{\sqrt{nh_n^d}}{1+\sqrt{nh_n^d}} \in [0,1]$, by (4.8)-(4.10),

$$\begin{pmatrix} r_n \widehat{\mathbf{L}}_n^* \\ \sqrt{n^2 h_n^d} \widehat{\mathbf{W}}_n^* \end{pmatrix} \rightsquigarrow_{\mathbb{P}} \text{Normal}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \pi_0^2 \boldsymbol{\Sigma}_0 + (1-\pi_0)^2 4\boldsymbol{\Delta}_0(K_{\mathbf{c}}) & \mathbf{O} \\ \mathbf{O} & 2\boldsymbol{\Delta}_0(K_{\mathbf{c}}) \end{bmatrix} \right).$$

If $\pi_0 \in (0,1]$, $\lim_n nh_n^d > 0$ and $\lim_n \sqrt{n}/r_n = 1/\pi_0$. Write κ for $\lim_n nh_n^d$, which equals $\left(\frac{\pi_0}{1-\pi_0}\right)^2$. Denoting equality in law by $=_d$, we have

$$\begin{aligned} \sqrt{n}(\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}^* - \widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}) &= -\mathbf{H}_0^{-1} \left[\frac{\sqrt{n}}{r_n} r_n \widehat{\mathbf{L}}_n^* + \frac{1-\pi_n}{\pi_n} \sqrt{n^2 h_n^d} \widehat{\mathbf{W}}_n^* \right] + o_{\mathbb{P}}(1) \\ &\rightsquigarrow_{\mathbb{P}} -\mathbf{H}_0^{-1} \left[\left(\mathbf{L} + \sqrt{\frac{2}{\kappa}} \mathbf{W} \right) + \sqrt{\frac{1}{\kappa}} \mathbf{W}_2 \right] =_d -\mathbf{H}_0^{-1} \left[\mathbf{L} + \sqrt{\frac{3}{\kappa}} \mathbf{W} \right] \end{aligned}$$

where $(\mathbf{L}' \ \mathbf{W}')'$ be a mean-zero joint normal random vector with the covariance matrix in (4.2) and $\mathbf{W}_2$ is a mean-zero normal vector with the covariance matrix $2\boldsymbol{\Delta}_0$, independent of $\mathbf{L}, \mathbf{W}$. That is,

$\sqrt{n}(\widehat{\boldsymbol{\theta}}^*_{n,\mathbf{c}} - \widehat{\boldsymbol{\theta}}_{n,\mathbf{c}})$ is asymptotically normal with the asymptotic covariance

$$\mathbf{H}_0^{-1}\left[\boldsymbol{\Sigma}_0 + \frac{6}{\kappa}\boldsymbol{\Delta}_0(K_{\mathbf{c}})\right]\mathbf{H}_0^{-1}$$

which equals $\lim_{n\to\infty} n\mathbf{V}_{n,\mathbf{c}}(3^{-1/d}h_n)$ because, when $\kappa = \lim_n nh_n^d > 0$,

$$\lim_{n\to\infty} n\mathbf{V}_{n,\mathbf{c}}(h_n) = \mathbf{H}_0^{-1}\left[\boldsymbol{\Sigma}_0 + \lim_{n\to\infty}\frac{2}{nh_n^d}\boldsymbol{\Delta}_0(K_{\mathbf{c}})\right]\mathbf{H}_0^{-1}.$$

If $\pi_0 = 0$, $\lim_n \sqrt{n^2h_n^d}/r_n = 1$ and

$$\begin{aligned}\sqrt{n^2h_n^d}\big(\widehat{\boldsymbol{\theta}}^*_n - \widehat{\boldsymbol{\theta}}_n\big) &= -\mathbf{H}_0^{-1}\left[\frac{\sqrt{n^2h_n^d}}{r_n}r_n\widehat{\mathbf{L}}^*_n + \sqrt{n^2h_n^d}\widehat{\mathbf{W}}^*_n\right] + o_{\mathbb{P}}(1)\\ &\rightsquigarrow_{\mathbb{P}} -\mathbf{H}_0^{-1}\left[\sqrt{2}\mathbf{W} + \mathbf{W}_2\right] =_d -\mathbf{H}_0^{-1}\left[\sqrt{3}\mathbf{W}\right].\end{aligned}$$

Thus, with $\pi_0 = 0$, $\sqrt{n^2h_n^d}(\widehat{\boldsymbol{\theta}}^*_n - \widehat{\boldsymbol{\theta}}_n)$ is asymptotically normal with the covariance matrix

$$\mathbf{H}_0^{-1}[6\boldsymbol{\Delta}_0(K_{\mathbf{c}})]\mathbf{H}_0^{-1} = \lim_{n\to\infty} n^2h_n^d\mathbf{V}_{n,\mathbf{c}}(3^{-1/d}h_n).$$

In both cases, for each $\mathbf{t} = (t_1 \dots t_k)' \in \mathbb{R}^k$,

$$\mathbb{P}^*\left[\mathbf{V}_{n,\mathbf{c}}(3^{-1/3}h_n)^{-1/2}\left(\widehat{\boldsymbol{\theta}}^*_{n,\mathbf{c}}(h_n) - \widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(h_n)\right) \le \mathbf{t}\right] - \Phi(\mathbf{t}) = o_{\mathbb{P}}(1) \tag{4.11}$$

where Φ is the cdf of a k-dimensional standard normal random vector.

For the general case in which nh_n^d may not have a limit on the extended real, we argue along subsequences to prove (4.11).

Verifying (4.8) Write

$$\widehat{\boldsymbol{\ell}}_n(\mathbf{z}) = \sum_{l=0}^{L/2}\lambda_l\widehat{\boldsymbol{\ell}}_{n,l}(\mathbf{z}), \qquad \widehat{\boldsymbol{\omega}}_n(\mathbf{z}_1,\mathbf{z}_2) = \sum_{l=0}^{L/2}\lambda_l\widehat{\boldsymbol{\omega}}_{n,l}(\mathbf{z}_1,\mathbf{z}_2).$$

By $\mathbb{E}^*[\widehat{\boldsymbol{\ell}}_n(\mathbf{z}^*_i)\widehat{\boldsymbol{\omega}}_n(\mathbf{z}^*_i,\mathbf{z}^*_j)'] = \mathbf{O}$ for $i > j$ and $\mathbb{E}^*[\widehat{\boldsymbol{\omega}}_n(\mathbf{z}^*_i,\mathbf{z}^*_j)\widehat{\boldsymbol{\omega}}_n(\mathbf{z}^*_i,\mathbf{z}^*_p)'] = \mathbf{O}$ for distinct i, j, p,

$$\widehat{\varsigma}_n^2 = \frac{4r_n^2}{n}\frac{1}{n}\sum_{i=1}^n\left(\boldsymbol{\mu}_1'\widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i)\right)^2 + \frac{2h_n^dn}{(n-1)}\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left(\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i,\mathbf{z}_j)\right)^2.$$

For the first sum on the right-hand side, by the hypothesis $\mathbf{s}(\mathbf{z}_i,\mathbf{z}_i;\boldsymbol{\theta}) = 0$ for $i = 1,\ldots,n$ and $\boldsymbol{\theta} \in \Theta_0^\delta$,

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i)\widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i)' = \frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\sum_{p=1,p\neq j,i}^{n}\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\right)\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_p)\right)'$$
$$+\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\right)\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\right)' - \left(\sum_{l=0}^{L/2}\lambda_l\bar{\mathbf{s}}_{n,l}\right)\left(\sum_{l=0}^{L/2}\lambda_l\bar{\mathbf{s}}_{n,l}\right)'.$$

Note $\bar{\mathbf{s}}_{n,l} = \frac{n-1}{n}\widehat{\mathbf{U}}_{n,l} + \mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)] = o(1)$. Using Lemma 9, with $\pi_n = \frac{\sqrt{nh_n^d}}{1+\sqrt{nh_n^d}}$,

$$\frac{4r_n^2}{n}\frac{1}{n}\sum_{i=1}^{n}\left(\boldsymbol{\mu}_1'\widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i)\right)^2 = \boldsymbol{\mu}_1'\left[\pi_n^2\boldsymbol{\Sigma}_0 + (1-\pi_n)^2 4\boldsymbol{\Delta}_0\right]\boldsymbol{\mu}_1 + o_{\mathbb{P}}(1).$$

Since $\pi_n \in [0,1]$, this variance term is asymptotically bounded from above and bounded away from zero.

For the term $\sum_{i=1}^n\sum_{j=1}^n(\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i,\mathbf{z}_j))^2$, note that for $l,\tilde{l} \in \{0,\ldots,L/2\}$,

$$\sum_{i=1}^{n}\widehat{\boldsymbol{\ell}}_{n,l}(\mathbf{z}_i)\bar{\mathbf{s}}_{n,\tilde{l}}' = \mathbf{0},$$

and

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\widehat{\boldsymbol{\ell}}_{n,\tilde{l}}(\mathbf{z}_i)' = n\sum_{i=1}^{n}\widehat{\boldsymbol{\ell}}_{n,l}(\mathbf{z}_i)\widehat{\boldsymbol{\ell}}_{n,\tilde{l}}(\mathbf{z}_i)'.$$

Then, using the calculations in the proof of Lemma 9,

$$\begin{aligned}
&\frac{h_n^d}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i,\mathbf{z}_j)\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i,\mathbf{z}_j)' \\
&= \frac{h_n^d}{n^2}\sum_{i=1}^{n}\sum_{j=1,j\neq i}^{n}\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\right)\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\right)' \\
&\quad -\frac{2h_n^d}{n}\sum_{i=1}^{n}\left(\sum_{l=0}^{L/2}\lambda_l\widehat{\boldsymbol{\ell}}_{n,l}(\mathbf{z}_i)\right)\left(\sum_{l=0}^{L/2}\lambda_l\widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i)\right)' - h_n^d\left(\sum_{l=0}^{L/2}\lambda_l\bar{\mathbf{s}}_{n,l}\right)\left(\sum_{l=0}^{L/2}\lambda_l\bar{\mathbf{s}}_{n,l}\right)' \\
&= \boldsymbol{\Delta}_0(K_{\mathbf{c}}) + o_{\mathbb{P}}(1).
\end{aligned}$$

Verifying (4.9) By $(x+y)^4 \le 8(x^4+y^4)$ for $x, y \in \mathbb{R}$,

$$\mathbb{E}^*|g_{in}^*(\boldsymbol{\mu})|^4 \le C\frac{r_n^4}{n^4}\mathbb{E}^*|\boldsymbol{\mu}_1'\widehat{\boldsymbol{\ell}}_n(\mathbf{z}^*)|^4 + C\frac{h_n^{2d}}{n^4}\mathbb{E}^*\left|\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\right|^4.$$

We calculate the stochastic order of each term on the right-hand side. To ease notational burden, in this subsection and the next, we write $\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2) = \sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)$.

$$\begin{aligned}
\mathbb{E}^*|\boldsymbol{\mu}_1'\widehat{\boldsymbol{\ell}}_n(\mathbf{z}^*)|^4 &\le C\mathbb{E}^*\left[\mathbb{E}^*[\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1^*,\mathbf{z}_2^*)|\mathbf{z}_1^*]^4\right] \\
&\le C\frac{1}{n}\sum_{i=1}^n\left|\frac{1}{n}\sum_{j\neq i}\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)\right|^4 \\
&\le C\frac{1}{n^5}\sum_{i=1}^n\left|\sum_{j\neq i}|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|^2 + \sum_{j\neq i}\sum_{p\neq i,j}\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_p)\right|^2 \\
&\le C\frac{1}{n^5}\sum_{i=1}^n\left|\sum_{j\neq i}|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|^2\right|^2 + C\frac{1}{n^5}\sum_{i=1}^n\left|\sum_{j\neq i}\sum_{p\neq i,j}\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_p)\right|^2 \\
&= C\frac{1}{n^5}\sum_{i=1}^n\sum_{j\neq i}|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|^4 + C\frac{1}{n^5}\sum_{i=1}^n\sum_{j\neq i}\sum_{p\neq i,j}|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|^2|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_p)|^2 \\
&\quad + C\frac{1}{n^5}\sum_{i=1}^n\sum_{j\neq i}\sum_{p\neq i,j}\sum_{q\neq i,j,p}|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|^2\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_p)\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_q) \\
&\quad + C\frac{1}{n^5}\sum_{i=1}^n\sum_{j\neq i}\sum_{p\neq i,j}\sum_{q\neq i,j,p}\sum_{r\neq i,j,p,q}\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_p)\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_q)\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_r) \\
&= O_{\mathbb{P}}\Big(n^{-3}\mathbb{E}\big[|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^4\big] + n^{-2}\mathbb{E}\big[\mathbb{E}[|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^2|\mathbf{z}_1]^2\big] \\
&\qquad + n^{-1}\mathbb{E}\big[|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^2\mathbb{E}[\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)|\mathbf{z}_1]^2\big] + \mathbb{E}\big[\mathbb{E}[\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]^4\big]\Big)
\end{aligned}$$

and

$$\begin{aligned}
\frac{r_n^4}{n^4}\sum_{i=1}^n\mathbb{E}^*|\boldsymbol{\mu}_1'\widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i^*)|^4 = O_{\mathbb{P}}\Bigg(&\frac{h_n^{2d}\mathbb{E}\big[|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^4\big]}{n^2(1+\sqrt{nh_n^d})^4} + \frac{h_n^{2d}\mathbb{E}\big[\mathbb{E}[|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^2|\mathbf{z}_1]^2\big]}{n(1+\sqrt{nh_n^d})^4} \\
&+ \frac{h_n^{2d}\mathbb{E}\big[|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^2\mathbb{E}[\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)|\mathbf{z}_1]^2\big]}{(1+\sqrt{nh_n^d})^4} \\
&+ \frac{nh_n^{2d}\mathbb{E}\big[\mathbb{E}[\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]^4\big]}{(1+\sqrt{nh_n^d})^4}\Bigg) = o_{\mathbb{P}}(1)
\end{aligned}$$

where we use Lemmas 5 and 8, and $(nh_n^d)^2/(1+\sqrt{nh_n^d})^4 \le 1$.

For the other term, by $\mathbb{E}^*[\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_p^*)'] = \mathbf{O}$ for $j \neq p$,

$$\begin{aligned}
\mathbb{E}^*\left|\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\right|^4 &= \sum_{j=1}^{i-1}\mathbb{E}^*|\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)|^4 + 12\sum_{j=2}^{i-1}\sum_{p=1}^{j-1}\mathbb{E}^*\left[|\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)|^2|\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_p^*)|^2\right] \\
&\leq Ci\mathbb{E}^*|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1^*,\mathbf{z}_2^*)|^4 + Ci^2\mathbb{E}^*\left[|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1^*,\mathbf{z}_2^*)|^2|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1^*,\mathbf{z}_3^*)|^2\right] \\
&= Ci\frac{1}{n^2}\sum_{j=1}^{n}\sum_{p\neq j}|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_j,\mathbf{z}_p)|^4 + Ci^2\frac{1}{n^3}\sum_{j=1}^{n}\sum_{p\neq j}|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_j,\mathbf{z}_p)|^4 \\
&\quad + Ci^2\frac{1}{n^3}\sum_{j=1}^{n}\sum_{p\neq j}\sum_{q\neq j,p}|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_j,\mathbf{z}_p)|^2|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_j,\mathbf{z}_q)|^2
\end{aligned}$$

and

$$\sum_{i=1}^{n}\frac{h_n^{2d}}{n^4}\mathbb{E}^*\left|\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\right|^4 = h_n^{2d}O_{\mathbb{P}}\Big(n^{-2}\mathbb{E}|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^4 + n^{-1}\mathbb{E}\big[\mathbb{E}[|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^2|\mathbf{z}_1]^2\big]\Big)$$

which is $o_{\mathbb{P}}(1)$ by Lemma 5.

Verifying (4.10)

$$\begin{aligned}
\sum_{i=1}^{n}\sigma_{in}^{*2} - \varsigma_n^2 &= \frac{4r_n^2}{n}\mathbb{E}^*|\boldsymbol{\mu}_1'\widehat{\boldsymbol{\ell}}_n(\mathbf{z}^*)|^2 + \frac{4h_n^d}{(n-1)^2}\sum_{i=1}^{n}\mathbb{E}^*\left[\left|\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\right|^2\middle|\mathcal{F}_{i-1}^*\right] \\
&\quad + \frac{8r_n\sqrt{h_n^d}}{n(n-1)}\sum_{i=1}^{n}\mathbb{E}^*\left[\boldsymbol{\mu}_1'\widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i^*)\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\middle|\mathcal{F}_{i-1}^*\right] - \widehat{\varsigma}_n^2 \\
&= \frac{4h_n^d}{(n-1)^2}\sum_{i=1}^{n}\left(\mathbb{E}^*\left[\left|\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\right|^2\middle|\mathcal{F}_{i-1}^*\right] - \mathbb{E}^*\left[\left|\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\right|^2\right]\right) \\
&\quad + \frac{8r_n\sqrt{h_n^d}}{n(n-1)}\sum_{i=1}^{n}\mathbb{E}^*\left[\boldsymbol{\mu}_1'\widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i^*)\sum_{j=1}^{i-1}\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\middle|\mathcal{F}_{i-1}^*\right] \\
&\equiv I_1 + I_2.
\end{aligned}$$

Then, it suffices to show $\mathbb{E}^*[I_1^2] + \mathbb{E}^*[I_2^2] = o_{\mathbb{P}}(1)$. For the term I_1,

$$\begin{aligned}
I_1 &= \frac{4h_n^d}{(n-1)^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\Big(\mathbb{E}^*\Big[|\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)|^2\Big|\mathcal{F}_{i-1}^*\Big] - \mathbb{E}^*\big[|\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)|^2\big]\Big) \\
&\quad + \frac{4h_n^d}{(n-1)^2}\sum_{i=2}^{n}\sum_{j=1}^{i-1}\sum_{p=1,p\neq j}^{i-1}\mathbb{E}^*\Big[\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_p^*)\Big|\mathcal{F}_{i-1}^*\Big] \equiv J_1 + J_2
\end{aligned}$$

and

$$
\begin{aligned}
\mathbb{E}^* J_1^2 &\le C\frac{h_n^{2d}}{n^4}\sum_{i_1=2}^{n}\sum_{i_2=2}^{n}\sum_{j=1}^{i_1\wedge i_2-1}\mathbb{E}^*\left|\mathbb{E}^*\left[|\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)|^2\middle|\mathcal{F}_{i-1}^*\right]-\mathbb{E}^*\left[|\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)|^2\right]\right|^2\\
&\le C\frac{h_n^{2d}}{n}\mathbb{E}^*\left[\mathbb{E}^*\left[|\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)|^2\middle|\mathcal{F}_{i-1}^*\right]^2\right]\\
&\le C\frac{h_n^{2d}}{n}\mathbb{E}^*\left[\mathbb{E}^*\left[|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1^*,\mathbf{z}_2^*)|^2\middle|\mathbf{z}_1^*\right]^2\right]\\
&= C\frac{h_n^{2d}}{n}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{n}\sum_{j=1}^{n}|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|^2\right)^2\\
&= C\frac{h_n^{2d}}{n}\left(\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|^4+\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{p\neq i,j}|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_p)|^2\right)\\
&= O_{\mathbb{P}}\left(\frac{h_n^{2d}\mathbb{E}[|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^4]}{n^2}+\frac{h_n^{2d}\mathbb{E}\big[\mathbb{E}[|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^2|\mathbf{z}_1]^2\big]}{n}\right)
\end{aligned}
$$

which is $o_{\mathbb{P}}(1)$ by Lemma 5. For the term J_2, for $j\neq p$ and $q\neq r$,

$$
\mathbb{E}^*\left[\mathbb{E}^*\left[\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_j^*)\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_p^*)\middle|\mathcal{F}_{i-1}^*\right]\mathbb{E}^*\left[\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_q^*)\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*,\mathbf{z}_r^*)\middle|\mathcal{F}_{i-1}^*\right]\right]=0
$$

unless $j=q,p=r$ or $j=r,p=q$. Then,

$$
\begin{aligned}
\mathbb{E}^* J_2^2 &\le Ch_n^{2d}\mathbb{E}^*\left[\mathbb{E}^*\left[\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_1^*,\mathbf{z}_2^*)\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_1^*,\mathbf{z}_3^*)\middle|\mathbf{z}_2^*,\mathbf{z}_3^*\right]^2\right]\\
&\le Ch_n^{2d}\mathbb{E}^*\left[\mathbb{E}^*\left[\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1^*,\mathbf{z}_2^*)\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1^*,\mathbf{z}_3^*)\middle|\mathbf{z}_2^*,\mathbf{z}_3^*\right]^2\right]\\
&= Ch_n^{2d}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\frac{1}{n}\sum_{p=1,p\neq i,j}^{n}\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_p,\mathbf{z}_i)\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_p,\mathbf{z}_j)\right)^2\\
&= Ch_n^{2d}\frac{1}{n^4}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{p\neq i,j}\sum_{q\neq i,j,p}\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_p)\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_j,\mathbf{z}_p)\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_q)\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_j,\mathbf{z}_q)\\
&\quad+\frac{Ch_n^{2d}}{n}\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j\neq i}\sum_{p\neq i,j}|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|^2|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_p)|^2+\frac{Ch_n^{2d}}{n^2}\frac{1}{n^2}\sum_{i=1}^{n}\sum_{j\neq i}|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|^4\\
&= O_{\mathbb{P}}\Bigg(h_n^{2d}\mathbb{E}\Big[\mathbb{E}\big[\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)|\mathbf{z}_2,\mathbf{z}_3\big]^2\Big]+\frac{h_n^{2d}\mathbb{E}\big[\mathbb{E}[|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^2|\mathbf{z}_1]^2\big]}{n}\\
&\qquad+\frac{h_n^{2d}\mathbb{E}\big[|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^4\big]}{n^2}\Bigg)=o_{\mathbb{P}}(1)
\end{aligned}
$$

by Lemmas 5 and 8.

Finally, for the term I_2, with $j \neq p$

$$\mathbb{E}^* \left[\mathbb{E}^* \left[\boldsymbol{\mu}_1' \widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i^*) \boldsymbol{\mu}_2' \widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*, \mathbf{z}_j^*) \middle| \mathcal{F}_{i-1}^* \right] \mathbb{E}^* \left[\boldsymbol{\mu}_1' \widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i^*) \boldsymbol{\mu}_2' \widehat{\boldsymbol{\omega}}_n(\mathbf{z}_i^*, \mathbf{z}_p^*) \middle| \mathcal{F}_{i-1}^* \right] \right] = 0$$

and we have

$$\begin{aligned}
\mathbb{E}^*[I_2^2] &\leq C \frac{r_n^2 h_n^d}{n} \mathbb{E}^* \left[\mathbb{E}^* \left[\boldsymbol{\mu}_1' \widehat{\boldsymbol{\ell}}_n(\mathbf{z}_1^*) \boldsymbol{\mu}_2' \widehat{\boldsymbol{\omega}}_n(\mathbf{z}_1^*, \mathbf{z}_2^*) \middle| \mathbf{z}_2^* \right]^2 \right] \\
&\leq C \frac{r_n^2 h_n^d}{n} \Bigg(\mathbb{E}^* \left[\mathbb{E}^* \left[\mathbb{E}^*[\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*) | \mathbf{z}_1^*] \boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*) \middle| \mathbf{z}_2^* \right]^2 \right] + \mathbb{E}^* \left[\mathbb{E}^*[\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*) | \mathbf{z}_1^*]^2 \right]^2 \\
&\qquad + \mathbb{E}^*[\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*)]^4 + \mathbb{E}^*[\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*)]^2 \mathbb{E}^* \left[\mathbb{E}^*[\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*) | \mathbf{z}_1^*]^2 \right] \Bigg).
\end{aligned}$$

Note $\mathbb{E}^*[\mathbb{E}^*[\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*) | \mathbf{z}_1^*]^2] = \frac{1}{n} \sum_{i=1}^n |\boldsymbol{\mu}_1' \widehat{\boldsymbol{\ell}}_n(\mathbf{z}_i)|^2 + |\boldsymbol{\mu}_1' \bar{\mathbf{s}}_n^L|^2 = O_{\mathbb{P}}(1 + (nh_n^d)^{-1})$ where $\bar{\mathbf{s}}_n^L = \sum_{l=0}^{L/2} \lambda_l \bar{\mathbf{s}}_{n,l}$, and $\mathbb{E}^*[\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*)] = \boldsymbol{\mu}_1' \bar{\mathbf{s}}_n^L = o_{\mathbb{P}}(1)$. Then, it remains to calculate the magnitude of $\mathbb{E}^*[\mathbb{E}^*[\mathbb{E}^*[\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*) | \mathbf{z}_1^*] \boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*) | \mathbf{z}_2^*]^2]$. Writing $s_{a,jp} = \boldsymbol{\mu}_a' \mathbf{s}_n^L(\mathbf{z}_j, \mathbf{z}_p)$ for $a = 1, 2$, the bootstrap expectation equals

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n} \sum_{j=1, j \neq i}^n \frac{1}{n} \sum_{p=1, p \neq j}^n s_{1,jp} s_{2,ij} \right)^2 \\
&\leq 2 \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n^2} \sum_{j=1, j \neq i}^n \sum_{p=1, p \neq i,j}^n s_{1,jp} s_{2,ij} \right)^2 + 2 \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{n^2} \sum_{j=1, j \neq i}^n s_{1,ij} s_{2,ij} \right)^2 \\
&= \frac{2}{n^5} \sum_{i,j,p,q,r} s_{2,ij} s_{1,jp} s_{2,iq} s_{1,qr} + \frac{2}{n^5} \sum_{i,j,p,q} \{ s_{2,ij} s_{1,jp} s_{2,iq} s_{1,qj} + s_{2,ij} s_{1,jp} s_{2,iq} s_{1,qp} \} \\
&\quad + \frac{2}{n^5} \sum_{i,j,p,r} \{ s_{2,ij}^2 s_{1,jp} s_{1,jr} + s_{2,ij} s_{1,jp} s_{2,ip} s_{1,pr} \} + \frac{2}{n^5} \sum_{i,j,p} \{ s_{2,ij}^2 s_{1,jp}^2 + s_{2,ij} s_{1,jp}^2 s_{2,ip} \} \\
&\quad + \frac{2}{n^5} \sum_{i,j,p} s_{1,ij} s_{2,ij} s_{1,ip} s_{2,ip} + \frac{2}{n^5} \sum_{i,j} (s_{1,ij} s_{2,ij})^2
\end{aligned}$$

where $\sum_{i,j,p,q,r}$ is understood as summation over $\{1 \leq i, j, p, q, r \leq n : \text{ no two same indices}\}$. Then,

$$\begin{aligned}
&\mathbb{E}^* \left[\mathbb{E}^* \left[\mathbb{E}^*[\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*) | \mathbf{z}_1^*] \boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*) \middle| \mathbf{z}_2^* \right]^2 \right] \\
&= O_{\mathbb{P}} \Big(\mathbb{E} |\boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_2) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_2, \mathbf{z}_3) \boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_4) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_4, \mathbf{z}_5)| \\
&\quad + n^{-1} \mathbb{E} |\boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_2) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_2, \mathbf{z}_3) \boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_4) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_2, \mathbf{z}_4)| \\
&\quad + n^{-1} \mathbb{E} |\boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_2) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_2, \mathbf{z}_3) \boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_4) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_3, \mathbf{z}_4)| \\
&\quad + n^{-1} \mathbb{E} |(\boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_2))^2 \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_2, \mathbf{z}_3) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_2, \mathbf{z}_4)| \\
&\quad + n^{-1} \mathbb{E} |\boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_2) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_2, \mathbf{z}_3) \boldsymbol{\mu}_2' \mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_3) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_3, \mathbf{z}_4)|
\end{aligned}$$

$$
\begin{aligned}
&+ n^{-2}\big\{\mathbb{E}(\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2))^2(\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_2,\mathbf{z}_3))^2 + \mathbb{E}|\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)(\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_2,\mathbf{z}_3))^2\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)|\big\} \\
&+ n^{-2}\mathbb{E}|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)| \\
&+ n^{-3}\mathbb{E}|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|^2\Big)
\end{aligned}
$$

and $\mathbb{E}^*[I_2^2] = o_{\mathbb{P}}(1)$ follows from Lemma 8.

4.3 Technical Lemmas

Lemma 2. *Suppose that Assumption 1 holds. Then, there exists $\bar{h} > 0$ such that the minimization problem $\min_{\boldsymbol{\theta}\in\Theta} M(\boldsymbol{\theta}; h)$ has a solution for each $h \in (0, \bar{h})$. Furthermore, $\lim_{h\downarrow 0}\boldsymbol{\theta}(h) = \boldsymbol{\theta}_0$.*

Proof. By change-of-variables,

$$
M(\boldsymbol{\theta}; h) = \int \mathbb{E}[m(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|\mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w} - \mathbf{u}h] f_{\mathbf{w}}(\mathbf{w}) f_{\mathbf{w}}(\mathbf{w} - \mathbf{u}h) K(\mathbf{u}) d\mathbf{u} d\mathbf{w}.
$$

By the hypothesis, the dominated convergence theorem implies $M(\boldsymbol{\theta}; h) \to M_0(\boldsymbol{\theta})$ as $h \downarrow 0$. By convexity of $\boldsymbol{\theta} \mapsto M(\boldsymbol{\theta}; h)$, the convergence is uniform on any compact set.

Now, fix $\epsilon \in (0, \delta)$. By the hypothesis, $\eta = \inf_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|=\epsilon} M_0(\boldsymbol{\theta}) - M_0(\boldsymbol{\theta}_0) > 0$. By uniform convergence of $M(\boldsymbol{\theta}; h)$, there exists $\bar{h} > 0$ such that $\sup_{\boldsymbol{\theta}\in\Theta_0^\epsilon} |M(\boldsymbol{\theta}; h) - M_0(\boldsymbol{\theta})| < \eta/2$ for $h \in (0, \bar{h})$.

Given $\boldsymbol{\theta}_1$ with $\|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0\| > \epsilon$, let $\lambda \in (0, 1)$ be such that $\|\lambda\boldsymbol{\theta}_1 + (1-\lambda)\boldsymbol{\theta}_0\| = \epsilon$. Write $\boldsymbol{\vartheta} = \lambda\boldsymbol{\theta}_1 + (1-\lambda)\boldsymbol{\theta}_0$. Now,

$$
\begin{aligned}
& M(\boldsymbol{\vartheta}; h) \le \lambda M(\boldsymbol{\theta}_1; h) + (1-\lambda) M(\boldsymbol{\theta}_0; h) \\
\Rightarrow\quad & M_0(\boldsymbol{\vartheta}) - M_0(\boldsymbol{\theta}_0) - \eta/2 \le \lambda\big(M(\boldsymbol{\theta}_1; h) - M(\boldsymbol{\theta}_0; h)\big) \\
\Rightarrow\quad & \eta/2\lambda \le M(\boldsymbol{\theta}_1; h) - M(\boldsymbol{\theta}_0; h)
\end{aligned}
$$

where the last inequality implies that $\inf_{\boldsymbol{\theta}\in\Theta} M(\boldsymbol{\theta}; h) = \min_{\boldsymbol{\theta}:\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\le\epsilon} M(\boldsymbol{\theta}; h)$. Then, using convexity (continuity) of $\boldsymbol{\theta} \mapsto M(\boldsymbol{\theta}; h)$ on the compact set $\{\boldsymbol{\theta} \in \mathbb{R}^k : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le \epsilon\}$, a minimizer exists.

The above argument also shows that for each $\epsilon \in (0, \delta)$, there exists $h_\epsilon > 0$ such that $\|\boldsymbol{\theta}(h) - \boldsymbol{\theta}_0\| \le \epsilon$ for $h \in (0, h_\epsilon)$. □

Lemma 3. *Under Assumption 1(i)(iii) and Assumption 2(iii)(iv), for $\eta_n = o(1)$, $\tau_n = o(1)$, $\boldsymbol{\vartheta}_n = \boldsymbol{\theta}_0 + o(1)$,*

$$
\tau_n^{-2}\big\{M(\boldsymbol{\vartheta}_n + \mathbf{t}\tau_n; \eta_n) - M(\boldsymbol{\vartheta}_n; \eta_n) - \mathbb{E}[\mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\vartheta}_n) K_{\eta_n}(\mathbf{w}_1 - \mathbf{w}_2)]'\mathbf{t}\tau_n\big\} - \frac{1}{2}\mathbf{t}'\mathbf{H}_0\mathbf{t} = o(1)
$$

for each $\mathbf{t} \in \mathbb{R}^k$. In addition, $\mathbb{E}[\mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}(h)) K_h(\mathbf{w}_1 - \mathbf{w}_2)] = \mathbf{0}$ for $\boldsymbol{\theta}(h) \in \Theta_0^\delta$.

Proof. For $\mathbf{t} \in \mathbb{R}^k$,

$$\frac{\left|M(\boldsymbol{\theta}+\mathbf{t}\tau;h)-M(\boldsymbol{\theta};h)-\mathbb{E}[\mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})K_h(\mathbf{w}_1-\mathbf{w}_2)]'\mathbf{t}\tau-\frac{\tau^2}{2}\mathbf{t}'\mathbb{E}[\mathbf{H}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t})K_h(\mathbf{w}_1-\mathbf{w}_2)]\mathbf{t}\right|}{\tau^2}$$
$$\leq \left|\mathbb{E}[e_2(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t},\tau)K_h(\mathbf{w}_1-\mathbf{w}_2)]\right|$$
$$= \left|\int e_2(\mathbf{w},\mathbf{w}-\mathbf{u}h;\boldsymbol{\theta},\mathbf{t},\tau)f_{\mathbf{w}}(\mathbf{w})f_{\mathbf{w}}(\mathbf{w}-\mathbf{u}h)K(\mathbf{u})d\mathbf{w}d\mathbf{u}\right|$$

where the integral in the last line converges to 0 as $(\tau,\boldsymbol{\theta},h)\to(0,\boldsymbol{\theta}_0,0)$ by the dominated convergence theorem under Assumption 1(i)(iii) and Assumption 2(iii)(iv). Now,

$$\mathbb{E}[\mathbf{H}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t})K_h(\mathbf{w}_1-\mathbf{w}_2)] = \int \mathbf{H}(\mathbf{w},\mathbf{w}-\mathbf{u}h;\boldsymbol{\theta},\mathbf{t})f_{\mathbf{w}}(\mathbf{w})f_{\mathbf{w}}(\mathbf{w}-\mathbf{u}h)K(\mathbf{u})d\mathbf{w}d\mathbf{u}$$

and by the dominated convergence theorem,

$$\left\|\mathbb{E}[\mathbf{H}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\vartheta}_n,\mathbf{t})K_h(\mathbf{w}_1-\mathbf{w}_2)]-\int \mathbf{H}(\mathbf{w},\mathbf{w};\boldsymbol{\vartheta}_n,\mathbf{t})f_{\mathbf{w}}(\mathbf{w})^2d\mathbf{w}\int K(\mathbf{u})d\mathbf{u}\right\| = o(1).$$

Again by the dominated convergence theorem,

$$\int \mathbf{H}(\mathbf{w},\mathbf{w};\boldsymbol{\vartheta}_n,\mathbf{t})f_{\mathbf{w}}(\mathbf{w})^2d\mathbf{w} = \int \mathbf{H}(\mathbf{w},\mathbf{w};\boldsymbol{\theta}_0)f_{\mathbf{w}}(\mathbf{w})^2d\mathbf{w}+o(1).$$

Combining above arguments, we obtain the first conclusion.

For the second conclusion,

$$\left|\frac{M(\boldsymbol{\theta}+\mathbf{t};h)-M(\boldsymbol{\theta};h)-\mathbb{E}[\mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})K_h(\mathbf{w}_1-\mathbf{w}_2)]'\mathbf{t}\tau}{\tau}\right| \leq \left|\mathbb{E}[e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta},\mathbf{t},\tau)K_h(\mathbf{w}_1-\mathbf{w}_2)]\right|$$

and for $\boldsymbol{\theta}\in\Theta_0^\delta$, as $\tau\to 0$, the right-hand side term goes to zero by the dominated convergence theorem, which implies that $\boldsymbol{\theta}\mapsto M(\boldsymbol{\theta};h)$ is (directionally) differentiable on Θ_0^δ. Then, the desired result follows because $\boldsymbol{\theta}(h)$ is a local minimizer of $M(\boldsymbol{\theta};h)$. □

Lemma 4. *Suppose Assumption 1(i)(iii) and Assumption 2(iii)(iv) hold. For $\tau_n\to 0$, $\boldsymbol{\vartheta}_n = \boldsymbol{\theta}_0+o(1)$, and $\eta_n = o(1)$,*

$$\mathbb{E}[\mathbb{E}[e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\vartheta}_n,\mathbf{t},\tau_n)K_{\eta_n}(\mathbf{w}_1-\mathbf{w}_2)|\mathbf{z}_1]^2] = o(1), \quad \eta_n^d\mathbb{E}[|e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\vartheta}_n,\mathbf{t},\tau_n)K_{\eta_n}(\mathbf{w}_1-\mathbf{w}_2)|^2] = o(1).$$

Proof. By change-of-variables,

$$\mathbb{E}[e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\vartheta}_n,\mathbf{t},\tau_n)K_{\eta_n}(\mathbf{w}_1-\mathbf{w}_2)|\mathbf{z}_1]$$
$$= \int \mathbb{E}[e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\vartheta}_n,\mathbf{t},\tau_n)|\mathbf{z}_1,\mathbf{w}_2=\mathbf{w}_1-\mathbf{u}\eta_n]f_{\mathbf{w}}(\mathbf{w}_1-\mathbf{u}\eta_n)K(\mathbf{u})d\mathbf{u}$$

and under the hypothesis, the dominated convergence theorem implies the first result. For the other result, by change-of-variables,

$$\begin{aligned}
&\eta_n^d \mathbb{E}[|e_1(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\vartheta}_n, \mathbf{t}, \tau_n) K_{\eta_n}(\mathbf{w}_1 - \mathbf{w}_2)|^2] \\
&= \int \mathbb{E}[e_1(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\vartheta}_n, \mathbf{t}, \tau_n)^2 | \mathbf{w}_1, \mathbf{w}_2 = \mathbf{w}_1 - \mathbf{u}\eta_n] f_{\mathbf{w}}(\mathbf{w}_1) f_{\mathbf{w}}(\mathbf{w}_1 - \mathbf{u}\eta_n) K^2(\mathbf{u}) d\mathbf{u} d\mathbf{w}_1
\end{aligned}$$

and by the hypothesis, we can apply the dominated convergence theorem to conclude that the desired result holds. □

Lemma 5. *Let* $\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2) = \mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}(c_l h_n)) K_{c_l h_n}(\mathbf{w}_1 - \mathbf{w}_2)$. *Suppose* $h_n \to 0$, $n^2 h_n^d \to \infty$, *and* $\lim_{h \downarrow 0} \boldsymbol{\theta}(h) = \boldsymbol{\theta}_0$. *Assumptions 1(i)(iii) and 2(i) imply that for* $l, \tilde{l} \in \{0, \ldots, L/2\}$,

$$\begin{aligned}
&\mathbb{E}\|\mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)|\mathbf{z}_1]\|^4 = o(n), \\
&h_n^{2d} \mathbb{E}\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\|^4 = o(n^2), \\
&h_n^{2d} \mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\|^2 | \mathbf{z}_1]^2] = o(n), \\
&h_n^{2d} \mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2) \mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1, \mathbf{z}_3)'\| | \mathbf{z}_2, \mathbf{z}_3]^2] = o(1), \\
&h_n^{d} \mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2) \mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1, \mathbf{z}_3)'\| | \mathbf{z}_2]^2] = o(1).
\end{aligned}$$

Proof. For h_n small enough, $\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\| \le b(\mathbf{z}_1) b(\mathbf{z}_2) K_{c_l h_n}(\mathbf{w}_1 - \mathbf{w}_2)$ by Assumption 2(i).

Verification of $\mathbb{E}\|\mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)|\mathbf{z}_1]\|^4 = o(n)$. By $\mathbb{E}[b(\mathbf{z})|\mathbf{w}] f(\mathbf{w}) \le C$,

$$\|\mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)|\mathbf{z}_1]\| \le \int b(\mathbf{z}_1) \mathbb{E}[b(\mathbf{z})|\mathbf{w}] f(\mathbf{w}) K_{c_l h_n}(\mathbf{w}_1 - \mathbf{w}) d\mathbf{w} \le C b(\mathbf{z}_1) \int K(\mathbf{u}) d\mathbf{u},$$

and $\mathbb{E}\|\mathbb{E}[\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)|\mathbf{z}_1]\|^4 \le C$ follows from $\mathbb{E} b(\mathbf{z})^4 < \infty$.

Verification of $h_n^2 \mathbb{E}\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\|^4 = o(n^2)$.

$$\begin{aligned}
&\mathbb{E}\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\|^4 \le \mathbb{E} b(\mathbf{z}_1)^4 b(\mathbf{z}_2)^4 K_{c_l h_n}(\mathbf{w}_1 - \mathbf{w}_2)^4 \\
&\le |c_l|^{-3} h_n^{-3d} \int \mathbb{E}[b(\mathbf{z}_2)^4 | \mathbf{w}_2 = \mathbf{w}_1 + \mathbf{u}h] f(\mathbf{w}_1 + \mathbf{u}h_n) \mathbb{E}[b(\mathbf{z}_1)^4 | \mathbf{w}_1] f(\mathbf{w}_1) K(\mathbf{u})^4 d\mathbf{u} d\mathbf{w}_1 \\
&\le C h_n^{-3d} \mathbb{E}[b(\mathbf{z}_1)^4] \int K^4(\mathbf{u}) d\mathbf{u}.
\end{aligned}$$

Then, $h_n^2 \mathbb{E}\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\|^4 \le C h_n^{-d} = o(n^2)$ as $n^2 h_n^d \to \infty$.

Verification of $h_n^{2d} \mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\|^2 | \mathbf{z}_1]^2] = o(n)$.

$$\begin{aligned}
\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\|^2 | \mathbf{z}_1] &\le 3(c_l h_n)^{-2d} \int b(\mathbf{z}_1)^2 \mathbb{E}[b(\mathbf{z}_2)^2 | \mathbf{w}_2] f(\mathbf{w}_2) K\left(\frac{\mathbf{w}_1 - \mathbf{w}_2}{c_l h_n}\right)^2 d\mathbf{w}_2 \\
&\le C b(\mathbf{z}_1)^2 h_n^{-d} \int K(\mathbf{u})^2 d\mathbf{u}
\end{aligned}$$

so, we have $h_n^{2d}\mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\|^2|\mathbf{z}_1]^2] \leq C$.

Verification of $h_n^{2d}\mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1, \mathbf{z}_3)'\||\mathbf{z}_2, \mathbf{z}_3]^2] = o(1)$.

$$\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1, \mathbf{z}_3)'\| \leq b(\mathbf{z}_1)^2 b(\mathbf{z}_2)b(\mathbf{z}_3)K_{c_l h_n}(\mathbf{w}_1 - \mathbf{w}_2)K_{c_{\tilde{l}h_n}}(\mathbf{w}_1 - \mathbf{w}_3)$$

and

$$\begin{aligned}
&\mathbb{E}\big[b(\mathbf{z}_1)^2 b(\mathbf{z}_2)b(\mathbf{z}_3)K_{c_l h_n}(\mathbf{w}_1 - \mathbf{w}_2)K_{c_{\tilde{l}h_n}}(\mathbf{w}_1 - \mathbf{w}_3)\big|\mathbf{z}_2, \mathbf{z}_3\big] \\
&\leq b(\mathbf{z}_2)b(\mathbf{z}_3)h_n^{-d}\int \mathbb{E}[b(\mathbf{z}_1)^2|\mathbf{w}_1 = \mathbf{w}_2 + \mathbf{u}h]f(\mathbf{w}_2 + \mathbf{u}h)K_{c_l}(\mathbf{u})K_{c_{\tilde{l}}}\left(\frac{\mathbf{w}_3 - \mathbf{w}_2}{h_n} - \mathbf{u}\right)d\mathbf{u} \\
&\leq h_n^{-d}Cb(\mathbf{z}_2)b(\mathbf{z}_3)\bar{K}\left(\frac{\mathbf{w}_3 - \mathbf{w}_2}{h_n}\right)
\end{aligned}$$

where $\bar{K}(\mathbf{w}) = \int K_{c_l}(\mathbf{u})K_{c_{\tilde{l}}}(\mathbf{w} - \mathbf{u})d\mathbf{u}$. Now,

$$\mathbb{E}|b(\mathbf{z}_2)b(\mathbf{z}_3)\bar{K}_{h_n}(\mathbf{w}_3 - \mathbf{w}_2)|^2 \leq Ch_n^{-d}\int \bar{K}(\mathbf{u})^2 d\mathbf{u}\mathbb{E}[b(\mathbf{z}_2)]$$

and thus, $h_n^{2d}\mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1, \mathbf{z}_3)'\||\mathbf{z}_2, \mathbf{z}_3]^2] \leq Ch_n^d = o(1)$.

Verification of $h_n^{d}\mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1, \mathbf{z}_3)'\||\mathbf{z}_2]^2] = o(1)$.

$$\begin{aligned}
&\mathbb{E}\big[b(\mathbf{z}_1)^2 b(\mathbf{z}_2)b(\mathbf{z}_3)K_{c_l h_n}(\mathbf{w}_1 - \mathbf{w}_2)K_{c_{\tilde{l}}h_n}(\mathbf{w}_1 - \mathbf{w}_3)\big|\mathbf{z}_2\big] \\
&\leq b(\mathbf{z}_2)\int \mathbb{E}[b(\mathbf{z}_1)^2|\mathbf{w}_1 = \mathbf{w}_2 + \mathbf{u}h]f(\mathbf{w}_2 + \mathbf{u}h)\mathbb{E}[b(\mathbf{z}_3)|\mathbf{w}_3 = \mathbf{w}_2 + \mathbf{v}h]f(\mathbf{w}_3 + \mathbf{v}h) \\
&\qquad \times K_{c_l}(\mathbf{u})K_{c_{\tilde{l}}}(\mathbf{v} - \mathbf{u})\,d\mathbf{u}d\mathbf{v} \\
&\leq Cb(\mathbf{z}_2)\int \bar{K}(\mathbf{v})\,d\mathbf{v}
\end{aligned}$$

and $\mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1, \mathbf{z}_2)\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1, \mathbf{z}_3)'\||\mathbf{z}_2]^2] \leq C$, which implies the desired result. □

Lemma 6. *For $r \in \mathbb{N}$, $\mathbb{E}[|\boldsymbol{\mu}'\boldsymbol{\omega}_n(\mathbf{z}_1, \mathbf{z}_2)|^r|\mathbf{z}_1] \leq 4^r\mathbb{E}[|\boldsymbol{\mu}'\mathbf{s}_n(\mathbf{z}_1, \mathbf{z}_2)|^r|\mathbf{z}_1]$ almost surely.*

Proof. For $i < j < p < q$,

$$\boldsymbol{\omega}_n(\mathbf{z}_i, \mathbf{z}_j) = \mathbb{E}[\mathbf{s}_n(\mathbf{z}_i, \mathbf{z}_j) - \mathbf{s}_n(\mathbf{z}_i, \mathbf{z}_p) - \mathbf{s}_n(\mathbf{z}_j, \mathbf{z}_p) + \mathbf{s}_n(\mathbf{z}_p, \mathbf{z}_q)|\mathbf{z}_i, \mathbf{z}_j].$$

Then, by Jenssen's inequality,

$$\mathbb{E}[|\boldsymbol{\mu}'\boldsymbol{\omega}_n(\mathbf{z}_i, \mathbf{z}_j)|^r|\mathbf{z}_i] \leq \mathbb{E}\left[\left(|\boldsymbol{\mu}'\mathbf{s}_n(\mathbf{z}_i, \mathbf{z}_j)| + |\boldsymbol{\mu}'\mathbf{s}_n(\mathbf{z}_i, \mathbf{z}_p)| + |\boldsymbol{\mu}'\mathbf{s}_n(\mathbf{z}_j, \mathbf{z}_p)| + |\boldsymbol{\mu}'\mathbf{s}_n(\mathbf{z}_p, \mathbf{z}_q)|\right)^r|\mathbf{z}_i\right]$$

and by $(x_1+x_2+x_3+x_4)^r \le 4^{r-1}\sum_{i=1}^4 |x_i|^r$,

$$\mathbb{E}[|\boldsymbol{\mu}'\boldsymbol{\omega}_n(\mathbf{z}_i,\mathbf{z}_j)|^r|\mathbf{z}_i] \le 4^r\mathbb{E}[|\boldsymbol{\mu}'\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_2)|^r|\mathbf{z}_i].$$

□

Lemma 7. *Suppose that Assumption 1(i)(iii) and Assumption 2(iii)(iv) hold. Also, assume that $h_n \to 0$, $n^2h_n^d \to \infty$, $\boldsymbol{\theta}_{n,l} \to \boldsymbol{\theta}_0$, and that for any $\mathbf{z}$ in the support of $\mathbf{z}$, $\mathbf{s}(\mathbf{z},\mathbf{z}) = \mathbf{0}$ and $m(\mathbf{z},\mathbf{z};\boldsymbol{\theta}) = 0$ for $\boldsymbol{\theta} \in \Theta_0^\delta$. Then, for $\mathbf{t} \in \mathbb{R}^k$ and $l = 0,\dots,L/2$,*

$$\binom{n}{2}^{-1}\sum_{i<j} e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_j^*;\mathbf{t}) - \frac{1}{n^2}\sum_{i,j} e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t}) = o_{\mathbb{P}}\left(r_n^{-1}\right).$$

Proof. By Hoeffding decomposition,

$$\begin{aligned}
&\binom{n}{2}^{-1}\sum_{i<j} e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_j^*;\mathbf{t}) - \frac{1}{n^2}\sum_{i,j} e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t})\\
&= \frac{1}{n}\sum_{i=1}^n 2\left(\frac{1}{n}\sum_{j=1}^n e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_j;\mathbf{t}) - \frac{1}{n^2}\sum_{i,j} e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t})\right)\\
&+ \binom{n}{2}^{-1}\sum_{i<j}\left\{e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_j^*;\mathbf{t}) - \frac{1}{n}\sum_{p=1}^n e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_p;\mathbf{t}) - \frac{1}{n}\sum_{p=1}^n e_{n,l}(\mathbf{z}_j^*,\mathbf{z}_p;\mathbf{t}) + \frac{1}{n^2}\sum_{p,q} e_{n,l}(\mathbf{z}_p,\mathbf{z}_q;\mathbf{t})\right\}.
\end{aligned}$$

The variance of $\frac{1}{n}\sum_{j=1}^n e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_j;\mathbf{t}) - \frac{1}{n^2}\sum_{i,j} e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t})$ with respect to the bootstrap distribution is

$$\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{p=1}^n e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t})e_{n,l}(\mathbf{z}_i,\mathbf{z}_p;\mathbf{t}) - \left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t})\right)^2 = o_{\mathbb{P}}\Big(1 + (nh_n^d)^{-1/2}\Big)$$

where the last equality follows from $\mathbb{E}[\mathbb{E}[e_{n,l}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{t})|\mathbf{z}_1]^2] = o(1)$, $h_n^d\mathbb{E}[e_{n,l}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{t})^2] = o(1)$ (both follow from Lemma 4), and $\frac{1}{n(n-1)}\sum_{i=1}^n\sum_{j=1}^n e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t}) = \mathbb{E}[e_{n,l}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{t})] + o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$ where the last equality holds because $\mathbb{E}[e_{n,l}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{t})] = r_n[M_{n,l}(\boldsymbol{\theta}_{n,l}+\mathbf{t}r_n^{-1} - M_{n,l}(\boldsymbol{\theta}_{n,l})] - \mathbb{E}[s_{n,l}(\mathbf{z}_1,\mathbf{z}_2)]'\mathbf{t} = o(1)$. Thus, by Markov inequality,

$$\frac{1}{n}\sum_{i=1}^n 2\left(\frac{1}{n}\sum_{j=1}^n e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_j;\mathbf{t}) - \frac{1}{n^2}\sum_{i,j} e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t})\right) = o_{\mathbb{P}}\Big(n^{-1/2} + (n^2h_n^d)^{-1/2}\Big).$$

The variance of $e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_j^*;\mathbf{t}) - \frac{1}{n}\sum_{p=1}^n e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_p;\mathbf{t}) - \frac{1}{n}\sum_{p=1}^n e_{n,l}(\mathbf{z}_j^*,\mathbf{z}_p;\mathbf{t}) + \frac{1}{n^2}\sum_{p,q} e_{n,l}(\mathbf{z}_p,\mathbf{z}_q;\mathbf{t})$ with respect to the bootstrap distribution is

$$\frac{1}{n^2}\sum_{i,j} e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t})^2 - 2\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{p=1}^n e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t})e_{n,l}(\mathbf{z}_i,\mathbf{z}_p;\mathbf{t}) + \left(\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n e_{n,l}(\mathbf{z}_i,\mathbf{z}_j;\mathbf{t})\right)^2$$

which is $o_{\mathbb{P}}(h_n^{-d})$ by $h_n^d\mathbb{E}[e_{n,l}(\mathbf{z}_1,\mathbf{z}_2;\mathbf{t})^2]=o(1)$. By Markov inequality,

$$\binom{n}{2}^{-1}\sum_{i<j}\left\{e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_j^*;\mathbf{t})-\frac{1}{n}\sum_{p=1}^n e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_p;\mathbf{t})-\frac{1}{n}\sum_{p=1}^n e_{n,l}(\mathbf{z}_j^*,\mathbf{z}_p;\mathbf{t})+\frac{1}{n^2}\sum_{p,l}e_{n,l}(\mathbf{z}_p,\mathbf{z}_l;\mathbf{t})\right\}$$
$$=o_{\mathbb{P}}\Big((n^2h_n^d)^{-1/2}\Big).$$

The desired result follows from combining the two stochastic orders. $\square$

Lemma 8. *Suppose $h_n\to 0$, $n^2h_n^d\to\infty$, and $\lim_{h\downarrow 0}\boldsymbol{\theta}(h)=\boldsymbol{\theta}_0$. Assumptions 1(i)(iii) and 2(i) imply*

$$\begin{aligned}
&\mathbb{E}[\|\mathbb{E}[\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'|\mathbf{z}_3]\|^2]=o(n),\\
&h_n^d\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\|^2\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)|\mathbf{z}_1]\|^2]=o(n^2h_n^d\wedge n),\\
&h_n^{2d}\mathbb{E}[\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)'|\mathbf{z}_1]\|^2]=o(n),\\
&h_n^{2d}\mathbb{E}[\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'|\mathbf{z}_2,\mathbf{z}_3]\|^2]=o((nh_n^d)^2\wedge 1),\\
&h_n^{2d}\mathbb{E}[\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'|\mathbf{z}_2,\mathbf{z}_3]\|\|\mathbf{s}_n^L(\mathbf{z}_2,\mathbf{z}_3)\|^2]=o(n),\\
&h_n^d\mathbb{E}[\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'|\mathbf{z}_2,\mathbf{z}_3]\|\|\mathbf{s}_n^L(\mathbf{z}_2,\mathbf{z}_3)\|\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_3,\mathbf{z}_1)|\mathbf{z}_3]\|]=o(n),\\
&h_n^d\mathbb{E}[\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'|\mathbf{z}_2,\mathbf{z}_3]\|\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_2]\|\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)|\mathbf{z}_3]\|]=o(1).
\end{aligned}$$

Proof. We have

$$\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\|\le b(\mathbf{z}_1)b(\mathbf{z}_2)\mathcal{K}_{h_n}(\mathbf{w}_1-\mathbf{w}_2),\qquad \mathcal{K}(\mathbf{u})=\sum_{l=0}^{L/2}|\lambda_l|K_{c_l}(\mathbf{u})$$

if $\boldsymbol{\theta}_{n,l}\in\Theta_0^\delta$ for each $l=0,\dots,L/2$ (which occurs for sufficiently large n by Lemma 2). Note that $\mathcal{K}$ is non-negative, bounded, symmetric, and integrable with respect to the Lebesgue measure.

Verification of $\mathbb{E}\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]'|\mathbf{z}_3]\|^2]=o(n)$. As shown in the proof of Lemma 5, $\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\||\mathbf{z}_1]\le Cb(\mathbf{z}_1)$. Then, $\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]'\|\le Cb(\mathbf{z}_1)^2b(\mathbf{z}_3)\mathcal{K}_{h_n}(\mathbf{w}_1-\mathbf{w}_3)$ and

$$\mathbb{E}[b(\mathbf{z}_1)^2b(\mathbf{z}_3)\mathcal{K}_{h_n}(\mathbf{w}_1-\mathbf{w}_3)|\mathbf{z}_3]=b(\mathbf{z}_3)\int\mathbb{E}[b(\mathbf{z}_1)^2|\mathbf{w}_1=\mathbf{w}_3+\mathbf{u}h]f(\mathbf{w}_3+\mathbf{u}h)\mathcal{K}(\mathbf{u})d\mathbf{u}\le Cb(\mathbf{z}_3).$$

Thus, $\mathbb{E}\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]'|\mathbf{z}_3]\|^2]\le C$.

Verification of $h_n^d\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\|^2\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)|\mathbf{z}_1]\|^2]=o(n^2h_n^d\wedge n)$. As above,

$$\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\|^2\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)|\mathbf{z}_1]\|^2\le Cb(\mathbf{z}_1)^4b(\mathbf{z}_2)^2\mathcal{K}_{h_n}(\mathbf{w}_1-\mathbf{w}_2)^2.$$

Then,

$$\mathbb{E}[b(\mathbf{z}_1)^4 b(\mathbf{z}_2)^2 \mathcal{K}_{h_n}(\mathbf{w}_1 - \mathbf{w}_2)^2] \leq h_n^{-d} C \int \mathbb{E}[b(\mathbf{z}_2)^2 | \mathbf{w}_2 = \mathbf{w}] f(\mathbf{w}) d\mathbf{w} \int \mathcal{K}(\mathbf{u})^2 d\mathbf{u}$$

and $h_n^d \mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_2)\|^2 \|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1, \mathbf{z}_3)|\mathbf{z}_1]\|^2] \leq C$.

Verification of $h_n^{2d}\mathbb{E}[\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)'|\mathbf{z}_1]\|^2] = O(1)$. We have

$$\|\mathbf{s}_n(\mathbf{z}_1, \mathbf{z}_2)\mathbf{s}_n(\mathbf{z}_1, \mathbf{z}_2)'\| \leq b(\mathbf{z}_1)^2 b(\mathbf{z}_2)^2 \mathcal{K}_{h_n}(\mathbf{w}_1 - \mathbf{w}_2)^2.$$

Then, $\mathbb{E}[b(\mathbf{z}_1)^2 b(\mathbf{z}_2)^2 \mathcal{K}_{h_n}(\mathbf{w}_1 - \mathbf{w}_2)^2 | \mathbf{z}_1] \leq C h_n^{-d} b(\mathbf{z}_1)^2$ and $h_n^{2d}\mathbb{E}[\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)'|\mathbf{z}_1]\|^2] \leq C$.

Verification of $h_n^{2d}\mathbb{E}[\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]^2] = o((nh_n^d)^2 \wedge 1)$. Using the argument for verifying $h_n^{2d}\mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]^2]] = o(1)$ in Lemma 5, we can show

$$h_n^d \mathbb{E}[\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]^2] \leq C$$

and

$$(nh_n^d)^{-2} h_n^{2d}\mathbb{E}[\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]^2] \leq \frac{C}{n^2 h_n^d} = o(1).$$

Verification of $h_n^{2d}\mathbb{E}[\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]\|\mathbf{s}_n(\mathbf{z}_2,\mathbf{z}_3)\|^2] = o(n)$.

$$\|\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_3)'\| \leq b(\mathbf{z}_1)^2 b(\mathbf{z}_2) b(\mathbf{z}_3) \mathcal{K}_{h_n}(\mathbf{w}_1 - \mathbf{w}_2)\mathcal{K}_{h_n}(\mathbf{w}_1 - \mathbf{w}_3)$$

and as in the proof of Lemma 5 (verification of $h_n^{2d}\mathbb{E}[\mathbb{E}[\|\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]^2] = o(1)$), we can show that

$$\mathbb{E}\big[b(\mathbf{z}_1)^2 b(\mathbf{z}_2) b(\mathbf{z}_3)\mathcal{K}_{h_n}(\mathbf{w}_1 - \mathbf{w}_2)\mathcal{K}_{h_n}(\mathbf{w}_1 - \mathbf{w}_3)\big|\mathbf{z}_2,\mathbf{z}_3\big] \leq h_n^{-d} C b(\mathbf{z}_2) b(\mathbf{z}_3) \bar{\mathcal{K}}\left(\frac{\mathbf{w}_3 - \mathbf{w}_2}{h_n}\right)$$

where $\bar{\mathcal{K}}(\mathbf{u}) = \int \mathcal{K}(\mathbf{v})\mathcal{K}(\mathbf{u} - \mathbf{v}) d\mathbf{v}$. Then,

$$\mathbb{E}[\|\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]\|\mathbf{s}_n(\mathbf{z}_2,\mathbf{z}_3)\|^2 \leq b(\mathbf{z}_2)^3 b(\mathbf{z}_3)^3 \bar{\mathcal{K}}_{h_n}(\mathbf{w}_3 - \mathbf{w}_2)\mathcal{K}_{h_n}(\mathbf{w}_2 - \mathbf{w}_3)^2$$

and

$$\mathbb{E}[b(\mathbf{z}_2)^3 b(\mathbf{z}_3)^3 \bar{\mathcal{K}}_{h_n}(\mathbf{w}_3 - \mathbf{w}_2)\mathcal{K}_{h_n}(\mathbf{w}_2 - \mathbf{w}_3)^2] \leq C h_n^{-2d}\mathbb{E}[b(\mathbf{z}_2)^3] \int \bar{\mathcal{K}}(\mathbf{u})\mathcal{K}(\mathbf{u})^2 d\mathbf{u}.$$

Then, $h_n^{2d}\mathbb{E}[\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]\|\mathbf{s}_n(\mathbf{z}_2,\mathbf{z}_3)\|^2] \leq C$.

Verification of $h_n^d\mathbb{E}[\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]\|\mathbf{s}_n^L(\mathbf{z}_2,\mathbf{z}_3)\|\|\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_3,\mathbf{z}_1)|\mathbf{z}_3]\|] = o(n)$.

$$\mathbb{E}[\|\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3] \leq C b(\mathbf{z}_2) b(\mathbf{z}_3)\bar{\mathcal{K}}_{h_n}(\mathbf{w}_3 - \mathbf{w}_2)$$

and $\|\mathbb{E}[\mathbf{s}_n(\mathbf{z}_3,\mathbf{z}_1)|\mathbf{z}_3]\| \le Cb(\mathbf{z}_3)$. Then,

$$\begin{aligned}&\mathbb{E}[\|\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_3)'\||\mathbf{z}_2,\mathbf{z}_3]\|\mathbf{s}_n(\mathbf{z}_2,\mathbf{z}_3)\|\|\mathbb{E}[\mathbf{s}_n(\mathbf{z}_3,\mathbf{z}_1)|\mathbf{z}_3]\|\\&\le Cb(\mathbf{z}_2)^2 b(\mathbf{z}_3)^3\bar{\mathcal{K}}_{h_n}(\mathbf{w}_3-\mathbf{w}_2)\mathcal{K}_{h_n}(\mathbf{w}_2-\mathbf{w}_3)\end{aligned}$$

and $\mathbb{E}\big[b(\mathbf{z}_2)^2 b(\mathbf{z}_3)^3\bar{\mathcal{K}}_{h_n}(\mathbf{w}_3-\mathbf{w}_2)\mathcal{K}_{h_n}(\mathbf{w}_2-\mathbf{w}_3)\big] \le Ch_n^{-d}$. Thus, the desired conclusion holds.

Verification of $h_n^d\mathbb{E}[\|\mathbb{E}[\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_3)'|\mathbf{z}_2,\mathbf{z}_3]\|\|\mathbb{E}[\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_2]\|\|\mathbb{E}[\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_3)|\mathbf{z}_3]\|] = o(1)$.

$$\|\mathbb{E}[\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_2)\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_3)'|\mathbf{z}_2,\mathbf{z}_3]\|\|\mathbb{E}[\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_2]\|\|\mathbb{E}[\mathbf{s}_n(\mathbf{z}_1,\mathbf{z}_3)|\mathbf{z}_3]\| \le Cb(\mathbf{z}_2)^2 b(\mathbf{z}_3)^2\bar{\mathcal{K}}_{h_n}(\mathbf{w}_2-\mathbf{w}_3)$$

and $\mathbb{E}[b(\mathbf{z}_2)^2 b(\mathbf{z}_3)^2\bar{\mathcal{K}}_{h_n}(\mathbf{w}_2-\mathbf{w}_3)] \le C$. Thus, the desired conclusion follows. □

Lemma 9. *Under Assumptions 1 and 2,*

$$\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\sum_{k=1,k\neq j,i}^n\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\right)\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_k)\right)' = \frac{1}{4}\mathbf{\Sigma}_0 + o_{\mathbb{P}}\left(\frac{1}{nh_n^d}+1\right)$$

and

$$\frac{1}{n^3}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\right)\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\right)' = \frac{1}{nh_n^d}\big[\mathbf{\Delta}_0(K_{\mathbf{c}}) + o_{\mathbb{P}}(1)\big].$$

Proof. By Hoeffding decomposition,

$$\begin{aligned}&\frac{1}{n(n-1)(n-2)}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\sum_{p=1,p\neq j,i}^n\boldsymbol{\mu}_1'\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\right)\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_p)\right)'\boldsymbol{\mu}_1\\
&=\boldsymbol{\mu}_1'\mathbb{E}\left[\sum_{l,\tilde{l}=0}^{L/2}\lambda_l\lambda_{\tilde{l}}\mathbb{E}\big[\mathbf{s}_{n,l}(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1\big]\mathbb{E}\big[\mathbf{s}_{n,\tilde{l}}(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1\big]'\right]\boldsymbol{\mu}_1\\
&\quad+\frac{1}{n}\sum_{i=1}^n 3\left(\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i]-\bar{\zeta}_1\right)+\binom{n}{2}^{-1}\sum_{i<j}3\left(\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i,\mathbf{z}_j]-\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i]-\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_j]+\bar{\zeta}_1\right)\\
&\quad+\binom{n}{3}^{-1}\sum_{i<j<p}\big(\zeta_{1,ijp}-\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i,\mathbf{z}_j]-\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i,\mathbf{z}_p]-\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_j,\mathbf{z}_p]+\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i]\\
&\qquad\qquad+\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_j]+\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_p]-\bar{\zeta}_1\big)\end{aligned}$$

where

$$\zeta_{1,ijp} = \boldsymbol{\mu}_1'\frac{\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_p)'+\mathbf{s}_n^L(\mathbf{z}_j,\mathbf{z}_i)\mathbf{s}_n^L(\mathbf{z}_j,\mathbf{z}_p)'+\mathbf{s}_n^L(\mathbf{z}_p,\mathbf{z}_j)\mathbf{s}_n^L(\mathbf{z}_p,\mathbf{z}_i)'}{3}\boldsymbol{\mu}_1,$$

$\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2) = \sum_{l=0}^L\lambda_l\mathbf{s}_{n,l}L(\mathbf{z}_1,\mathbf{z}_2)$, and $\bar{\zeta}_1 = \mathbb{E}[\zeta_{1,ijp}]$. Using identical arguments for verifying (4.3), the expectation term in the above Hoeffding decomposition converges to $\boldsymbol{\mu}_1'\boldsymbol{\Sigma}_0\boldsymbol{\mu}_1/4$. For the mean-zero U-statistic terms, it suffices to show that their variances are $o_{\mathbb{P}}(1+(nh_n^d)^{-1})$. Lemmas 5 and 8 imply

$$\mathbb{V}[\mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_i]] \leq \mathbb{E}\left[\mathbb{E}\big[\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|\mathbf{z}_i\big]^4\right] + \mathbb{E}\left[\mathbb{E}\big[\mathbb{E}\big[\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_j,\mathbf{z}_k)|\mathbf{z}_j\big]\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_j,\mathbf{z}_i)|\mathbf{z}_i\big]^2\right] = o(n).$$

Lemma 8 implies

$$\begin{aligned}&\mathbb{V}[\mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_i,\mathbf{z}_j] - \mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_i] - \mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_j] + \bar{\zeta}_1] \leq C\mathbb{E}\big[\mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_i,\mathbf{z}_j]^2\big]\\ &\leq C\mathbb{E}\big[(\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j))^2\mathbb{E}[\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|\mathbf{z}_i]^2\big] + C\mathbb{E}\big[\mathbb{E}[\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_k,\mathbf{z}_i)\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_k,\mathbf{z}_j)|\mathbf{z}_i,\mathbf{z}_j]^2\big] = o(n^2).\end{aligned}$$

Lemma 8 implies

$$\begin{aligned}&\mathbb{V}\big[\zeta_{1,ijk} - \mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_i,\mathbf{z}_j] - \mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_i,\mathbf{z}_k] - \mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_j,\mathbf{z}_k] + \mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_i] + \mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_j] + \mathbb{E}[\zeta_{1,ijk}|\mathbf{z}_k]\big]\\ &\leq C\mathbb{E}[\zeta_{1,ijk}^2] \leq C\mathbb{E}\big[\mathbb{E}[|\boldsymbol{\mu}_1'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)|^2|\mathbf{z}_i]^2\big] = o(nh_n^{-2d}).\end{aligned}$$

By Hoeffding decomposition,

$$\begin{aligned}&\binom{n}{2}^{-1}\sum_{i<j}\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)'\boldsymbol{\mu}_2\\ &= \boldsymbol{\mu}_2'\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)']\boldsymbol{\mu}_2 + \frac{1}{n}\sum_{i=1}^n 2\big(\mathbb{E}[(\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j))^2|\mathbf{z}_i] - \mathbb{E}[(\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2))^2]\big)\\ &\quad + \binom{n}{2}^{-1}\sum_{i<j}(\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j))^2 - \mathbb{E}[(\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j))^2|\mathbf{z}_i] - \mathbb{E}[(\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j))^2|\mathbf{z}_j] + \mathbb{E}[(\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j))^2].\end{aligned}$$

Using identical arguments for verifying (4.3), $h_n^d\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j)']$ converges to $\boldsymbol{\Delta}_0(K_{\mathbf{c}})$, and the remaining U-statistic terms are $o_{\mathbb{P}}(n)$ by $\mathbb{E}[\mathbb{E}[(\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j))^2|\mathbf{z}_i]^2] = o(n^3)$ and $\mathbb{E}[(\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j))^4] = o(n^2h_n^{-d})$ because $n^2h_n^d \to \infty$. □

4.4 Proof of Proposition 1

Since

$$\frac{M_0(\boldsymbol{\theta}_0+\mathbf{t}\tau) - M_0(\boldsymbol{\theta}_0) - \mathbb{E}[\boldsymbol{\psi}(\mathbf{w},\mathbf{w})]'\mathbf{t}\tau}{\tau} = \int \mathbb{E}[e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0,\mathbf{t},\tau)|\mathbf{w}_1=\mathbf{w},\mathbf{w}_2=\mathbf{w}]f_{\mathbf{w}}(\mathbf{w})^2d\mathbf{w},$$

Assumption 2(iii)(iv) implies that $\mathbb{E}[\boldsymbol{\psi}(\mathbf{w},\mathbf{w})]$ is the (directional) derivative of $M_0(\boldsymbol{\theta})$ at $\boldsymbol{\theta}_0$, and thus, $\mathbb{E}[\boldsymbol{\psi}(\mathbf{w},\mathbf{w})] = \mathbf{0}$ because M_0 is minimized at $\boldsymbol{\theta}_0$, which lies in the interior of Θ.

For the following result, we use multi-index notation as introduced in the paragraph before Proposition 1.

Lemma 10. *Let $\boldsymbol{\psi}(\mathbf{w}_1,\mathbf{w}_2) = \mathbb{E}[\mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)|\mathbf{w}_1,\mathbf{w}_2]$ and $L \geq 2$ be an even integer. Suppose that Assumption 1(i)(iii) holds and that $\mathbf{v} \mapsto \boldsymbol{\psi}(\mathbf{w},\mathbf{v})$ is L-times continuously differentiable with*

$\mathbb{E}[\sup_{\mathbf{v}\in\mathcal{W}} \|\partial_{\mathbf{v}}^{\boldsymbol{\alpha}} \boldsymbol{\psi}(\mathbf{w},\mathbf{v})\| < \infty$ *for each* $|\boldsymbol{\alpha}| \le L$*. Then, there exist non-random vectors* $\mathbf{b}_{2l}^M \in \mathbb{R}^k$ $l = 1, \ldots, L/2$ *such that*

$$\mathbb{E}[\mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)K_h(\mathbf{w}_1-\mathbf{w}_2)] = \sum_{l=1}^{L/2} \mathbf{b}_{2l}^M h^{2l} + o(h^L).$$

Proof. As just shown, $\mathbb{E}[\psi(\mathbf{w},\mathbf{w})] = \mathbf{0}$. Then

$$\begin{aligned}\mathbb{E}[\mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)K_h(\mathbf{w}_1-\mathbf{w}_2)] &= \sum_{|\boldsymbol{\alpha}|\le L} h^{|\boldsymbol{\alpha}|}(\boldsymbol{\alpha}!)^{-1}\mathbb{E}\Big[\partial_{\mathbf{v}}^{\boldsymbol{\alpha}}\boldsymbol{\psi}(\mathbf{w},\mathbf{v})\big|_{\mathbf{v}=\mathbf{w}}\Big]\int \mathbf{u}^{\boldsymbol{\alpha}}K(\mathbf{u})d\mathbf{u}\\ &+ h^L \sum_{|\boldsymbol{\alpha}|=L}(\boldsymbol{\alpha}!)^{-1}\int\Big(\partial_{\mathbf{v}}^{\boldsymbol{\alpha}}\boldsymbol{\psi}(\mathbf{w},\mathbf{v})\big|_{\mathbf{v}=\mathbf{w}-\tau\mathbf{u}h} - \partial_{\mathbf{v}}^{\boldsymbol{\alpha}}\boldsymbol{\psi}(\mathbf{w},\mathbf{v})\big|_{\mathbf{v}=\mathbf{w}}\Big) f_{\mathbf{w}}(\mathbf{w})d\mathbf{w}\mathbf{u}^{\boldsymbol{\alpha}}K(\mathbf{u})d\mathbf{u}\end{aligned}$$

where $\tau \in (0,1)$ denotes a mean value which may depend on $\mathbf{w}$ and $\mathbf{u}h$. The desired result follows from the dominated convergence theorem. Note that $\int \mathbf{u}^{\boldsymbol{\alpha}}K(\mathbf{u}) = 0$ when at least of one element of $\boldsymbol{\alpha} \in \mathbb{Z}_+^d$ is odd. □

In the context of Proposition 1,

$$\mathbf{b}_2^M = \sum_{i=1}^d 2^{-1}\int \frac{\partial^2\boldsymbol{\psi}(\mathbf{w},\mathbf{v})}{\partial v_i^2}\Big|_{\mathbf{v}=\mathbf{w}} f_{\mathbf{w}}(\mathbf{w})d\mathbf{w}\int u_i^2 K(\mathbf{u})d\mathbf{u}.$$

The above expansion and Lemma 3 imply that as $h \downarrow 0$,

$$h^{-4}\big[M(\boldsymbol{\theta}_0 + \mathbf{t}h^2; h_n) - M(\boldsymbol{\theta}_0; h)\big] - \mathbf{t}'\mathbf{b}_2^M - \frac{1}{2}\mathbf{t}'\mathbf{H}_0\mathbf{t} = o(1).$$

Since $\mathbf{t} \mapsto M(\boldsymbol{\theta}_0 + \mathbf{t}h^2; h) - M(\boldsymbol{\theta}_0; h)$ is convex and its minimizer is $h^{-2}(\boldsymbol{\theta}(h) - \boldsymbol{\theta}_0)$, the corollary following Lemma 2 of Hjort and Pollard (1993) implies

$$h^{-2}\big(\boldsymbol{\theta}(h) - \boldsymbol{\theta}_0\big) - \big(-\mathbf{H}_0^{-1}\mathbf{b}_2^M\big) = o(1) \quad \Rightarrow \quad \boldsymbol{\theta}(h) = \boldsymbol{\theta}_0 - \mathbf{H}_0^{-1}\mathbf{b}_2^M h^2 + o(h^2).$$

□

4.5 Bias Expansion

We demonstrate how to verify Assumption 3 with $L = 4$ under the following primitive conditions.

Assumption 4. Assumptions 1 and 2 hold. There exist $\mathbb{R}^k$-valued functions $\dot{\mathbf{h}}_{ij}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t})$, $1 \le i \le j \le k$, such that $\dot{\mathbf{h}}_{ij}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta}_0,\mathbf{t})$ is continuous with respect to $\mathbf{w}_2$ with probability one, $\mathbb{E}\sup_{\mathbf{v}\in\mathcal{W}}\|\dot{\mathbf{h}}_{ij}(\mathbf{w}_1,\mathbf{v};\boldsymbol{\theta}_0,\mathbf{t})f_{\mathbf{w}}(\mathbf{v})\| < \infty$, and letting $\mathbf{E}_3(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t},\tau)$ be the $k\times k$ symmetric matrix whose (i,j)th element is

$$\frac{H_{ij}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta}+\mathbf{t}\tau,\mathbf{t}) - H_{ij}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t}) - \dot{\mathbf{h}}_{ij}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t})'\mathbf{t}\tau}{\tau},$$

we have $\lim_{(\mathbf{u},\boldsymbol{\theta},\tau)\to(\mathbf{0},\boldsymbol{\theta}_0,0)} \mathbf{E}_3(\mathbf{w}_1,\mathbf{w}_1+\mathbf{u};\boldsymbol{\theta},\mathbf{t},\tau) = \mathbf{O}$ with probability one and for $\mathbf{t}\in\mathbb{R}^k$, there is $\bar\tau>0$ such that

$$\mathbb{E} \sup_{\mathbf{v}\in\mathcal{W},\boldsymbol{\theta}\in\Theta_0^\delta,\tau\in(0,\bar\tau)} \big\|\mathbf{E}_3(\mathbf{w}_1,\mathbf{v};\boldsymbol{\theta},\mathbf{t},\tau)\big\| < \infty.$$

$\mathbf{v}\mapsto\boldsymbol{\psi}(\mathbf{w},\mathbf{w}+\mathbf{v})$ satisfies the hypothesis of Lemma 10 with $L=4$.

$\mathbf{v}\mapsto f_{\mathbf{w}}(\mathbf{v})$ and $\mathbf{v}\mapsto\mathbf{H}(\mathbf{w},\mathbf{v};\boldsymbol{\theta}_0,\mathbf{t})$ are twice continuously differentiable and for $|\boldsymbol{\alpha}|\le 2$,

$$\mathbb{E}\sup_{\mathbf{v}\in\mathcal{W}} \big\|\partial_{\mathbf{v}}^{\boldsymbol{\alpha}}\big\{\mathbf{H}(\mathbf{w}_1,\mathbf{v};\boldsymbol{\theta}_0,\mathbf{t})f_{\mathbf{w}}(\mathbf{v})\big\}\big\| < \infty.$$

The following proposition generalizes Proposition 1 to the case of $L=4$. In particular, it demonstrates that the term associated with the third power of h is equal to zero.

Proposition 2. *Under Assumption 4, Assumption 3 holds with $L=4$.*

Proof. We have $\mathbf{b}_2 = -\mathbf{H}_0^{-1}\mathbf{b}_2^M$. Let $\partial M(\boldsymbol{\theta};h) = \mathbb{E}[\mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})K_h(\mathbf{w}_1-\mathbf{w}_2)]$ and $\mathbf{H}(\boldsymbol{\theta},\mathbf{t};\eta) = \mathbb{E}[\mathbf{H}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta}_0,\mathbf{t})K_\eta(\mathbf{w}_1-\mathbf{w}_2)]$. For $h,\eta>0$ close to zero, by Taylor expansion,

$$\partial M(\boldsymbol{\theta}_0+\mathbf{b}_2h^2;\eta) = \partial M(\boldsymbol{\theta}_0;\eta) + \mathbf{H}(\boldsymbol{\theta}_0,\mathbf{b}_2;\eta)\mathbf{b}_2h^2 + \mathfrak{b}_{4,1}h^4 + o(h^4) \tag{4.12}$$

where $\mathfrak{b}_{4,1} = \sum_{i=1}^d 2^{-1}\int \partial\mathbf{H}(\mathbf{w},\mathbf{w};\boldsymbol{\theta};\mathbf{b}_2)/\partial\theta_i|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} f_{\mathbf{w}}(\mathbf{w})^2 d\mathbf{w}\mathbf{b}_2 b_{2,i}$. By the expansion $\mathbf{H}(\boldsymbol{\theta}_0,\mathbf{b}_2;\eta) = \mathbf{H}_0 + O(\eta^2)$,

$$\mathbf{H}(\boldsymbol{\theta}_0,\mathbf{b}_2;\eta)\mathbf{b}_2 = -\mathbf{b}_2^M + \mathfrak{b}_{4,2}\eta^2 + o(\eta^2)$$

where $\mathfrak{b}_{4,2} = -\sum_{i=1}^d 2^{-1}\int \partial^2\mathbf{H}(\mathbf{w},\mathbf{v};\boldsymbol{\theta}_0,\mathbf{b}_2)/\partial v_i^2|_{\mathbf{v}=\mathbf{w}} f_{\mathbf{w}}(\mathbf{w})^2 d\mathbf{w}\int u_i^2K(\mathbf{u})d\mathbf{u}\mathbf{H}_0^{-1}\mathbf{b}_2^M$. Then, using Lemma 10 and (4.12),

$$\partial M(\boldsymbol{\theta}_0+\mathbf{b}_2h^2;\eta) = \mathbf{b}_2^M\eta^2 + \mathbf{b}_4^M\eta^4 - \mathbf{b}_2^Mh^2 + \mathfrak{b}_{4,1}h^4 + \mathfrak{b}_{4,2}\eta^2h^2 + o\big(\eta^4+h^4\big).$$

Then,

$$\partial M(\boldsymbol{\theta}_0+\mathbf{b}_2h^2;h) = \big(\mathbf{b}_4^M + \mathfrak{b}_{4,1} + \mathfrak{b}_{4,2}\big)h^4 + o(h^4).$$

By Lemma 3, as $h\downarrow 0$,

$$h^{-8}\big\{M(\boldsymbol{\theta}_0+\mathbf{b}_2h^2+\mathbf{t}h^4;h) - M(\boldsymbol{\theta}_0+\mathbf{b}_2h^2;h)\big\} - h^{-4}\partial M(\boldsymbol{\theta}_0+\mathbf{b}_2h^2;h)'\mathbf{t} = \frac{1}{2}\mathbf{t}'\mathbf{H}_0\mathbf{t} + o(1).$$

Since $h^{-4}\partial M(\boldsymbol{\theta}_0+\mathbf{b}_2h^2;h)'\mathbf{t} + \frac{1}{2}\mathbf{t}'\mathbf{H}_0\mathbf{t} = Q(t) + o(1)$ where

$$Q(\mathbf{t}) = \big(\mathbf{b}_4^M + \mathfrak{b}_{4,1} + \mathfrak{b}_{4,2}\big)'\mathbf{t} + \frac{1}{2}\mathbf{t}'\mathbf{H}_0\mathbf{t},$$

the corollary following Lemma 2 of Hjort and Pollard (1993) implies

$$h^{-4}(\boldsymbol{\theta}(h) - \boldsymbol{\theta}_0 - \mathbf{b}_2h^2) + \mathbf{H}_0^{-1}\big(\mathbf{b}_4^M + \mathfrak{b}_{4,1} + \mathfrak{b}_{4,2}\big) = o(1)$$

and we have the desired conclusion with $\mathbf{b}_4 = -\mathbf{H}_0^{-1}\big(\mathbf{b}_4^M + \mathfrak{b}_{4,1} + \mathfrak{b}_{4,2}\big)$. □

Propositions 1 and 2 verified Assumption 3 for $L = 2$ and $L = 4$, respectively, under primitive conditions. The approach underlying those propositions could be extended to verify Assumption 3 for $L > 4$ at the expense of additional cumbersome notation and technical work.

5 Sufficient Conditions for Motivating Examples

To demonstrate the plausibility of Assumptions 1 and 2, we revisit the examples of Section 2. In each example, Assumptions 1(ii) holds and Assumption 1(iii) is fairly primitive, so we focus on giving primitive sufficient conditions for Assumptions 1(iv)-(v) and 2.

5.1 Partially Linear Regression Model

We take $\mathbf{s} = \mathbf{s}_{\text{PLR}}$ and $\mathbf{H} = \mathbf{H}_{\text{PLR}}$, where

$$\begin{aligned}\mathbf{s}_{\text{PLR}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} m_{\text{PLR}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) \\ &= -\dot{\mathbf{x}}_{i,j}(\dot{y}_{i,j} - \dot{\mathbf{x}}_{i,j}' \boldsymbol{\theta})\end{aligned}$$

and

$$\begin{aligned}\mathbf{H}_{\text{PLR}}(\mathbf{w}_i, \mathbf{w}_j) &= \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \mathbb{E}\left[m_{\text{PLR}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) \middle| \mathbf{w}_i, \mathbf{w}_j \right] = \frac{\partial}{\partial \boldsymbol{\theta}'} \mathbb{E}\left[s_{\text{PLR}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) \middle| \mathbf{w}_i, \mathbf{w}_j \right] \\ &= \mathbb{E}\left[\dot{\mathbf{x}}_{i,j} \dot{\mathbf{x}}_{i,j}' \middle| \mathbf{w}_i, \mathbf{w}_j \right],\end{aligned}$$

the latter depending on neither $\boldsymbol{\theta}$ nor $\mathbf{t}$ (because $m_{\text{PLR}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta})$ is quadratic in $\boldsymbol{\theta}$).

Under mild conditions, Assumptions 1(iv)-(v) and 2 hold with

$$\boldsymbol{\xi}_0(\mathbf{z}) = 2\mathbb{E}[\mathbf{s}_{\text{PLR}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0) | \mathbf{z}_1 = \mathbf{z}, \mathbf{w}_2 = \mathbf{w}] f_{\mathbf{w}}(\mathbf{w}) = -2\left(\mathbf{x} - \mathbb{E}[\mathbf{x}|\mathbf{w}]\right) f_{\mathbf{w}}(\mathbf{w})\varepsilon,$$

$$\begin{aligned}\boldsymbol{\Xi}_0(\mathbf{w}) &= \mathbb{E}[\mathbf{s}_{\text{PLR}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)\mathbf{s}_{\text{PLR}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)' | \mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}] f_{\mathbf{w}}(\mathbf{w}) \\ &= 2\mathbb{V}\left[\dot{\mathbf{x}}_{1,2}\varepsilon_1 \middle| \mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w} \right] f_{\mathbf{w}}(\mathbf{w}),\end{aligned}$$

and

$$\mathbf{G}_0(\mathbf{w}) = \mathbf{H}_{\text{PLR}}(\mathbf{w}, \mathbf{w}) f_{\mathbf{w}}(\mathbf{w}) = 2\mathbb{V}[\mathbf{x}|\mathbf{w}] f_{\mathbf{w}}(\mathbf{w}).$$

For instance, it suffices to set $b(\mathbf{z}) = (1 + \|\mathbf{x}\|)(1 + |\varepsilon| + |\eta_0(\mathbf{w})| + \|\mathbf{x}\|)$ and to assume that

(i) The functions $\mathbf{w} \mapsto \eta_0(\mathbf{w})$, $\mathbf{w} \mapsto \mathbb{E}[\mathbf{x}|\mathbf{w}]$, $\mathbf{w} \mapsto \mathbb{E}[\mathbf{x}\mathbf{x}'|\mathbf{w}]$, $\mathbf{w} \mapsto \mathbb{E}[\varepsilon^2|\mathbf{w}]$, $\mathbf{w} \mapsto \mathbb{E}[\mathbf{x}\varepsilon^2|\mathbf{w}]$, and $\mathbf{w} \mapsto \mathbb{E}[\mathbf{x}\mathbf{x}'\varepsilon^2|\mathbf{w}]$ are continuous on $\mathcal{W}$.

(ii)
$$\mathbb{E}\left[(1 + \|\mathbf{x}\|^4)\varepsilon^4\right] + \sup_{\mathbf{w} \in \mathcal{W}} \mathbb{E}\left[(1 + \|\mathbf{x}\|^4)\varepsilon^4 \middle| \mathbf{w}\right] f_{\mathbf{w}}(\mathbf{w}) < \infty$$

and

$$\mathbb{E}\left[(1+\|\mathbf{x}\|^4)\eta_0(\mathbf{w})^4+\|\mathbf{x}\|^8\right]+\sup_{\mathbf{w}\in\mathcal{W}}\mathbb{E}\left[(1+\|\mathbf{x}\|^4)\eta_0(\mathbf{w})^4+\|\mathbf{x}\|^8\middle|\,\mathbf{w}\right]f_{\mathbf{w}}(\mathbf{w})<\infty.$$

(iii) With probability one, $\mathbb{V}[\mathbf{x}|\mathbf{w}]$ is positive definite and $\mathbb{V}[\varepsilon|\mathbf{x},\mathbf{w}]>0$.

5.2 Partially Linear Logit Model

We take $\mathbf{s}=\mathbf{s}_{\text{PLL}}$ and $\mathbf{H}=\mathbf{H}_{\text{PLL}}$, where

$$\begin{aligned}\mathbf{s}_{\text{PLL}}(\mathbf{z}_i,\mathbf{z}_j;\boldsymbol{\theta})&=\frac{\partial}{\partial\boldsymbol{\theta}}m_{\text{PLL}}(\mathbf{z}_i,\mathbf{z}_j;\boldsymbol{\theta})\\&=-\mathbb{1}\{\dot{y}_{i,j}\neq 0\}\dot{\mathbf{x}}_{i,j}\left(y_i-\Lambda(\dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta})\right)\end{aligned}$$

and

$$\begin{aligned}\mathbf{H}_{\text{PLL}}(\mathbf{w}_i,\mathbf{w}_j;\boldsymbol{\theta})&=\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}'}\mathbb{E}\left[m_{\text{PLL}}(\mathbf{z}_i,\mathbf{z}_j;\boldsymbol{\theta})|\,\mathbf{w}_i,\mathbf{w}_j\right]=\frac{\partial}{\partial\boldsymbol{\theta}'}\mathbb{E}\left[s_{\text{PLL}}(\mathbf{z}_i,\mathbf{z}_j;\boldsymbol{\theta})|\,\mathbf{w}_i,\mathbf{w}_j\right]\\&=\mathbb{E}\left[\dot{\mathbf{x}}_{i,j}\dot{\mathbf{x}}_{i,j}'\lambda(\dot{\mathbf{x}}_{i,j}'\boldsymbol{\theta})\mathbb{1}\{\dot{y}_{i,j}\neq 0\}\middle|\,\mathbf{w}_i,\mathbf{w}_j\right],\end{aligned}$$

where the latter does not depend on $\mathbf{t}$ (because $m_{\text{PLL}}(\mathbf{z}_i,\mathbf{z}_j;\boldsymbol{\theta})$ is twice differentiable in $\boldsymbol{\theta}$).

Under mild conditions, Assumptions 1(iv)-(v) and 2 hold with

$$\boldsymbol{\xi}_0(\mathbf{z})=2\mathbb{E}[\mathbf{s}_{\text{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)|\mathbf{z}_1=\mathbf{z},\mathbf{w}_2=\mathbf{w}]f_{\mathbf{w}}(\mathbf{w}),$$

$$\boldsymbol{\Xi}_0(\mathbf{w})=\mathbb{E}[\mathbf{s}_{\text{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)\mathbf{s}_{\text{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)'|\mathbf{w}_1=\mathbf{w}_2=\mathbf{w}]f_{\mathbf{w}}(\mathbf{w}),$$

and

$$\mathbf{G}_0(\mathbf{w})=\mathbf{H}_{\text{PLL}}(\mathbf{w},\mathbf{w};\boldsymbol{\theta}_0)f_{\mathbf{w}}(\mathbf{w}).$$

For instance, it suffices to set $b(\mathbf{z})=1+\|\mathbf{x}\|$ and to assume that, for some $\delta>0$,

(i) The function $\mathbf{w}\mapsto\eta_0(\mathbf{w})$ is continuous on $\mathcal{W}$. Also, the conditional distribution of $\mathbf{x}$ given $\mathbf{w}$ admits a density $f_{\mathbf{x}|\mathbf{w}}$ with respect to some measure ρ such that $\mathbf{w}\mapsto f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w})$ is continuous on $\mathcal{W}$ (with probability one) and

$$\int_{\mathbb{R}^k}(1+\|\mathbf{x}\|^2)\sup_{\|\mathbf{u}\|\leq\delta}f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w}+\mathbf{u})d\rho(\mathbf{x})<\infty\qquad\text{for every }\mathbf{w}\in\mathcal{W}.$$

(ii) $\mathbb{E}\left[\|\mathbf{x}\|^4\right]+\sup_{\mathbf{w}\in\mathcal{W}}\mathbb{E}\left[\|\mathbf{x}\|^4\middle|\,\mathbf{w}\right]f_{\mathbf{w}}(\mathbf{w})<\infty.$

(iii) With probability one, $\mathbb{V}[\mathbf{x}|\mathbf{w}]$ is positive definite.

5.3 Partially Linear Tobit Model

We take $\mathbf{s} = \mathbf{s}_{\text{PLT}}$ and $\mathbf{H} = \mathbf{H}_{\text{PLT}}$, where (in slight abuse of notation)

$$\begin{aligned}\mathbf{s}_{\text{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) &= \frac{\partial}{\partial \boldsymbol{\theta}} m_{\text{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) \\ &= \dot{\mathbf{x}}_{i,j} \mathbb{1}\{y_j > \max\left(y_i - \dot{\mathbf{x}}_{i,j}' \boldsymbol{\theta}, 0\right)\} - \dot{\mathbf{x}}_{i,j} \mathbb{1}\{y_i > \max\left(y_j + \dot{\mathbf{x}}_{i,j}' \boldsymbol{\theta}, 0\right)\}\end{aligned}$$

and

$$\begin{aligned}&\mathbf{H}_{\text{PLT}}(\mathbf{w}_i, \mathbf{w}_j; \boldsymbol{\theta}) = \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \mathbb{E}\left[m_{\text{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) \middle| \mathbf{w}_i, \mathbf{w}_j \right] = \frac{\partial}{\partial \boldsymbol{\theta}'} \mathbb{E}\left[s_{\text{PLT}}(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta}) \middle| \mathbf{w}_i, \mathbf{w}_j \right] \\ &= \mathbb{E}\left[\dot{\mathbf{x}}_{i,j} \dot{\mathbf{x}}_{i,j}' \int_0^\infty f_{\varepsilon|\mathbf{w}}(\varepsilon - \mathbf{x}_j' \boldsymbol{\theta}_0 - \eta_0(\mathbf{w}_j) | \mathbf{w}_j) f_{\varepsilon|\mathbf{w}}(\varepsilon - \mathbf{x}_i' \boldsymbol{\theta}_0 - \eta_0(\mathbf{w}_i) + \dot{\mathbf{x}}_{i,j}' \boldsymbol{\theta} | \mathbf{w}_i) d\varepsilon \middle| \mathbf{w}_i, \mathbf{w}_j \right] \\ &\quad + \mathbb{E}\left[\dot{\mathbf{x}}_{i,j} \dot{\mathbf{x}}_{i,j}' \int_0^\infty f_{\varepsilon|\mathbf{w}}(\varepsilon - \mathbf{x}_i' \boldsymbol{\theta}_0 - \eta_0(\mathbf{w}_i) | \mathbf{w}_i) f_{\varepsilon|\mathbf{w}}(\varepsilon - \mathbf{x}_j' \boldsymbol{\theta}_0 - \eta_0(\mathbf{w}_j) - \dot{\mathbf{x}}_{i,j}' \boldsymbol{\theta} | \mathbf{w}_j) d\varepsilon \middle| \mathbf{w}_i, \mathbf{w}_j \right].\end{aligned}$$

Under mild conditions, Assumptions 1(iv)-(v) and 2 hold with

$$\boldsymbol{\xi}_0(\mathbf{z}) = 2\mathbb{E}[\mathbf{s}_{\text{PLT}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0) | \mathbf{z}_1 = \mathbf{z}, \mathbf{w}_2 = \mathbf{w}] f_{\mathbf{w}}(\mathbf{w}),$$

$$\boldsymbol{\Xi}_0(\mathbf{w}) = \mathbb{E}[\mathbf{s}_{\text{PLT}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0) \mathbf{s}_{\text{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)' | \mathbf{w}_1 = \mathbf{w}_2 = \mathbf{w}] f_{\mathbf{w}}(\mathbf{w}),$$

and

$$\mathbf{G}_0(\mathbf{w}) = \mathbf{H}_{\text{PLT}}(\mathbf{w}, \mathbf{w}; \boldsymbol{\theta}_0) f_{\mathbf{w}}(\mathbf{w}).$$

For instance, it suffices to set $b(\mathbf{z}) = 1 + \|\mathbf{x}\|$ and to assume that, for some $\delta > 0$,

(i) The function $\mathbf{w} \mapsto \eta_0(\mathbf{w})$ is continuous on $\mathcal{W}$. Also, the conditional distribution of $\mathbf{x}$ given $\mathbf{w}$ admits a a density $f_{\mathbf{x}|\mathbf{w}}$ with respect to some measure ρ such that $\mathbf{w} \mapsto f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w})$ is continuous on $\mathcal{W}$ (with probability one) and

$$\int_{\mathbb{R}^k} (1 + \|\mathbf{x}\|^2) \sup_{\|\mathbf{u}\| \le \delta} f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w} + \mathbf{u}) d\rho(\mathbf{x}) < \infty \qquad \text{for every } \mathbf{w} \in \mathcal{W}.$$

In addition, the function $(\varepsilon, \mathbf{w}) \mapsto f_{\varepsilon|\mathbf{w}}(\varepsilon|\mathbf{w})$ is continuous and bounded and the function

$$(\mathbf{x}, \mathbf{w}) \mapsto \int_{\mathbb{R}} \sup_{|u| + \|\mathbf{u}\| \le \delta} f_{\varepsilon|\mathbf{w}}(\varepsilon - \mathbf{x}' \boldsymbol{\theta}_0 - \eta_0(\mathbf{w}) + u | \mathbf{w} + \mathbf{u}) d\varepsilon$$

is bounded.

(ii) $\mathbb{E}\left[\|\mathbf{x}\|^4\right] + \sup_{\mathbf{w} \in \mathcal{W}} \left(1 + \mathbb{E}\left[\|\mathbf{x}\|^4 \middle| \mathbf{w}\right]\right) f_{\mathbf{w}}(\mathbf{w}) < \infty$.

(iii) With probability one, $\mathbb{V}[\mathbf{x}|\mathbf{w}]$ is positive definite.

6 Conclusion

This paper has developed bandwidth robust distribution theory and bootstrap-based inference procedures for a broad class of convex pairwise difference estimators. Our theoretical work is based on small bandwidth asymptotics and carefully leverages convexity. The theory is illustrated by means of three prominent examples. In addition to expanding the scope of small bandwidth asymptotics, our results lay the groundwork for several promising avenues of future research. First, our methods could be generalized to develop bandwidth selection based on higher-order stochastic expansions. Second, they could be expanded to allow for pairwise difference estimators based on generated regressors, a class of estimators that sometimes arises in the context of control function and related econometric methods. Third, when the objective function is smooth, plug-in variance estimation could be developed as an alternative to bootstrap inference. Finally, our current results do not cover settings where the objective function is sufficiently non-smooth to result in non-Gaussian distributional approximations. We plan to investigate these research directions in upcoming work.

References

Ahn, H. and J. L. Powell (1993): "Semiparametric Estimation of Censored Selection Models with a Nonparametric Selection Mechanism," *Journal of Econometrics*, 58, 3–29.

Aradillas-Lopez, A. (2012): "Pairwise-Difference Estimation of Incomplete Information Games," *Journal of Econometrics*, 168, 120–140.

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

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

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

——— (2013): "Generalized Jackknife Estimators of Weighted Average Derivatives (with Discussions and Rejoinder)," *Journal of the American Statistical Association*, 108, 1243–1268.

——— (2014a): "Small Bandwidth Asymptotics for Density-Weighted Average Derivatives," *Econometric Theory*, 30, 176–200.

——— (2014b): "Bootstrapping Density-Weighted Average Derivatives," *Econometric Theory*, 30, 1135–1164.

Cattaneo, M. D., M. H. Farrell, M. Jansson, and R. P. Masini (2025a): "Higher-Order Refinements of Small Bandwidth Asymptotics for Density-Weighted Average Derivative Estimators," *Journal of Econometrics, forthcoming.*

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

——— (2022): "Average Density Estimators: Efficiency and Bootstrap Consistency," *Econometric Theory*, 38, 1140–1174.

Cattaneo, M. D., M. Jansson, and K. Nagasawa (2024): "Bootstrap-Assisted Inference for Generalized Grenander-type Estimators," *Annals of Statistics*, 52, 1509–1533.

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

Cattaneo, M. D., J. M. Klusowski, and W. G. Underwood (2025b): "Inference with Mondrian Random Forests," *arXiv preprint arXiv:2310.09702.*

Heyde, C. C. and B. M. Brown (1970): "On the Departure from Normality of a Certain Class of Martingales," *Annals of Mathematical Statistics*, 41, 2161–2165.

Hjort, N. L. and D. Pollard (1993): "Asymptotics for Minimisers of Convex Processes," *arXiv preprint arXiv:1107.3806.*

Hong, H. and M. Shum (2010): "Pairwise-Difference Estimation of a Dynamic Optimization Model," *Review of Economic Studies*, 77, 273–304.

Honoré, B. E. (1992): "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects," *Econometrica*, 60, 533–565.

Honoré, B. E., E. Kyriazidou, and C. Udry (1997): "Estimation of Type 3 Tobit Models using Symmetric Rrimming and Pairwise Comparisons," *Journal of Econometrics*, 76, 107–128.

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

——— (2005): "Pairwise Difference Estimators for Nonlinear Models," in *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, ed. by D. W. K. Andrews and J. H. Stock, Cambridge University Press, 520–553.

Jochmans, K. (2013): "Pairwise-Comparison Estimation with Non-parametric Controls," *Econometrics Journal*, 16, 340–372.

Kyriazidou, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65, 1335–1364.

Matsushita, Y. and T. Otsu (2021): "Jackknife Empirical Likelihood: Small Bandwidth, Sparse Network and High-Dimensional Asymptotics," *Biometrika*, 108, 661–674.

Pollard, D. (1991): "Asymptotics for Least Absolute Deviation Regression Estimators," *Econometric Theory*, 7, 186–199.

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

Powell, J. L. (1994): "Estimation of Semiparametric Models," in *Handbook of Econometrics, Volume IV*, ed. by R. F. Engle and D. L. McFadden, Elsevier, 2443–2521.

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

Schucany, W. R. and J. P. Sommers (1977): "Improvement of Kernel Type Density Estimators," *Journal of the American Statistical Association*, 72, 420–423.

Shao, J. and D. Tu (2012): *The Jackknife and Bootstrap*, Springer.

van der Vaart, A. W. (1998): *Asymptotic Statistics*, Cambridge University Press.