Robust Inference for Convex Pairwise Difference Estimators*

Matias D. Cattaneo[†] Michael Jansson[‡] Kenichi Nagasawa[§]

July 2, 2025

Abstract

This paper develops new distributional theory and bootstrap-based inference methods for a broad class of convex pairwise difference estimators. These estimators minimize a kernel-weighted objective function over observation pairs that are similar in terms of 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 distributional properties of the following *convex* pairwise difference estimator:

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

where $\Theta \subseteq \mathbb{R}^k$ is the parameter space, $\boldsymbol{\theta} \mapsto m(\mathbf{z}_i, \mathbf{z}_j; \boldsymbol{\theta})$ is a convex objective 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 widely used to address heterogeneity in nonlinear models. See Powell (1994), Honoré and Powell (2005), and Aradillas-Lopez et al. (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 arises from 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

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

where d is the dimension of $\mathbf{w}$, it is well known that the pairwise difference estimator is asymptotically linear, admitting a representation of the form

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} 2\boldsymbol{\xi}(\mathbf{z}_i) + o_{\mathbb{P}}(1) \rightsquigarrow \mathsf{N}(\mathbf{0}, \mathbb{V}[2\boldsymbol{\xi}(\mathbf{z}_i)]), \tag{1.1}$$

where $\boldsymbol{\theta}_0$ is the population parameter of interest and $2\boldsymbol{\xi}(\cdot)$ is the influence function. (The exact form of $\boldsymbol{\xi}$ is given below.) Here, 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 must take negative values.

This paper makes three main contributions:

1. *Small bandwidths.* Building on the idea of *small bandwidth asymptotics* introduced by Cattaneo et al. (2014a), we establish a more robust Gaussian distributional approximation for the pairwise difference estimator that allows for higher levels of localization: It remains 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.* Building on the idea of *generalized jackknifing* introduced by Schucany and Sommers (1977), and following Honoré and Powell (2005), we debias the pairwise difference estimator, thereby allowing 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 crucial 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. *Bootstrap Inference.* Building on insights in Cattaneo et al. (2014b), we develop valid bootstrap-based distributional approximation for the debiased pairwise difference estimator. 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, which is robust to small bandwidths.

In combination, our three contributions therefore 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 helps establishing our theoretical developments under weaker technical conditions. 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), Ahn et al. (2018), Aradillas-Lopez (2012), Blundell and Powell (2004), Hong and Shum (2010), Honoré (1992), Honoré et al. (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 et al. (2010), Cattaneo et al. (2014b), Cattaneo et al. (2018), Cattaneo and Jansson (2018), Matsushita and Otsu (2021), Cattaneo and Jansson (2022), Cattaneo et al. (2025a), and references therein. The generalized jackknife has been successfully used for debiasing in density weighted average derivative estimation (Powell et al., 1989), asymptotically linear pairwise difference estimation (Honoré and Powell, 2005), nonlinear semiparametric estimation (Cattaneo et al., 2013), monotone estimation (Cattaneo et al., 2024), and random forest estimation (Cattaneo et al., 2025c), 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 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. 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. The appendix includes the proofs of our theoretical results and other technical details.

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, while the other two examples do not. The second example has a smooth objective function, while the third example does not. All three examples have convex objective functions 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

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

where $g : \mathbb{R}^d \to \mathbb{R}$ is an unknown function, treated as a nuisance parameter, and the parameter of interest is $\boldsymbol{\theta}_0$. A pairwise difference estimator of $\boldsymbol{\theta}_0$ can be based on the objective function

$$m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = m_{\mathrm{PL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = \frac{1}{2}(y_1 - y_2 - (\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta})^2.$$

The objective function is convex in $\boldsymbol{\theta}$, and setting $\Theta = \mathbb{R}^k$ the pairwise estimator admits a closed form solution:

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

2.2 Partially Linear Logit Model

The partially linear logit model studies here is of the form

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

where g is an unknown nuisance function, $\boldsymbol{\theta}_0$ is the parameter of interest, and where

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

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

$$m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = m_{\mathtt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = -\mathbb{1}\{y_1 \neq y_2\}[y_2 \ln \Lambda(\mathbf{x}_2'\boldsymbol{\theta} - \mathbf{x}_1'\boldsymbol{\theta}) + y_1 \ln \Lambda(\mathbf{x}_1'\boldsymbol{\theta} - \mathbf{x}_2'\boldsymbol{\theta})].$$

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

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

where g is an unknown nuisance function, $\boldsymbol{\theta}_0$ is the parameter of interest, $\mathbf{x}_i \perp\!\!\!\perp \varepsilon_i | \mathbf{w}_i$, and the conditional distribution $\varepsilon_i | \mathbf{w}_i$ admits a Lebesgue density. The associated pairwise difference estimator of $\boldsymbol{\theta}_0$ employs $\Theta = \mathbb{R}^k$ and the function

$$m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = m_{\mathtt{PLT}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = \tilde{m}_{\mathtt{PLT}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) - \tilde{m}_{\mathtt{PLT}}(\mathbf{z}_1, \mathbf{z}_2; \mathbf{0}),$$

where

$$\tilde{m}_{\mathtt{PLT}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = \begin{cases} |y_1| - ((\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta} + y_2) \operatorname{sgn}(y_1) & \text{if } (\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta} \leq -y_2 \\ |y_1 - y_2 - (\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta}| & \text{if } -y_2 < (\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta} < y_1 \\ |y_2| + ((\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta} - y_1) \operatorname{sgn}(y_2) & \text{if } y_1 \leq (\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta} \end{cases}.$$

Because $\tilde{m}_{\mathtt{PLT}}(\mathbf{z}_1, \mathbf{z}_2; \mathbf{0})$ does not depend on $\boldsymbol{\theta}$, the presence of $\tilde{m}_{\mathtt{PLT}}(\mathbf{z}_1, \mathbf{z}_2; \mathbf{0})$ in $m_{\mathtt{PLT}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})$ 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

$$m_{\texttt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}) = \begin{cases} |y_1 - y_2 - (\mathbf{x}_1-\mathbf{x}_2)'\boldsymbol{\theta}| - |y_1 - y_2| & \text{if } y_1>0, y_2>0 \\ \max\{y_1 - (\mathbf{x}_1-\mathbf{x}_2)'\boldsymbol{\theta}, 0\} - \max\{y_1, 0\} & \text{if } y_1>0, y_2=0 \\ \max\{y_2 + (\mathbf{x}_1-\mathbf{x}_2)'\boldsymbol{\theta}, 0\} - \max\{y_2, 0\} & \text{if } y_1=0, y_2>0 \\ 0 & \text{if } y_1=0, y_2=0 \end{cases}.$$

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 the kernel function is non-negative.

3 Distributional Approximation and Bootstrap Inference

As it is standard in the literature, we define our estimator $\widehat{\boldsymbol{\theta}}_n = \widehat{\boldsymbol{\theta}}_n(h_n)$ to be a sequence of measurable random variables satisfying

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

where $\widehat{M}_n$ is the objective function discussed in the introduction:

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

Assuming basic regularity conditions, the objective function $\widehat{M}_n$ is a sample counterpart of

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

and this function $M(\cdot;h)$ approximates, as $h \downarrow 0$, the function

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

where $f_{\mathbf{w}}$ is the Lebesgue density of $\mathbf{w}$. This function M_0 plays an important role in pairwise difference estimation problems because the parameter of interest $\boldsymbol{\theta}_0$ can often be characterized as a unique solution to the minimization problem $\min_{\boldsymbol{\theta}\in\Theta} M_0(\boldsymbol{\theta})$.

Introducing the functions $M(\boldsymbol{\theta};h)$ and $M_0(\boldsymbol{\theta})$ is useful for our distributional approximation theory because we can decompose $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0$ into the non-stochastic "bias" component and the "centered" stochastic component. Specifically, we define a non-random (fixed-h) "pseudo" parameter

$$\boldsymbol{\theta}_n = \boldsymbol{\theta}(h_n), \qquad \boldsymbol{\theta}(h) \in \operatorname*{arg\,min}_{\boldsymbol{\theta}\in\Theta} M(\boldsymbol{\theta};h).$$

Then, we decompose $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0$ into $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n$ ("mean-zero" stochastic term) and $\boldsymbol{\theta}_n - \boldsymbol{\theta}_0$ (non-random "bias" term). Using this decomposition, below we discuss the role of generalize jackknifing to achieve higher-order bias reduction without affecting the convexity of the objective function defining the estimator, nor the generalized distributional approximation based on small bandwidth asymptotics.

First, we provide a set of sufficient conditions which guarantees, among other things, $M(\boldsymbol{\theta}; h)$ converges to $M_0(\boldsymbol{\theta})$ as $h \downarrow 0$, which is crucial for the bias term to vanish asymptotically.

Assumption 1. For $\epsilon > 0$, define $\Theta_0^\epsilon = \{\boldsymbol{\theta} \in \Theta : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| < \epsilon\}$.

(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 $\mathbf{w}$ is absolutely continuous with respect to the Lebesgue measure. The density is bounded and continuous on its support.

(iv) For some $\delta > 0$, $\Theta_0^\delta \subset \Theta$. For each $\boldsymbol{\theta} \in \Theta$, $\mathbb{E}[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})] < \infty$,

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

with probability one, and, denoting the support of $\mathbf{w}$ by $\mathcal{W}$,

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

M_0 is uniquely minimized at $\boldsymbol{\theta}_0$ on Θ.

Next, we state regularity conditions under which we analyze asymptotic properties of $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_n$. The function $(\mathbf{z}_1, \mathbf{z}_2) \mapsto \mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) \in \mathbb{R}^k$ can be interpreted as a "derivative" of $\boldsymbol{\theta} \mapsto m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})$ at $\boldsymbol{\theta}$, although we do not require full differentiability: the partially linear Tobit example above has $m(\mathbf{z}_1, \mathbf{z}_2, \boldsymbol{\theta})$ that is not differentiable at points such that $(\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta} = 0$. Similarly, the function $(\mathbf{z}_1, \mathbf{z}_2) \mapsto \mathbf{H}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t}) \in \mathbb{R}^{k \times k}$ can be thought of as the second directional derivative of $\mathbb{E}[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) | \mathbf{w}_1, \mathbf{w}_2]$, where $\mathbf{t} \in \mathbb{R}^k$ is the direction of derivative. As the partially linear Tobit example indicates, the map $\boldsymbol{\theta} \mapsto m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})$ may not be twice directionally differentiable, but the conditional expectation $\mathbb{E}[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) | \mathbf{w}_1, \mathbf{w}_2]$ often enjoys the required smoothness with respect to $\boldsymbol{\theta}$ because the integral can smooth out kinks. Let $C > 0$ denote an absolute constant that may take different values in each case.

Assumption 2. Let $\delta > 0$. For $\boldsymbol{\theta} \in \Theta_0^\delta$, $\mathbf{t} \in \mathbb{R}^k$, and $\tau > 0$ small enough, define $e_1(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}, \mathbf{t}, \tau) = \tau^{-1}\{m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta} + \mathbf{t}\tau) - m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) - \mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})'\mathbf{t}\tau\}$ and $e_2(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t}, \tau) = \tau^{-2}\{\mathbb{E}[m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta} + \mathbf{t}\tau) - m(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) - \mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})'\mathbf{t}\tau | \mathbf{w}_1, \mathbf{w}_2] - \frac{1}{2}\mathbf{t}'\mathbf{H}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t})\mathbf{t}\tau^2\}$.

(i) There exists a real-valued function $b(\mathbf{z})$ such that for $\boldsymbol{\theta} \in \Theta_0^\delta$, $\|\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})\| \leq b(\mathbf{z}_1)b(\mathbf{z}_2)$, $\mathbb{E}[b(\mathbf{z})^4 | \mathbf{w}] f_{\mathbf{w}}(\mathbf{w}) \leq C < \infty$ with probability one, and $\mathbb{E}|b(\mathbf{z})|^4 < \infty$.

(ii) Let $\mathbf{S}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\vartheta},\boldsymbol{\theta}) = \mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\vartheta})\mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})'$. With probability one,

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

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

$$\mathbb{E}\left[\sup_{\tau\in(0,\delta),\boldsymbol{\theta}\in\Theta_0^\delta,\mathbf{w}_2\in\mathcal{W}} \left|\mathbb{E}\left[e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta},\mathbf{t},\tau)\middle|\mathbf{z}_1,\mathbf{w}_2\right]f_{\mathbf{w}}(\mathbf{w}_2)\right|^2\right]<\infty$$
$$\mathbb{E}\left[\sup_{\tau\in(0,\delta),\boldsymbol{\theta}\in\Theta_0^\delta,\mathbf{w}_2\in\mathcal{W}} \mathbb{E}\left[e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta},\mathbf{t},\tau)^2\middle|\mathbf{w}_1,\mathbf{w}_2\right]f_{\mathbf{w}}(\mathbf{w}_2)\right]<\infty$$
$$\mathbb{E}\left[\sup_{\tau\in(0,\delta),\boldsymbol{\theta}\in\Theta_0^\delta,\mathbf{w}_2\in\mathcal{W}} \left|e_2(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t},\tau)\right|f_{\mathbf{w}}(\mathbf{w}_2)\right]<\infty$$
$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta_0^\delta,\mathbf{w}_2\in\mathcal{W}} \|\mathbf{H}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t})\|\, f_{\mathbf{w}}(\mathbf{w}_2)\right]<\infty.$$

(iv) For $\mathbf{t}\in\mathbb{R}^k$, with probability one,

$$\lim_{(\tau,\boldsymbol{\theta},\mathbf{u})\to(0,\boldsymbol{\theta}_0,\mathbf{0})} \mathbb{E}\left[e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta},\mathbf{t},\tau)\middle|\mathbf{z}_1,\mathbf{w}_2=\mathbf{w}_1+\mathbf{u}\right]=0$$
$$\lim_{(\tau,\boldsymbol{\theta},\mathbf{u})\to(0,\boldsymbol{\theta}_0,\mathbf{0})} \mathbb{E}\left[e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta},\mathbf{t},\tau)^2\middle|\mathbf{w}_1,\mathbf{w}_2=\mathbf{w}_1+\mathbf{u}\right]=0$$
$$\lim_{(\tau,\boldsymbol{\theta},\mathbf{u})\to(0,\boldsymbol{\theta}_0,\mathbf{0})} e_2(\mathbf{w}_1,\mathbf{w}_1+\mathbf{u};\boldsymbol{\theta},\mathbf{t},\tau)=0$$
$$\lim_{(\mathbf{u},\tau)\to(\mathbf{0},0)}\sup_{\boldsymbol{\theta}\in\Theta_0^\tau} \|\mathbf{H}(\mathbf{w}_1,\mathbf{w}_1+\mathbf{u};\boldsymbol{\theta},\mathbf{t})-\mathbf{H}(\mathbf{w}_1,\mathbf{w}_1;\boldsymbol{\theta},\mathbf{t})\|=0$$
$$\lim_{\boldsymbol{\theta}\to\boldsymbol{\theta}_0}\mathbf{H}(\mathbf{w}_1,\mathbf{w}_1;\boldsymbol{\theta},\mathbf{t})=\mathbf{H}(\mathbf{w}_1,\mathbf{w}_1;\boldsymbol{\theta}_0,\mathbf{t})$$

and $\mathbf{H}(\mathbf{w}_1,\mathbf{w}_1;\boldsymbol{\theta}_0,\mathbf{t}) = \mathbf{H}(\mathbf{w}_1,\mathbf{w}_1;\boldsymbol{\theta}_0,\mathbf{s})$ almost surely for any $\mathbf{t},\mathbf{s}\in\mathbb{R}^k$, which allows for dropping the last argument of $\mathbf{H}$ when evaluated at $\mathbf{w}_1=\mathbf{w}_2,\boldsymbol{\theta}=\boldsymbol{\theta}_0$. The matrix

$$\mathbf{H}_0=\int\mathbf{H}(\mathbf{w},\mathbf{w};\boldsymbol{\theta}_0)f_{\mathbf{w}}(\mathbf{w})^2d\mathbf{w}$$

is positive definite.

3.1 Small Bandwidth Asymptotics

To describe the generalized asymptotic distribution of the estimator $\widehat{\boldsymbol{\theta}}_n$, define

$$\mathbf{V}_n=\mathbf{V}_n(h_n),\qquad \mathbf{V}_n(h)=\mathbf{H}_0^{-1}\left[n^{-1}\boldsymbol{\Sigma}_0+\binom{n}{2}^{-1}h^{-d}\boldsymbol{\Delta}_0(K)\right]\mathbf{H}_0^{-1},$$

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

$$\boldsymbol{\Sigma}_0 = 4\mathbb{E}\left[\boldsymbol{\xi}(\mathbf{z})\boldsymbol{\xi}(\mathbf{z})'\right] \quad \text{and} \quad \boldsymbol{\Delta}_0(K) = \mathbb{E}\big[\boldsymbol{\Xi}(\mathbf{w})\big] \int K^2(\mathbf{u}) d\mathbf{u}. \tag{3.1}$$

The notation $\boldsymbol{\Delta}_0(K)$ emphasizes its dependence on the kernel function. The following theorem gives the small bandwidth Gaussian distributional approximation for the canonical pairwise difference estimator. Let $\Phi(\mathbf{t})$ denote the distribution function of a k-dimensional standard Gaussian random vector.

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

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

The convergence rate of $\hat{\boldsymbol{\theta}}_n$ equals the magnitude of $\mathbf{V}_n^{-1/2}$:

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

Provided that $\mathbf{V}_n^{-1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = o(1)$, Theorem 1 encompasses the following three distinct large-sample Gaussian approximations.

- *Asymptotic Linearity*: $nh_n^d \to \infty$. Then, $\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ admits the linear representation (1.1).
- *Root-n Consistency*: $nh_n^d \to c \in (0, \infty)$. 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

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

- *Small Bandwidths*: $nh_n^d \to 0$. Then, $\sqrt{n^2 h^d}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ converges weekly to a mean-zero Gaussian distribution with asymptotic variance $2\boldsymbol{\Delta}_0(K)$.

Whenever the localization parameter is chosen so that $nh_n^d \not\to \infty$, the variance in the Gaussian approximation includes the small bandwidth component $\binom{n}{2}^{-1} h_n^{-d}\boldsymbol{\Delta}_0(K)$, capturing the additional uncertainty generated from increasing the localization of the observations pairs. Therefore, Theorem 1 gives a refined Gaussian distributional approximation for $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0$, removing the condition $nh_n^d \to \infty$, and enlarging the range of localization parameters by imposing the weaker condition $n^2 h_n^d \to 0$, provided the bias condition $\mathbf{V}_n^{-1/2}(\boldsymbol{\theta}_n - \boldsymbol{\theta}_0) = o(1)$ holds. As demonstrated by Cattaneo et al. (2025a), the first-order small bandwidth asymptotic theory can lead to a higher-order corrected distributional approximation.

3.2 Debiasing via Generalized Jackknifing

In Theorem 1, we centered the statistic at $\boldsymbol{\theta}_n = \boldsymbol{\theta}(h_n)$ to circumvent smoothing bias issues. This section focuses on the bias term, and introduces an automatic debiasing approach under high-level conditions. Assumption 1 implies that $\lim_{h\downarrow 0} \boldsymbol{\theta}(h) = \boldsymbol{\theta}_0$, but with additional mild conditions it is possible to show that the bias term is $O(h^2)$. We employ standard multi-index notation: $|\boldsymbol{\alpha}| = \sum_{j=1}^d \alpha_j$ for $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)' \in \mathbb{Z}_+^d$ and $\partial_{\mathbf{v}}^{\boldsymbol{\alpha}} f(\mathbf{w}, \mathbf{v}) = \frac{\partial^{|\boldsymbol{\alpha}|}}{\partial v_1^{\alpha_1} \cdots \partial v_d^{\alpha_d}} f(\mathbf{w}, \mathbf{v})$ for $\mathbf{v} = (v_1, \dots, v_d)' \in \mathbb{R}^d$.

Proposition 1. *Suppose Assumptions 1 and 2, and the following conditions hold.*

(i) $\int \|\mathbf{u}\|^2 K(\mathbf{u}) < \infty$.

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

Then, there exists a non-random vector $\mathbf{b}_2 \in \mathbb{R}^k$ *such that*

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

The bias expansion in Proposition 1 can be extended to a higher-order bias expansion i.e., characterizing leading bias terms up to $O(h_n^L)$ with $L > 2$ under appropriate smoothness conditions. We illustrate this in Section 4.5, where we establish a bias expansion of local M-estimators for $L = 4$ under primitive conditions. This result explicitly leverages convexity of the objective function, and thus appears to be new to the literature. Formalizing the bias expansion for an arbitrary order of $L > 4$ is cumbersome, so we instead impose the following high-level condition (cf., Honoré and Powell, 2005, Equation (3.6)).

Assumption 3. For $\mathbf{b}_0 = 0$ and some even $L \geq 0$, there exist non-random vectors $\mathbf{b}_l \in \mathbb{R}^k$, $l = 1, \dots, L$, such that

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

This assumption encompasses the setting under Assumptions 1 and 2 ($L = 0$), as well as the setting under the additional conditions imposed in Proposition 1 ($L = 2$). Assumption 3 sets the terms with odd powers of h equal to zero, which holds whenever a symmetric kernel function and appropriate smoothness conditions are imposed. See Section 4.5 for more discussion.

To describe the debiasing procedure via generalized jackknifing, let $c_0 = 1$ and $\mathbf{c} = (c_0, \dots, c_{L/2})'$ be distinct positive constants such that the $(L/2+1) \times (L/2+1)$ matrix

$$\mathbf{C} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & c_1^2 & \dots & c_{L/2}^2 \\ \vdots & & \ddots & \\ 1 & c_1^L & \dots & c_{L/2}^L \end{bmatrix}$$

is invertible, and let $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \ldots, \lambda_{L/2})' \in \mathbb{R}^{L/2+1}$ be a vector such that $\mathbf{C}\boldsymbol{\lambda} = (1, 0, \ldots, 0)'$. The debiased estimator is

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

This procedure involves solving $L/2+1$ convex optimization problems: $\widehat{\boldsymbol{\theta}}_n(c_l h)$, $l = 0, \ldots, L/2$. The debiased estimator is a generalization of the original pairwise difference estimator because, if $\mathbf{c} = 1$ (hence $\lambda_0 = 1$) and we take $L = 0$ in Assumption 3, then $\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}} = \widehat{\boldsymbol{\theta}}_n$.

Our next theorem generalizes Theorem 1 by establishing the small bandwidth Gaussian approximation for the debiased pairwise difference estimator $\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}$. Accordingly, let

$$\boldsymbol{\theta}_{n,\mathbf{c}} = \boldsymbol{\theta}_{\mathbf{c}}(h_n), \qquad \boldsymbol{\theta}_{\mathbf{c}}(h) = \sum_{l=0}^{L/2} \lambda_l \boldsymbol{\theta}(c_l h)$$

and

$$\mathbf{V}_{n,\mathbf{c}} = \mathbf{V}_{n,\mathbf{c}}(h_n), \qquad \mathbf{V}_{n,\mathbf{c}}(h) = \mathbf{H}_0^{-1} \left[n^{-1}\boldsymbol{\Sigma}_0 + \binom{n}{2}^{-1} h^{-d} \boldsymbol{\Delta}_0(K_{\mathbf{c}}) \right] \mathbf{H}_0^{-1},$$

where

$$K_l(\mathbf{u}) = c_l^{-d} K\left(\frac{\mathbf{u}}{c_l}\right), \qquad K_{\mathbf{c}}(\mathbf{u}) = \sum_{l=0}^{L/2} \lambda_l K_l(\mathbf{u}), \qquad K_{\mathbf{c},h}(\mathbf{u}) = h^{-d} K_{\mathbf{c}}\left(\frac{\mathbf{u}}{h}\right).$$

It follows that $\boldsymbol{\theta}_{\mathbf{c}}(h)$ and $\mathbf{V}_{n,\mathbf{c}}(h)$ are a generalization of $\boldsymbol{\theta}(h)$ and $\mathbf{V}_n(h)$, respectively, because $\boldsymbol{\theta}(h) = \boldsymbol{\theta}_{\mathbf{c}}(h)$ and $\mathbf{V}_n(h) = \mathbf{V}_{n,\mathbf{c}}(h)$ if $\mathbf{c} = 1$. Debiasing via generalized jackknifing affects the variance formula 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 $h_n \to 0$, then*

$$\sup_{\mathbf{t} \in \mathbb{R}^k} \left| \mathbb{P}\left[\mathbf{V}_{n,\mathbf{c}}^{-1/2} (\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}} - \boldsymbol{\theta}_{n,\mathbf{c}}) \leq \mathbf{t} \right] - \Phi(\mathbf{t}) \right| \to 0$$

If, in addition, Assumption 3 holds and $n h_n^{2L} \to 0$, then

$$\sup_{\mathbf{t} \in \mathbb{R}^k} \left| \mathbb{P}\left[\mathbf{V}_{n,\mathbf{c}}^{-1/2} (\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}} - \boldsymbol{\theta}_0) \leq \mathbf{t} \right] - \Phi(\mathbf{t}) \right| \to 0$$

In addition to establishing a valid small bandwidth Gaussian distributional approximation, this theorem explicitly deals with the smoothing bias term via generalized jackknifing: the debiased estimator is centered at the parameter of interest $\boldsymbol{\theta}_0$ rather than at the pseudo-true parameter sequence $\boldsymbol{\theta}_n$. While from an asymptotic perspective debiasing inflates variance only when $n h_n^d \not\to \infty$, through the term $\boldsymbol{\Delta}_0(K_{\mathbf{c}})$, the Gaussian approximation in Theorem captures a finite sample contribution of the debiasing procedure, and thus can offer a better finite-sample distributional approximation. See Cattaneo et al. (2025a) for more discussion.

It is interesting to note that $K_{\mathbf{c}}$ is a higher-order kernel, even though the construction of the debiased estimator $\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}$ only employs second-order kernels, hence retaining the desired convexity for implementation. More precisely, $\int K_{\mathbf{c}}(\mathbf{u})d\mathbf{u} = \sum_{l=0}^{L/2} \lambda_l \int K_{c_l}(\mathbf{u})d\mathbf{u} = 1$. For $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)' \in \mathbb{Z}_+^d$, let $\mathbf{u}^{\boldsymbol{\alpha}} = \prod_{j=1}^d u_j^{\alpha_j}$ for $\mathbf{u} \in \mathbb{R}^d$. Then, for $0 < |\boldsymbol{\alpha}| \leq L$,

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

where the second equality uses changes-of-variables $\mathbf{v} = c_l^{-1}\mathbf{u}$ and the last equality uses the defining property of $\boldsymbol{\lambda}$ and the symmetry of K.

3.3 Bootstrap Inference

To develop feasible inference procedures, we consider the nonparametric bootstrap approximation of the limiting distribution of the debiased pairwise difference estimator $\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}$. In particular, our results cover the canonical estimator $\widehat{\boldsymbol{\theta}}_n$ by setting $\mathbf{c} = 1$.

Let $\{\mathbf{z}_i^* : i = 1, \ldots, n\}$ be a bootstrap i.i.d. sample drawn from the empirical CDF computed from the original observations $\{\mathbf{z}_i : i = 1, \ldots, n\}$. By the plug-in approach, we construct the bootstrap pairwise difference estimator $\widehat{\boldsymbol{\theta}}_n^* = \widehat{\boldsymbol{\theta}}_n^*(h_n)$ as an approximate minimizer:

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

where

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

is the bootstrap-analogue objective function. Furthermore, the bootstrap (generalized jackknife) debiased estimator is

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

When $\mathbf{c} = 1$, $\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}^* = \widehat{\boldsymbol{\theta}}_n^*$ reduces to the bootstrap-analogue of the pairwise difference estimator $\widehat{\boldsymbol{\theta}}_n$.

The following theorem characterizes the distributional approximation obtained from the nonparametric bootstrap. In perfect analogy with the results in Cattaneo et al. (2014b), we find that the bootstrap distribution estimator consistently estimate the correct limit distribution only when $nh_n^d \to \infty$, but otherwise exhibits a variance inflation making the distributional approximation inconsistent. Let $\mathbb{P}^*[\cdot] = \mathbb{P}[\cdot | \mathbf{z}_1, \ldots, \mathbf{z}_n]$, and $\to_{\mathbb{P}}$ denote convergence in probability.

Theorem 3. *Suppose Assumptions 1 and 2 hold, and that* $m_n(\mathbf{z},\mathbf{z};\boldsymbol{\theta}) = 0$ *and* $\mathbf{s}_n(\mathbf{z},\mathbf{z};\boldsymbol{\theta}) = \mathbf{0}$ *for all* $\mathbf{z}$ *and* $\boldsymbol{\theta} \in \Theta_0^\delta$. *If* $n^2 h_n^d \to \infty$ *and* $h_n \to 0$, *then*

$$\sup_{\mathbf{t}\in\mathbb{R}^k} \left| \mathbb{P}^*\left[\mathbf{V}_{n,\mathbf{c}}(3^{-1/d}h_n)^{-1/2}(\widehat{\boldsymbol{\theta}}^*_{n,\mathbf{c}}(h_n) - \widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(h_n)) \le \mathbf{t}\right] - \Phi(\mathbf{t}) \right| \to_{\mathbb{P}} 0.$$

Letting $\mathbf{I}$ denote the k-dimensional identity matrix, it follows that $\mathbf{V}_{n,\mathbf{c}}(3^{-1/d}h_n)^{-1}\mathbf{V}_{n,\mathbf{c}}(h_n) \to \mathbf{I}$ if and only if $nh_n^d \to \infty$, which implies that

$$\sup_{\mathbf{t}\in\mathbb{R}^k} \left| \mathbb{P}^*\left[\widehat{\boldsymbol{\theta}}^*_{n,\mathbf{c}}(h_n) - \widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(h_n) \le \mathbf{t}\right] - \mathbb{P}\left[\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(h_n) - \boldsymbol{\theta}(h_n) \le \mathbf{t}\right] \right| \to_{\mathbb{P}} 0$$

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 it will be conservative in the sense that the leading variance under the bootstrap distribution is larger than the leading variance of the asymptotic distribution: $\mathbf{V}_{n,\mathbf{c}}(3^{-1/d}h_n) > \mathbf{V}_{n,\mathbf{c}}(h_n)$ in a positive definite sense.

The leading variance inflation generated by the nonparametric bootstrap under the small bandwidth regime, $\liminf_{n\to\infty} nh_n^d < \infty$, can be easily fixed by appropriately rescaling the bandwidth used for the bootstrap implementation of the pairwise estimator: according to Theorem 3, computing $\widehat{\boldsymbol{\theta}}^*_n(3^{1/d}h_n) - \widehat{\boldsymbol{\theta}}_n(3^{1/d}h_n)$, instead of the original plug-in bootstrap statistic employing the bandwidth choice h_n, automatically adjusts the bootstrap variance, and therefore leads to a consistent distributional approximation. The following theorem formalizes this result.

Corollary 1. *Suppose Assumptions 1 and 2 hold, and that* $m_n(\mathbf{z},\mathbf{z};\boldsymbol{\theta}) = 0$ *and* $\mathbf{s}_n(\mathbf{z},\mathbf{z};\boldsymbol{\theta}) = \mathbf{0}$ *for all* $\mathbf{z}$ *and* $\boldsymbol{\theta} \in \Theta_0^\delta$. *If* $n^2 h_n^d \to \infty$ *and* $h_n \to 0$, *then*

$$\sup_{\mathbf{t}\in\mathbb{R}^k} \left| \mathbb{P}^*\left[\widehat{\boldsymbol{\theta}}^*_{n,\mathbf{c}}(3^{1/d}h_n) - \widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(3^{1/d}h_n) \le \mathbf{t}\right] - \mathbb{P}\left[\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(h_n) - \boldsymbol{\theta}(h_n) \le \mathbf{t}\right] \right| \to_{\mathbb{P}} 0.$$

If, in addition, Assumption 3 holds and $nh_n^{2L} \to 0$, *then*

$$\sup_{\mathbf{t}\in\mathbb{R}^k} \left| \mathbb{P}^*\left[\widehat{\boldsymbol{\theta}}^*_{n,\mathbf{c}}(3^{1/d}h_n) - \widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(3^{1/d}h_n) \le \mathbf{t}\right] - \mathbb{P}\left[\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(h_n) - \boldsymbol{\theta}_0 \le \mathbf{t}\right] \right| \to_{\mathbb{P}} 0.$$

Corollary 1 emphasizes the rate-adaptive nature of the consistency property enjoyed by the bootstrap distributional approximation. This result has immediate implications for robust inference. For example, letting $\alpha \in (0,1)$ and $\mathbf{a} \in \mathbb{R}^k$ 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

$$\mathsf{CI}^*_{1-\alpha,n} = \left[\mathbf{a}'\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(h_n) - q^*_{1-\alpha/2,n}\ ,\ \mathbf{a}'\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(h_n) - q^*_{\alpha/2,n}\right]$$

with

$$q^*_{t,n} = \inf\left\{q \in \mathbb{R} : \mathbb{P}^*[\mathbf{a}'\widehat{\boldsymbol{\theta}}^*_{n,\mathbf{c}}(3^{1/d}h_n) - \mathbf{a}'\widehat{\boldsymbol{\theta}}_{n,\mathbf{c}}(3^{1/d}h_n) \le q] \ge t\right\},$$

which satisfies

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

under the substantially weaker conditions $n^2 h_n^d \to \infty$ and $nh_n^{2L} \to 0$.

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

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

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

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

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,

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

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

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 $(\sqrt{n}\widehat{\mathbf{L}}_n' \; \sqrt{n^2 h_n^d}\widehat{\mathbf{W}}_n')\boldsymbol{\mu} = \sum_{i=1}^n g_{in}(\boldsymbol{\mu})$. 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,

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

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

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

where the inequality uses Lemma 6. Then,

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

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

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

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

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

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

and

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

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

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

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

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

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

and Lemma 5.

For (4.7),

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

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 $\widehat{M}_{n,l}^*(\boldsymbol{\theta})=\widehat{M}_n^*(\boldsymbol{\theta};c_lh_n)$ 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,

$$
\widehat{M}_{n,l}^*(\boldsymbol{\theta}_{n,l}+\mathbf{t}r_n^{-1})-\widehat{M}_{n,l}^*(\boldsymbol{\theta}_{n,l})=r_n^{-1}\binom{n}{2}^{-1}\sum_{i<j}e_{n,l}(\mathbf{z}_i^*,\mathbf{z}_j^*;\mathbf{t})+r_n^{-1}\mathbf{t}'\binom{n}{2}^{-1}\sum_{i<j}\mathbf{s}_{n,l}(\mathbf{z}_i^*,\mathbf{z}_j^*)
$$

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

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

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

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

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

The above display and Hoeffding decomposition imply

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

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 $\{\mathbf{z}_1^*,\ldots,\mathbf{z}_i^*\}$, and for $\boldsymbol{\mu} = (\boldsymbol{\mu}_1'\ \boldsymbol{\mu}_2')' \in \mathbb{R}^{2k}$,

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

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

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

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

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

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 $\left(\frac{\pi_0}{1-\pi_0}\right)^2$. Then,

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

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

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

$$\rightsquigarrow_{\mathbb{P}} -\mathbf{H}_0^{-1}\left[\sqrt{2}\mathbf{W}+\mathbf{W}_2\right] =_d -\mathbf{H}_0^{-1}\left[\sqrt{3}\mathbf{W}\right].$$

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

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

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

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

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,

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

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

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

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 $l, \tilde{l} \in \{0, \dots, L/2\}$,

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

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 $(x+y)^4 \leq 8(x^4 + y^4)$ for $x, y \in \mathbb{R}$,

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

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

$$\begin{aligned}
\mathbb{E}^* |\boldsymbol{\mu}_1' \widehat{\boldsymbol{\ell}}_n(\mathbf{z}^*)|^4 &\leq C \mathbb{E}^* \left[\mathbb{E}^* [\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_1^*, \mathbf{z}_2^*) | \mathbf{z}_1^*]^4 \right] \\
&\leq C \frac{1}{n} \sum_{i=1}^n \left| \frac{1}{n} \sum_{j \neq i} \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_i, \mathbf{z}_j) \right|^4 \\
&\leq C \frac{1}{n^5} \sum_{i=1}^n \left| \sum_{j \neq i} |\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_i, \mathbf{z}_j)|^2 + \sum_{j \neq i} \sum_{p \neq i,j} \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_i, \mathbf{z}_j) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_i, \mathbf{z}_p) \right|^2 \\
&\leq C \frac{1}{n^5} \sum_{i=1}^n \left| \sum_{j \neq i} |\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_i, \mathbf{z}_j)|^2 \right|^2 + C \frac{1}{n^5} \sum_{i=1}^n \left| \sum_{j \neq i} \sum_{p \neq i,j} \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_i, \mathbf{z}_j) \boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_i, \mathbf{z}_p) \right|^2 \\
&= C \frac{1}{n^5} \sum_{i=1}^n \sum_{j \neq i} |\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_i, \mathbf{z}_j)|^4 + C \frac{1}{n^5} \sum_{i=1}^n \sum_{j \neq i} \sum_{p \neq i,j} |\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_i, \mathbf{z}_j)|^2 |\boldsymbol{\mu}_1' \mathbf{s}_n^L(\mathbf{z}_i, \mathbf{z}_p)|^2
\end{aligned}$$

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

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 $(nh_n^d)^2/(1+\sqrt{nh_n^d})^4 \leq 1$.

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 $j\neq p$,

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

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)

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

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

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

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

and

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

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,

$$
\mathbb{E}^*J_2^2\leq Ch_n^{2d}\mathbb{E}^*\left[\mathbb{E}^*\left[\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_1^*,\mathbf{z}_2^*)\boldsymbol{\mu}_2'\widehat{\boldsymbol{\omega}}_n(\mathbf{z}_1^*,\mathbf{z}_3^*)\middle|\mathbf{z}_2^*,\mathbf{z}_3^*\right]^2\right]
$$

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

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

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

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

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

$$+\frac{2}{n^5}\sum_{i,j,p,r}\{s_{2,ij}^2 s_{1,jp}s_{1,jr}+s_{2,ij}s_{1,jp}s_{2,ip}s_{1,pr}\}+\frac{2}{n^5}\sum_{i,j,p}\{s_{2,ij}^2 s_{1,jp}^2+s_{2,ij}s_{1,jp}^2 s_{2,ip}\}$$
$$+\frac{2}{n^5}\sum_{i,j,p}s_{1,ij}s_{2,ij}s_{1,ip}s_{2,ip}+\frac{2}{n^5}\sum_{i,j}(s_{1,ij}s_{2,ij})^2$$

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}}\bigg(\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)| \\
&\quad + 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\} \\
&\quad + 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)| \\
&\quad + 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\bigg)
\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,

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

$$\Rightarrow \quad \eta/2\lambda \leq M(\boldsymbol{\theta}_1; h) - M(\boldsymbol{\theta}_0; h)$$

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

The above argument also shows that for each $\epsilon \in (0, \delta)$, there exists $h_\epsilon > 0$ such that $\|\boldsymbol{\theta}(h) - \boldsymbol{\theta}_0\| \leq \epsilon$ 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 $\mathbf{t} \in \mathbb{R}^k$,

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

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,

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

Again by the dominated convergence theorem,

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

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 $M(\boldsymbol{\theta}; h)$. $\square$

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,

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

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. $\square$

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$\leq Cb(\mathbf{z}_2)\int \bar{K}(\mathbf{v})\, d\mathbf{v}$$

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. $\square$

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

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

$\square$

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 $l = 0,\ldots,L/2$,*

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

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,

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

The desired result follows from combining the two stochastic orders. □

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 $l=0,\ldots,L/2$ (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, $\mathbb{E}[\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)\||\mathbf{z}_1]\leq Cb(\mathbf{z}_1)$. Then, $\|\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_3)\mathbb{E}[\mathbf{s}_n^L(\mathbf{z}_1,\mathbf{z}_2)|\mathbf{z}_1]'\|\leq Cb(\mathbf{z}_1)^2b(\mathbf{z}_3)\mathcal{K}_{h_n}(\mathbf{w}_1-\mathbf{w}_3)$ and

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

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

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,

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

Then,

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

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

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

Then, $\mathbb{E}[b(\mathbf{z}_1)^2b(\mathbf{z}_2)^2\mathcal{K}_{h_n}(\mathbf{w}_1-\mathbf{w}_2)^2|\mathbf{z}_1]\leq Ch_n^{-d}b(\mathbf{z}_1)^2$ 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

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

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

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

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

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

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

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

and

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

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

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

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

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. $\square$

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,

$$\frac{1}{n(n-1)(n-2)}\sum_{i=1}^n\sum_{j=1,j\neq i}^n\sum_{p=1,p\neq j,i}^n\boldsymbol{\mu}_1'\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_j)\right)\left(\sum_{l=0}^{L/2}\lambda_l\mathbf{s}_{n,l}(\mathbf{z}_i,\mathbf{z}_p)\right)'\boldsymbol{\mu}_1$$

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

$$+ \frac{1}{n}\sum_{i=1}^{n}3\left(\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i]-\bar{\zeta}_1\right)+\binom{n}{2}^{-1}\sum_{i<j}3\left(\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i,\mathbf{z}_j]-\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i]-\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_j]+\bar{\zeta}_1\right)$$

$$+ \binom{n}{3}^{-1}\sum_{i<j<p}\big(\zeta_{1,ijp}-\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i,\mathbf{z}_j]-\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i,\mathbf{z}_p]-\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_j,\mathbf{z}_p]+\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_i]$$

$$+\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_j]+\mathbb{E}[\zeta_{1,ijp}|\mathbf{z}_p]-\bar{\zeta}_1\big)$$

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

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

Lemma 8 implies

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

Lemma 8 implies

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

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 $\mathbb{E}[(\boldsymbol{\mu}_2'\mathbf{s}_n^L(\mathbf{z}_i,\mathbf{z}_j))^4] = o(n^2 h_n^{-d})$ because $n^2 h_n^d \to \infty$. $\square$

4.4 Proof of Proposition 1

Since

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

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 $|\boldsymbol{\alpha}| \leq L$. 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

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

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. $\square$

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 for

$$e_3(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t}, \tau) = \frac{H_{ij}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta} + \mathbf{t}\tau, \mathbf{t}) - H_{ij}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t}) - \dot{\mathbf{h}}_{ij}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t})'\mathbf{t}\tau}{\tau},$$

we have $\lim_{(\mathbf{u}, \boldsymbol{\theta}, \tau)\to(\mathbf{0}, \boldsymbol{\theta}_0, 0)} e_3(\mathbf{w}_1, \mathbf{w}_1 + \mathbf{u}; \boldsymbol{\theta}, \mathbf{t}, \tau) = 0$ with probability one and

$$\mathbb{E} \sup_{\mathbf{v}\in\mathcal{W}, \boldsymbol{\theta}\in\Theta_0^\delta, \tau\in(0,\delta)} \big|e_3(\mathbf{w}_1, \mathbf{v}; \boldsymbol{\theta}, \mathbf{t}, \tau)\big| < \infty.$$

$\mathbf{v} \mapsto \boldsymbol{\psi}(\mathbf{w}, \mathbf{w} + \mathbf{v})$ 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

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

for $|\boldsymbol{\alpha}| \le 2$.

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,

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

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

Lemma 10 and (4.12),

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

Then,

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

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

$$h^{-8}\big\{M(\boldsymbol{\theta}_0 + \mathbf{b}_2 h^2 + \mathbf{t} h^4; h) - M(\boldsymbol{\theta}_0 + \mathbf{b}_2 h^2; h)\big\} - h^{-4}\partial M(\boldsymbol{\theta}_0 + \mathbf{b}_2 h^2; h)'\mathbf{t} = \frac{1}{2}\mathbf{t}'\mathbf{H}_0\mathbf{t} + o(1).$$

Since $h^{-4}\partial M(\boldsymbol{\theta}_0 + \mathbf{b}_2 h^2; h)'\mathbf{t} + \frac{1}{2}\mathbf{t}'\mathbf{H}_0\mathbf{t} = Q(t) + o(1)$ 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

$$h^{-4}(\boldsymbol{\theta}(h) - \boldsymbol{\theta}_0 - \mathbf{b}_2 h^2) + \mathbf{H}_0^{-1}\big(\mathbf{b}_4^M + \mathfrak{b}_{4,1} + \mathfrak{b}_{4,2}\big) = o(1)$$

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

We provide primitive sufficient conditions for each example in Section 2 to verify that Assumptions 1(iv) and 2 hold. Recall that Assumption 1(ii) holds in each of the examples.

5.1 Partially Linear Regression Model

In this example, the objective function $m_{\text{PL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})$ is twice-differentiable in $\boldsymbol{\theta}$ and we can take

$$\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = \mathbf{s}_{\text{PL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = -\big(y_1 - y_2 - (\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta}\big)(\mathbf{x}_1 - \mathbf{x}_2)$$

and

$$\mathbf{H}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t}) = \mathbf{H}_{\text{PL}}(\mathbf{w}_1, \mathbf{w}_2) = \mathbb{E}[(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)'|\mathbf{w}_1, \mathbf{w}_2]$$

where the $\mathbf{H}$ function does not depend on $\mathbf{t}$ and $\boldsymbol{\theta}$.

To verify our assumptions, we impose the following conditions.

Assumption 5. Let $C > 0$ be a finite constant.

(i) $\mathbb{E}[\varepsilon^4(1+\|\mathbf{x}\|^4)+\|\mathbf{x}\|^8)|\mathbf{w}]f_{\mathbf{w}}(\mathbf{w})+g(\mathbf{w})^4(\mathbb{E}[\|\mathbf{x}\|^4|\mathbf{w}]+1)f_{\mathbf{w}}(\mathbf{w})\leq C$ with probability one. Also, $\mathbb{E}[(\varepsilon^4+g(\mathbf{w})^4)(1+\|\mathbf{x}\|^4)+\|\mathbf{x}\|^8]<\infty$.

(ii) $\mathbf{H}_0=\int\mathbb{E}[(\mathbf{x}_1-\mathbf{x}_2)(\mathbf{x}_1-\mathbf{x}_2)'|\mathbf{w}_1=\mathbf{w},\mathbf{w}_2=\mathbf{w}]f_{\mathbf{w}}(\mathbf{w})^2d\mathbf{w}$ is positive definite.

(iii) With probability one, $\mathbb{E}[\varepsilon|\mathbf{x},\mathbf{w}]=0$.

(iv) The functions $\mathbf{w}\mapsto g(\mathbf{w})$, $\mathbf{w}\mapsto\mathbb{E}[\mathbf{x}|\mathbf{w}]$, $\mathbf{w}\mapsto\mathbb{E}[\mathbf{x}\mathbf{x}'|\mathbf{w}]$, $\mathbf{w}\mapsto\mathbb{E}[\varepsilon^2|\mathbf{w}]$, $\mathbf{w}\mapsto\mathbb{E}[\varepsilon^2\mathbf{x}|\mathbf{w}]$, $\mathbf{w}\mapsto\mathbb{E}[\varepsilon^2\mathbf{x}\mathbf{x}'|\mathbf{w}]$ are continuous at almost every $\mathbf{w}$.

Proposition 3. *Under Assumptions 1(i)(iii) and 5, Assumptions 1(iv) and 2 hold for the partially linear regression example.*

Proof.

Assumption 1(iv) Note that

$$|m_{\mathtt{PL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|\leq(\|\mathbf{x}_1\|^2+\|\mathbf{x}_2\|^2)\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^2+g(\mathbf{w}_1)^2+g(\mathbf{w}_2)^2+\varepsilon_1^2+\varepsilon_2^2$$

and by the hypothesis, $\mathbb{E}[|m_{\mathtt{PL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|]<\infty$ for each $\boldsymbol{\theta}$.

For each $\boldsymbol{\theta}\in\Theta$,

$$\begin{aligned}&\mathbb{E}[m_{\mathtt{PL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|\mathbf{w}_1,\mathbf{w}_2]\\&=\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\mathbb{E}[(\mathbf{x}_1-\mathbf{x}_2)(\mathbf{x}_1-\mathbf{x}_2)'|\mathbf{w}_1,\mathbf{w}_2](\boldsymbol{\theta}-\boldsymbol{\theta}_0)+\frac{1}{2}|g(\mathbf{w}_1)-g(\mathbf{w}_2)|^2\\&\quad+\frac{1}{2}\big(\mathbb{E}[\varepsilon_1^2|\mathbf{w}_1]+\mathbb{E}[\varepsilon_2^2|\mathbf{w}_2]\big)-(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\mathbb{E}[\mathbf{x}_1-\mathbf{x}_2|\mathbf{w}_1,\mathbf{w}_2](g(\mathbf{w}_1)-g(\mathbf{w}_2))\end{aligned}$$

and by the hypothesis, $\mathbf{w}_2\mapsto\mathbb{E}[m_{\mathtt{PL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|\mathbf{w}_1,\mathbf{w}_2]$ is continuous with probability one. Also,

$$\begin{aligned}&\sup_{\mathbf{w}\in\mathcal{W}}\big|\mathbb{E}[m_{\mathtt{PL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|\mathbf{w}_1,\mathbf{w}_2=\mathbf{w}]\big|f_{\mathbf{w}}(\mathbf{w})\\&\leq C\Big(\mathbb{E}[\|\mathbf{x}_1\|^2+\|\mathbf{x}_1\||\mathbf{w}_1]\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|^2+g(\mathbf{w}_1)^2+\mathbb{E}[\varepsilon_1^2|\mathbf{w}_1]+\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|(\mathbb{E}[\|\mathbf{x}_1\||\mathbf{w}_1]+1)(g(\mathbf{w}_1)+1)\Big)\end{aligned}$$

and the dominating function has a finite expectation for each $\boldsymbol{\theta}$.

M_0 function takes the form

$$M_0(\boldsymbol{\theta})=\frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\mathbf{H}_0(\boldsymbol{\theta}-\boldsymbol{\theta}_0)+\int\mathbb{E}[\varepsilon^2|\mathbf{w}]f_{\mathbf{w}}(\mathbf{w})^2d\mathbf{w}$$

which is uniquely minimized at $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ as $\mathbf{H}_0$ is positive definite.

Assumption 2(i) By

$$\|\mathbf{s}_{\mathtt{PL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})\|\leq\{|\varepsilon_1|+|\varepsilon_2|+|g(\mathbf{w}_1)|+g(\mathbf{w}_2)+(\|\mathbf{x}_1\|+\|\mathbf{x}_2\|)\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\}(\|\mathbf{x}_1\|+\|\mathbf{x}_2\|),$$

we can take

$$b(\mathbf{z}_1) = (1 + |\varepsilon_1| + |g(\mathbf{w}_1)| + \|\mathbf{x}_1\|)(1 + \|\mathbf{x}_1\|)$$

for some $\delta \in (0,1)$. By hypothesis, $\mathbb{E}[b(\mathbf{z})^4|\mathbf{w}]f_{\mathbf{w}}(\mathbf{w})$ is bounded and $\mathbb{E}[b(\mathbf{z})^4] < \infty$.

Assumption 2(ii) For $\boldsymbol{\theta} \in \Theta$,

$$|\mathbf{s}_{\mathrm{PL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) - \mathbf{s}_{\mathrm{PL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)| \le (\|\mathbf{x}_1\| + \|\mathbf{x}_2\|)^2 \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le 2(\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2)\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$$

and with probability one, as $(\boldsymbol{\theta}, \mathbf{u}) \to (\boldsymbol{\theta}_0, \mathbf{0})$,

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

Also, under the maintained hypothesis,

$$\begin{aligned}\mathbb{E}[\mathbf{s}_{\mathrm{PL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)|\mathbf{z}_1, \mathbf{w}_2 = \mathbf{w}_1 + \mathbf{u}] &= -(g(\mathbf{w}_1) - g(\mathbf{w}_1 + \mathbf{u}) + \varepsilon_1)(\mathbf{x}_1 - \mathbb{E}[\mathbf{x}_2|\mathbf{w}_2 = \mathbf{w}_1 + \mathbf{u}]) \\ &\to -\varepsilon_1(\mathbf{x}_1 - \mathbb{E}[\mathbf{x}_1|\mathbf{w}_1]) = \mathbb{E}[\mathbf{s}_{\mathrm{PL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)|\mathbf{z}_1, \mathbf{w}_2 = \mathbf{w}_1]\end{aligned}$$

as $\mathbf{u} \to \mathbf{0}$. Thus, the first display holds.

Also, the dominated convergence theorem implies

$$\begin{aligned}\mathbb{E}[\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)'|\mathbf{w}_1, \mathbf{w}_2] &= (g(\mathbf{w}_1) - g(\mathbf{w}_2))^2 \mathbb{E}[(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)'|\mathbf{w}_1, \mathbf{w}_2] \\ &\quad + \mathbb{E}[(\varepsilon_1^2 + \varepsilon_2^2)(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)'|\mathbf{w}_1, \mathbf{w}_2]\end{aligned}$$

converges to $\mathbb{E}[\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)'|\mathbf{w}_1, \mathbf{w}_2 = \mathbf{w}_1]$ as $\mathbf{u} \to \mathbf{0}$. Combined with this observation, the bound

$$\begin{aligned}&\|\mathbf{s}_{\mathrm{PL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\vartheta}) - \mathbf{s}_{\mathrm{PL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)\|\|\mathbf{s}_{\mathrm{PL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})\| \\ &\le \{(\|\mathbf{x}_1\| + \|\mathbf{x}_2\|)^3(|\varepsilon_1| + |\varepsilon_2| + |g(\mathbf{w}_1)| + |g(\mathbf{w}_2)| + (\|\mathbf{x}_1\| + \|\mathbf{x}_2\|)\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|\}\|\boldsymbol{\vartheta} - \boldsymbol{\theta}_0\|\end{aligned}$$

implies that the second display holds.

Assumption 2(iii) We have

$$e_1(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}, \mathbf{t}, \tau) = \frac{1}{2}\tau|\mathbf{t}'(\mathbf{x}_1 - \mathbf{x}_2)|^2$$

and

$$\begin{aligned}&\mathbb{E}[e_1(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}, \mathbf{t}, \tau)|\mathbf{z}_1, \mathbf{w}_2] \le \tau\|\mathbf{t}\|(\|\mathbf{x}_1\|^2 + \mathbb{E}[\|\mathbf{x}_2\|^2|\mathbf{w}_2]) \\ &\mathbb{E}[e_1(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}, \mathbf{t}, \tau)^2|\mathbf{w}_1, \mathbf{w}_2] \le 2\tau^2\|\mathbf{t}\|^2(\mathbb{E}[\|\mathbf{x}_1\|^4|\mathbf{w}_1] + \mathbb{E}[\|\mathbf{x}_2\|^4|\mathbf{w}_2]).\end{aligned} \tag{5.1}$$

Thus, the first and second displays hold. The third display holds because $e_2(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t}, \tau) = 0$ almost surely. Finally, since $\|\mathbf{H}_{\text{PL}}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t})\| \leq 2(\mathbb{E}[\|\mathbf{x}_1\|^2|\mathbf{w}_1] + \mathbb{E}[\|\mathbf{x}_2\|^2|\mathbf{w}_2])$,

$$\mathbb{E} \sup_{\boldsymbol{\theta}\in\Theta_0^\delta} \sup_{\mathbf{w}\in\mathcal{W}} \|\mathbf{H}_{\text{PL}}(\mathbf{w}_1, \mathbf{w}; \boldsymbol{\theta}, \mathbf{t})\| f_{\mathbf{w}}(\mathbf{w}) \leq C(\mathbb{E}[\|\mathbf{x}_1\|^2] + 1) < \infty.$$

Assumption 2(iv) By (5.1), the first two displays hold. The third display holds as $e_2 = 0$ almost surely. The fourth and fifth displays hold as $\mathbf{H}_{\text{PL}}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t})$ does not depend on $\boldsymbol{\theta}$ and is continuous with respect to $\mathbf{w}_2$.

□

5.2 Partially Linear Logistic Model

In this example, the objective function is twice-differentiable in $\boldsymbol{\theta}$ and we have

$$\mathbf{s}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = \mathbf{s}_{\text{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) = -\mathbb{1}\{y_1 \neq y_2\}\big[y_1\Lambda(\mathbf{x}_2'\boldsymbol{\theta} - \mathbf{x}_1'\boldsymbol{\theta}) - y_2\Lambda(\mathbf{x}_1'\boldsymbol{\theta} - \mathbf{x}_2'\boldsymbol{\theta}))\big](\mathbf{x}_1 - \mathbf{x}_2)$$

and

$$\mathbf{H}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}, \mathbf{t}) = \mathbf{H}_{\text{PLL}}(\mathbf{w}_1, \mathbf{w}_2; \boldsymbol{\theta}) = \mathbb{E}\big[\mathbb{1}\{y_1 \neq y_2\}\lambda(\mathbf{x}_1'\boldsymbol{\theta} - \mathbf{x}_2'\boldsymbol{\theta})(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)'\big|\mathbf{w}_1, \mathbf{w}_2\big]$$

where $\lambda(u) = \Lambda(u)(1 - \Lambda(u))$ and the $\mathbf{H}$ function does not depend on $\mathbf{t}$.

Assumption 6. Let $C > 0$ be a finite constant.

(i) $\mathbb{E}[\|\mathbf{x}\|^4|\mathbf{w}]f_{\mathbf{w}}(\mathbf{w}) \leq C$ with probability one. $\mathbb{E}[\|\mathbf{x}\|^4] < \infty$.

(ii) $\mathbf{H}_0 = \int \mathbf{H}_{\text{PLL}}(\mathbf{w}, \mathbf{w}; \boldsymbol{\theta}_0) f_{\mathbf{w}}^2(\mathbf{w})d\mathbf{w}$ is positive definite.

(iii) $\mathbf{w} \mapsto g(\mathbf{w})$ is continuous for almost all $\mathbf{w}$. The conditional distribution of $\mathbf{x}$ given $\mathbf{w}$ has a density with respect to some measure ρ. Denoting the conditional density by $f_{\mathbf{x}|\mathbf{w}}$, $\int(1 + \|\mathbf{x}\|^2)\sup_{\|\mathbf{v}\|\leq\delta} f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w} + \mathbf{v})d\rho(\mathbf{x})$ is finite for almost all $\mathbf{w}$. Also, $\mathbf{w} \mapsto f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w})$ is continuous almost surely.

(iv) $\mathbb{E}[(\mathbf{x}_1 - \mathbf{x}_2)(\mathbf{x}_1 - \mathbf{x}_2)'|\mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w}]$ is invertible for almost all $\mathbf{w}$.

Proposition 4. *Under Assumptions 1(i)(iii) and 6, Assumptions 1(iv) and 2 hold for the partially linear logistic regression example.*

Proof.

Assumption 1(iv) We have $-\ln\Lambda \geq 0$ on $\mathbb{R}$ and by $\ln\Lambda(u) = u - \ln(1 + \exp(u))$,

$$\begin{aligned} 0 &\leq m_{\text{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) \leq -\ln\Lambda(\mathbf{x}_2'\boldsymbol{\theta} - \mathbf{x}_1'\boldsymbol{\theta}) - \ln\Lambda(\mathbf{x}_1'\boldsymbol{\theta} - \mathbf{x}_2'\boldsymbol{\theta}) \\ &= \ln\big(1 + \exp(\mathbf{x}_2'\boldsymbol{\theta} - \mathbf{x}_1'\boldsymbol{\theta})\big) + \ln\big(1 + \exp(\mathbf{x}_1'\boldsymbol{\theta} - \mathbf{x}_2'\boldsymbol{\theta})\big) \leq 2\ln 2 + 2\|\mathbf{x}_1 - \mathbf{x}_2\|\|\boldsymbol{\theta}\| \end{aligned}$$

where the last inequality uses for $u \geq 0$, $\ln(1+\exp(u)) \leq \ln 2 + u$. Thus, $\mathbb{E}[|m_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})|] < \infty$ for each $\boldsymbol{\theta}$.

The above bound and $\mathbb{E}[\|\mathbf{x}\||\mathbf{w}]f_{\mathbf{w}}(\mathbf{w}) \leq C$ imply

$$0 \leq \mathbb{E}[m_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})|\mathbf{w}_1, \mathbf{w}_2 = \mathbf{w}] \leq C(1 + \mathbb{E}[\|\mathbf{x}_1\||\mathbf{w}_1]\|\boldsymbol{\theta}\| + \|\boldsymbol{\theta}\|)$$

which in turn implies the second display.

For the first display,

$$\begin{aligned}
&\mathbb{E}\big[\mathbb{1}\{y_1 \neq y_2\}y_2 \ln \Lambda(\mathbf{x}_2'\boldsymbol{\theta} - \mathbf{x}_1'\boldsymbol{\theta})\big|\mathbf{w}_1, \mathbf{w}_2\big] \\
&= \int \{1 - \Lambda(\mathbf{x}_1'\boldsymbol{\theta}_0 + g(\mathbf{w}_1))\}\Lambda(\mathbf{x}_2'\boldsymbol{\theta}_0 + g(\mathbf{w}_2)) \ln \Lambda(\mathbf{x}_2'\boldsymbol{\theta} - \mathbf{x}_1'\boldsymbol{\theta}) f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_1|\mathbf{w}_1) f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}_2) d\rho(\mathbf{x}_1) d\rho(\mathbf{x}_2)
\end{aligned}$$

where the integrand is continuous in $\mathbf{w}_1, \mathbf{w}_2$ and the integrand times (-1) is non-negative and bounded above by $(\ln 2 + |\mathbf{x}_2'\boldsymbol{\theta} - \mathbf{x}_1'\boldsymbol{\theta}|) \sup_{\|\mathbf{v}\| \leq \delta} f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_1|\mathbf{w}_1 + \mathbf{v}) \sup_{\|\mathbf{v}\| \leq \delta} f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}_2 + \mathbf{v})$ which is integrable by the hypothesis. Then, the dominated convergence theorem implies $\mathbb{E}[m_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})|\mathbf{w}_1, \mathbf{w}_2]$ is continuous in $\mathbf{w}_2$ almost surely.

Finally, to show $\boldsymbol{\theta}_0$ uniquely minimizes M_0, note that for $d \in \{0, 1\}$,

$$\mathbb{P}[y_1 = d, y_2 = 1 - d|y_1 + y_2 = 1, \mathbf{x}_1, \mathbf{x}_2, \mathbf{w}_1, \mathbf{w}_2] = \frac{\exp(d\{(\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta}_0 + g(\mathbf{w}_1) - g(\mathbf{w}_2)\})}{1 + \exp((\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta}_0 + g(\mathbf{w}_1) - g(\mathbf{w}_2))}$$

and $\mathbb{E}[m_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)|\mathbf{x}_1, \mathbf{x}_2, \mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w}]$ equals the conditional log-likelihood of (y_1, y_2) given $y_1 + y_2 = 1, \mathbf{x}_1, \mathbf{x}_2, \mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w}$. By the hypothesis, $\mathbb{P}[(\mathbf{x}_1 - \mathbf{x}_2)'\boldsymbol{\theta} = c|\mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w}] = 0$ for any $c \in \mathbb{R}$, which in turn implies $\mathbb{P}[\mathbb{E}[m_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})|\mathbf{x}_1, \mathbf{x}_2, \mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w}] \neq \mathbb{E}[m_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)|\mathbf{x}_1, \mathbf{x}_2, \mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w}]] > 0$ for $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. Then, by standard arguments, the desired conclusion follows.

Assumption 2(i) We have $\|\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})\| \leq \|\mathbf{x}_1\| + \|\mathbf{x}_2\|$, and we can take $b(\mathbf{z}) = 1 + \|\mathbf{x}\|$.

Assumption 2(ii) Note $\|\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}) - \mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)\| \leq \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|\|\mathbf{x}_1 - \mathbf{x}_2\|^2$ and for any $\mathbf{w}$,

$$\begin{aligned}
&\|\mathbb{E}[\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})|\mathbf{z}_1, \mathbf{w}_2 = \mathbf{w}] - \mathbb{E}[\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)|\mathbf{z}_1, \mathbf{w}_2 = \mathbf{w}]\| f_{\mathbf{w}}(\mathbf{w}) \\
&\leq \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| C\big(\|\mathbf{x}_1\|^2 + 1\big)
\end{aligned}$$

$$\begin{aligned}
&\big\|\mathbb{E}[\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\vartheta})\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})'|\mathbf{w}_1, \mathbf{w}_2 = \mathbf{w}] - \mathbb{E}[\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)'|\mathbf{w}_1, \mathbf{w}_2 = \mathbf{w}]\big\| f_{\mathbf{w}}(\mathbf{w}) \\
&\leq \|\boldsymbol{\vartheta} - \boldsymbol{\theta}_0\| C\big(\|\mathbf{x}_1\|^3 + 1\big) + \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| C\big(\|\mathbf{x}_1\|^3 + 1\big)
\end{aligned}$$

where we use $(\mathbb{E}[\|\mathbf{x}\|^4|\mathbf{w}] + 1)f_{\mathbf{w}}(\mathbf{w}) \leq C$. Then, it suffices to show that

$$\mathbf{w}_2 \mapsto \mathbb{E}[\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)|\mathbf{z}_1, \mathbf{w}_2] \qquad \mathbf{w}_2 \mapsto \mathbb{E}[\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)\mathbf{s}_{\texttt{PLL}}(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0)'|\mathbf{w}_1, \mathbf{w}_2]$$

are continuous with probability one.

We have

$$\mathbb{E}[\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)|\mathbf{z}_1,\mathbf{w}_2] = (1-y_1)\int \Lambda(\mathbf{x}_2'\boldsymbol{\theta}_0+g(\mathbf{w}_2))\Lambda(\mathbf{x}_2'\boldsymbol{\theta}_0-\mathbf{x}_1'\boldsymbol{\theta}_0)(\mathbf{x}_1-\mathbf{x}_2)f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}_2)d\rho(\mathbf{x}_2)$$
$$- y_1\int \Lambda(\mathbf{x}_1'\boldsymbol{\theta}_0-\mathbf{x}_2'\boldsymbol{\theta}_0)(\mathbf{x}_1-\mathbf{x}_2)f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}_2)d\rho(\mathbf{x}_2)$$

and letting $\mathcal{L}(\mathbf{x}_1,\mathbf{w}_1,\mathbf{x}_2,\mathbf{w}_2)=\Lambda(\mathbf{x}_1'\boldsymbol{\theta}_0+g(\mathbf{w}_1))\{1-\Lambda(\mathbf{x}_2'\boldsymbol{\theta}_0+g(\mathbf{w}_2))\}\Lambda(\mathbf{x}_2'\boldsymbol{\theta}_0-\mathbf{x}_1'\boldsymbol{\theta}_0)^2$,

$$\mathbb{E}[\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)'|\mathbf{w}_1,\mathbf{w}_2]$$
$$=\int\{\mathcal{L}(\mathbf{x}_1,\mathbf{w}_1,\mathbf{x}_2,\mathbf{w}_2)+\mathcal{L}(\mathbf{x}_2,\mathbf{w}_2,\mathbf{x}_1,\mathbf{w}_1)\}(\mathbf{x}_1-\mathbf{x}_2)(\mathbf{x}_1-\mathbf{x}_2)'f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_1|\mathbf{w}_1)f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}_2)d\rho(\mathbf{x}_1)d\rho(\mathbf{x}_2).$$

Under the hypothesis, the dominated convergence theorem implies that both conditional expectations are continuous in $\mathbf{w}_2$ with probability one, proving the desired results.

Assumption 2(iii) For $\boldsymbol{\vartheta},\boldsymbol{\theta}$ with $\eta=\|\boldsymbol{\vartheta}-\boldsymbol{\theta}\|$,

$$\begin{aligned}&\left|m_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\vartheta})-m_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})-\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})'(\boldsymbol{\vartheta}-\boldsymbol{\theta})\right|\\ &\le \eta \sup_{\boldsymbol{\theta}_1:\|\boldsymbol{\theta}_1-\boldsymbol{\theta}\|\le\eta}\|\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_1)-\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})\|\le\eta^2\|\mathbf{x}_1-\mathbf{x}_2\|^2\end{aligned}$$

where the last inequality uses $\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})=-\mathbb{1}\{y_1\neq y_2\}(y_2-\Lambda(\mathbf{x}_2'\boldsymbol{\theta}-\mathbf{x}_1'\boldsymbol{\theta}))(\mathbf{x}_2-\mathbf{x}_1)$ and $|\Lambda(u)-\Lambda(v)|\le|u-v|$. Then, $|e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta},\mathbf{t},\tau)|\le\tau\|\mathbf{t}\|^2\|\mathbf{x}_1-\mathbf{x}_2\|^2$, and the first and second displays hold.

Using twice continuous differentiability of $\boldsymbol{\theta}\mapsto m_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})$, with $\|\boldsymbol{\vartheta}-\boldsymbol{\theta}\|=\eta$,

$$\begin{aligned}&\left|m_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\vartheta})-m_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})-\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})'(\boldsymbol{\vartheta}-\boldsymbol{\theta})-\frac{1}{2}(\boldsymbol{\vartheta}-\boldsymbol{\theta})\frac{\partial\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}(\boldsymbol{\vartheta}-\boldsymbol{\theta})\right|\\ &\le\eta^2\sup_{\boldsymbol{\theta}_1:\|\boldsymbol{\theta}_1-\boldsymbol{\theta}\|\le\eta}\left\|\frac{\partial\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_1}-\frac{\partial\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right\|\le\eta^3\|\mathbf{x}_1-\mathbf{x}_2\|^3.\end{aligned}$$

Then, noting $\mathbf{H}_{\mathrm{PLL}}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta})=\mathbb{E}[\partial\mathbf{s}_{\mathrm{PLL}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})/\partial\boldsymbol{\theta}'|\mathbf{w}_1,\mathbf{w}_2]$, we have

$$|e_2(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t},\tau)|\le\tau\|\mathbf{t}\|^3\mathbb{E}[\|\mathbf{x}_1-\mathbf{x}_2\|^3|\mathbf{w}_1,\mathbf{w}_2],$$

and the third display holds.

We have $\|\mathbf{H}_{\mathrm{PLL}}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta})\|\le2(\mathbb{E}[\|\mathbf{x}_1\|^2|\mathbf{w}_1]+\mathbb{E}[\|\mathbf{x}_2\|^2|\mathbf{w}_2])$ and the fourth display follows from $(\mathbb{E}[\|\mathbf{x}\|^4|\mathbf{w}]+1)f_{\mathbf{w}}(\mathbf{w})\le C$ and $\mathbb{E}[\|\mathbf{x}\|^4]<\infty$.

Assumption 2(iv) The bound $|e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta},\mathbf{t},\tau)|\le\tau\|\mathbf{t}\|^2\|\mathbf{x}_1-\mathbf{x}_2\|^2$ implies the first and second displays. Similarly, $|e_2(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t},\tau)|\le\tau\|\mathbf{t}\|^3\mathbb{E}[\|\mathbf{x}_1-\mathbf{x}_2\|^3|\mathbf{w}_1,\mathbf{w}_2]$ implies the third display.

For the fourth display, we have

$$\mathbf{H}_{\text{PLL}}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta})$$
$$= \int P(\mathbf{x}_1,\mathbf{w}_1,\mathbf{x}_2,\mathbf{w}_2)\lambda(\mathbf{x}_1'\boldsymbol{\theta}-\mathbf{x}_2'\boldsymbol{\theta})(\mathbf{x}_1-\mathbf{x}_2)(\mathbf{x}_1-\mathbf{x}_2)' f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_1|\mathbf{w}_1) f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}_2) d\rho(\mathbf{x}_1) d\rho(\mathbf{x}_2)$$

where

$$P(\mathbf{x}_1,\mathbf{w}_1,\mathbf{x}_2,\mathbf{w}_2) = \Lambda(\mathbf{x}_1'\boldsymbol{\theta}_0+g(\mathbf{w}_1))\Lambda(-\mathbf{x}_2'\boldsymbol{\theta}_0-g(\mathbf{w}_2))\} + \Lambda(-\mathbf{x}_1'\boldsymbol{\theta}_0-g(\mathbf{w}_1))\Lambda(\mathbf{x}_2'\boldsymbol{\theta}_0+g(\mathbf{w}_2))$$

is uniformly bounded and continuous in $\mathbf{w}_2$. We have

$$\begin{aligned}\|\mathbf{H}_{\text{PLL}}(\mathbf{w}_1,\mathbf{w}_1+\mathbf{u};\boldsymbol{\theta})-\mathbf{H}_{\text{PLL}}(\mathbf{w}_1,\mathbf{w}_1;\boldsymbol{\theta})\| \le \int &|P(\mathbf{x}_1,\mathbf{w}_1,\mathbf{x}_2,\mathbf{w}_1+\mathbf{u}) f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_1|\mathbf{w}_1) \\ &\times f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}_1+\mathbf{u}) - P(\mathbf{x}_1,\mathbf{w}_1,\mathbf{x}_2,\mathbf{w}_1) f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_1|\mathbf{w}_1) f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}_1)| \|\mathbf{x}_1-\mathbf{x}_2\|^2 d\rho(\mathbf{x}_1) d\rho(\mathbf{x}_2)\end{aligned}$$

where the dominating function does not depend on $\boldsymbol{\theta}$. Under the maintained hypothesis, the last integral goes to zero as $\mathbf{u}\to\mathbf{0}$ by the dominated convergence theorem. Thus, the fourth display holds.

For the fifth display, the dominated convergence theorem implies $\boldsymbol{\theta}\mapsto\mathbf{H}_{\text{PLL}}(\mathbf{w}_1,\mathbf{w}_1;\boldsymbol{\theta})$ is continuous with probability one. $\square$

5.3 Partially Linear Tobit Model

Let $\mathbf{x}_{12}=\mathbf{x}_1-\mathbf{x}_2$ and $y_{12}=y_1-y_2$. In this example, we have

$$\mathbf{s}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}) = \mathbf{s}_{\text{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}) = -\mathbf{x}_{12}\mathbb{1}\{y_{12}-\mathbf{x}_{12}'\boldsymbol{\theta}>0\}\mathbb{1}\{y_1>0\} + \mathbf{x}_{12}\mathbb{1}\{y_{12}-\mathbf{x}_{12}'\boldsymbol{\theta}<0\}\mathbb{1}\{y_2>0\}.$$

Below, we assume the existence of conditional density $f_{\varepsilon|\mathbf{w}}$ with respect to the Lebesgue measure. Define $\mu(\mathbf{x},\mathbf{w})=\mathbf{x}'\boldsymbol{\theta}_0+g(\mathbf{w})$ and

$$\begin{aligned}\eta(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}) = 2\int_0^\infty & f_{\varepsilon|\mathbf{w}}(u+\mathbf{x}_{12}'\boldsymbol{\theta}-\mu(\mathbf{x}_1,\mathbf{w}_1)|\mathbf{w}_1) f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x}_2,\mathbf{w}_2)|\mathbf{w}_2)du \\ &+ f_{\varepsilon|\mathbf{w}}(\mathbf{x}_{12}'\boldsymbol{\theta}-\mu(\mathbf{x}_1,\mathbf{w}_1)|\mathbf{w}_1)\int_{-\infty}^0 f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x}_2,\mathbf{w}_2)|\mathbf{w}_2)du.\end{aligned}$$

We take

$$\begin{aligned}\mathbf{H}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t}) = \mathbf{H}_{\text{PLT}}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t}) = \mathbb{E}\big[\{&\eta(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})(\mathbb{1}\{\mathbf{x}_{12}'\boldsymbol{\theta}>0\}+\mathbb{1}\{\mathbf{x}_{12}'\boldsymbol{\theta}=0,\mathbf{x}_{12}'\mathbf{t}>0\}) \\ &+\eta(\mathbf{z}_2,\mathbf{z}_1;\boldsymbol{\theta})(\mathbb{1}\{\mathbf{x}_{12}'\boldsymbol{\theta}<0\}+\mathbb{1}\{\mathbf{x}_{12}'\boldsymbol{\theta}=0,\mathbf{x}_{12}'\mathbf{t}<0\})\big\}\mathbf{x}_{12}\mathbf{x}_{12}'\big|\mathbf{w}_1,\mathbf{w}_2\big].\end{aligned}$$

Under the conditions stated below, $\mathbf{H}_{\text{PLT}}(\mathbf{w}_1,\mathbf{w}_1;\boldsymbol{\theta}_0,\mathbf{t})$ does not depend on $\mathbf{t}$. In this case, we drop the last argument.

To verify Assumption 2, we impose the following conditions.

Assumption 7. Let $\delta, C > 0$ be some fixed finite constants.

(i) With probability one $(\mathbb{E}[\|\mathbf{x}\|^4|\mathbf{w}]+1)f_{\mathbf{w}}(\mathbf{w}) \le C$, and $\mathbb{E}[\|\mathbf{x}\|^4] < \infty$.

(ii) The matrix $\mathbf{H}_0 = \int \mathbf{H}_{\mathtt{PLT}}(\mathbf{w},\mathbf{w};\boldsymbol{\theta}_0)f_{\mathbf{w}}(\mathbf{w})^2 d\mathbf{w}$ is positive definite.

(iii) $\mathbb{E}[\mathbf{x}_{12}\mathbf{x}_{12}'|\mathbf{w}_1 = \mathbf{w}, \mathbf{w}_2 = \mathbf{w}]$ is invertible for almost all $\mathbf{w}$. Conditional on $\mathbf{w}$, ε and $\mathbf{x}$ are statistically independent.

(iv) $\mathbf{w} \mapsto g(\mathbf{w})$ is continuous.

(v) The conditional distribution of ε given $\mathbf{w}$ has a Lebesgue density, denoted by $f_{\varepsilon|\mathbf{w}}$. For almost all $(e,\mathbf{w})$, $f_{\varepsilon|\mathbf{w}}(e|\mathbf{w}) \le C$, and $(e,\mathbf{w}) \mapsto f_{\varepsilon|\mathbf{w}}(e|\mathbf{w})$ is continuous. For each $\mathbf{x},\mathbf{w}$ in their support,

$$\int_{\mathbb{R}} \sup_{|v|,\|\mathbf{v}\|\le\delta} f_{\varepsilon|\mathbf{w}}(u - \mathbf{x}'\boldsymbol{\theta}_0 - g(\mathbf{w}) + v|\mathbf{w}+\mathbf{v})du \le C.$$

(vi) The conditional distribution of $\mathbf{x}$ given $\mathbf{w}$ has a density with respect to some measure ρ. Denoting the conditional density by $f_{\mathbf{x}|\mathbf{w}}$, $\int(1+\|\mathbf{x}\|^2)\sup_{\|\mathbf{v}\|\le\delta} f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w}+\mathbf{v})d\rho(\mathbf{x})$ is finite for almost all $\mathbf{w}$. Also, $\mathbf{w} \mapsto f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w})$ is continuous almost surely.

Proposition 5. *Under Assumptions 1(i)(iii) and 7, Assumptions 1(iv) and 2 hold for the partially linear Tobit example.*

Proof. Let $y_i^* = \mathbf{x}_i'\boldsymbol{\theta}_0 + g(\mathbf{w}_i) + \varepsilon_i$. Since the conditional distribution of ε given $\mathbf{x},\mathbf{w}$ has a Lebesgue density, we can write

$$f_{y^*|\mathbf{x},\mathbf{w}}(y^*|\mathbf{x},\mathbf{w}) = f_{\varepsilon|\mathbf{x},\mathbf{w}}(y^* - \mathbf{x}'\boldsymbol{\theta}_0 - g(\mathbf{w})|\mathbf{x},\mathbf{w}) = f_{\varepsilon|\mathbf{w}}(y^* - \mathbf{x}'\boldsymbol{\theta}_0 - g(\mathbf{w})|\mathbf{w}).$$

Assumption 1(iv) Since $|m_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})| \le (\|\mathbf{x}_1\|+\|\mathbf{x}_2\|)\|\boldsymbol{\theta}\|$, $\mathbb{E}[m_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})] < \infty$ for $\boldsymbol{\theta} \in \Theta$. Also, $\mathbb{E}[m(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|\mathbf{w}_1,\mathbf{w}_2]f_{\mathbf{w}}(\mathbf{w}_2) \le C\|\boldsymbol{\theta}\|(\mathbb{E}[\|\mathbf{x}_1\||\mathbf{w}_1]+1)$, which implies the second display.

For the first display, letting $y_{12} = y_1 - y_2$ and $\mathbf{x}_{12} = \mathbf{x}_1 - \mathbf{x}_2$,

$$\begin{aligned} m_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}) = &-\mathbb{1}\{y_1>0,y_2>0\}\mathbb{1}\{\operatorname{sgn}(y_{12}-\mathbf{x}_{12}'\boldsymbol{\theta}) = \operatorname{sgn}(y_{12})\}\mathbf{x}_{12}'\boldsymbol{\theta}\operatorname{sgn}(y_{12}) \\ &+ \mathbb{1}\{y_1>0,y_2>0\}\mathbb{1}\{\mathbf{x}_{12}'\boldsymbol{\theta} < y_{12} < 0\}\big(2y_{12} - \mathbf{x}_{12}'\boldsymbol{\theta}\big) \\ &+ \mathbb{1}\{y_1>0,y_2>0\}\mathbb{1}\{0 < y_{12} < \mathbf{x}_{12}'\boldsymbol{\theta}\}\big(\mathbf{x}_{12}'\boldsymbol{\theta} - 2y_{12}\big) \\ &- \mathbb{1}\{y_1>0,y_2=0\}\big(\mathbb{1}\{y_1>\mathbf{x}_{12}'\boldsymbol{\theta}\}\mathbf{x}_{12}'\boldsymbol{\theta} + \mathbb{1}\{0<y_1\le\mathbf{x}_{12}'\boldsymbol{\theta}\}y_1\big) \\ &+ \mathbb{1}\{y_1=0,y_2>0\}\big(\mathbb{1}\{y_2>-\mathbf{x}_{12}'\boldsymbol{\theta}\}\mathbf{x}_{12}'\boldsymbol{\theta} - \mathbb{1}\{0<y_2\le-\mathbf{x}_{12}'\boldsymbol{\theta}\}y_2\big). \end{aligned}$$

Then, $\mathbb{E}[m_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|\mathbf{w}_1,\mathbf{w}_2]$ is an integral with respect to

$$f_{y^*|\mathbf{x},\mathbf{w}}(y_1^*|\mathbf{x}_1,\mathbf{w}_1)f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_1|\mathbf{w}_1)f_{y^*|\mathbf{x},\mathbf{w}}(y_2^*|\mathbf{x}_2,\mathbf{w}_2)f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}_2|\mathbf{w}_2)dy_1^*dy_2^*d\mathbf{x}_1d\mathbf{x}_2$$

and the first display follows from the dominated convergence theorem under the maintained hypothesis.

$\boldsymbol{\theta}_0$ being a unique minimizer can be proven following the arguments in Honoré (1992) with appropriate modifications.

Assumption 2(i) We have $\|\mathbf{s}_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})\| \le \|\mathbf{x}_1\| + \|\mathbf{x}_2\|$, and we can take $b(\mathbf{z}) = 1 + \|\mathbf{x}\|$.

Assumption 2(ii) Note

$$\begin{aligned}\mathbb{E}[\mathbf{s}_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|\mathbf{z}_1,\mathbf{w}_2] = &-\mathbb{E}\Big[\mathbf{x}_{12}\int_0^{y_1-\mathbf{x}_{12}'\boldsymbol{\theta}} f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x}_2,\mathbf{w}_2)|\mathbf{w}_2)du\Big|\mathbf{z}_1,\mathbf{w}_2\Big]\mathbb{1}\{y_1>0\} \quad (5.2)\\ &-\mathbb{E}\Big[\mathbf{x}_{12}\mathbb{1}\{y_1>\mathbf{x}_{12}'\boldsymbol{\theta}\}\int_{-\infty}^{0} f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x}_2,\mathbf{w}_2)|\mathbf{w}_2)du\Big|\mathbf{z}_1,\mathbf{w}_2\Big]\mathbb{1}\{y_1>0\}\\ &+\mathbb{E}\Big[\mathbf{x}_{12}\int_{0\vee(y_1-\mathbf{x}_{12}'\boldsymbol{\theta})}^{\infty} f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x}_2,\mathbf{w}_2)|\mathbf{w}_2)du\Big|\mathbf{z}_1,\mathbf{w}_2\Big]\end{aligned}$$

Using that $f_{\varepsilon|\mathbf{w}}$ is bounded,

$$\begin{aligned}&\Big|\mathbb{E}[\mathbf{s}_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|\mathbf{z}_1,\mathbf{w}_2] - \mathbb{E}[\mathbf{s}_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)|\mathbf{z}_1,\mathbf{w}_2]\Big|\\ &\le C\big(\|\mathbf{x}_1\|^2+\mathbb{E}[\|\mathbf{x}_2\|^2|\mathbf{w}_2]\big)\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\| + \mathbb{E}\big[\|\mathbf{x}_{12}\|\mathbb{1}\{\mathbf{x}_{12}'\boldsymbol{\theta}<y_1\le\mathbf{x}_{12}'\boldsymbol{\theta}_0\}|\mathbf{z}_1,\mathbf{w}_2\big]\mathbb{1}\{y_1>0\}\\ &\quad+\mathbb{E}\big[\|\mathbf{x}_{12}\|\mathbb{1}\{\mathbf{x}_{12}'\boldsymbol{\theta}_0<y_1\le\mathbf{x}_{12}'\boldsymbol{\theta}\}|\mathbf{z}_1,\mathbf{w}_2\big]\mathbb{1}\{y_1>0\}\end{aligned}$$

and the terms after the inequality goes to zero almost surely as $\boldsymbol{\theta}\to\boldsymbol{\theta}_0$ by the dominated convergence theorem, using $\int(1+\|\mathbf{x}\|)\sup_{\|\mathbf{v}\|\le\delta} f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w}+\mathbf{v})d\rho(\mathbf{x})<\infty$ with probability one.

Now we verify $\mathbf{w}\mapsto\mathbb{E}[\mathbf{s}_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)|\mathbf{z}_1,\mathbf{w}_2=\mathbf{w}]$ is continuous almost surely.

$$\begin{aligned}&\mathbb{E}\Big[\mathbf{x}_{12}\int_0^{y_1-\mathbf{x}_{12}'\boldsymbol{\theta}_0} f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x}_2,\mathbf{w}_2)|\mathbf{w}_2)du\Big|\mathbf{z}_1,\mathbf{w}_2\Big]\\ &=\int\int_0^{y_1-(\mathbf{x}_1-\mathbf{x})'\boldsymbol{\theta}_0}(\mathbf{x}_1-\mathbf{x})f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x},\mathbf{w}_2)|\mathbf{w}_2)du f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w}_2)d\rho(\mathbf{x})\end{aligned}$$

and the integrand is continuous by the hypothesis. Thus the dominated convergence theorem implies the desired continuity. Doing similar calculations for each of the terms in (5.2), we see that

$$\lim_{\mathbf{u}\to\mathbf{0}}\Big|\mathbb{E}[\mathbf{s}_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)|\mathbf{z}_1,\mathbf{w}_2=\mathbf{w}_1+\mathbf{u}] - \mathbb{E}[\mathbf{s}_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta}_0)|\mathbf{z}_1,\mathbf{w}_2=\mathbf{w}_1]\Big| = 0$$

almost surely. Thus, the first display holds.

Letting $\tilde{f}(y,\mathbf{x},\mathbf{w}) = f_{\varepsilon|\mathbf{w}}(y-\mu(\mathbf{x},\mathbf{w})|\mathbf{w})f_{\mathbf{x}|\mathbf{w}}(\mathbf{x}|\mathbf{w})$, we have

$$\mathbb{E}[\mathbf{s}_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})\mathbf{s}_{\mathtt{PLT}}(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\vartheta})'|\mathbf{w}_1,\mathbf{w}_2]$$

$$
\begin{aligned}
&= \iint \int_{-\infty}^{\infty} \int_{0\vee(\max\{y_2^*,0\}+\max\{\mathbf{x}_{12}'\boldsymbol{\theta},\mathbf{x}_{12}'\boldsymbol{\vartheta}\})}^{\infty} \mathbf{x}_{12}\mathbf{x}_{12}'\tilde{f}(y_1^*,\mathbf{x}_1,\mathbf{w}_1)\tilde{f}(y_2^*,\mathbf{x}_2,\mathbf{w}_2)dy_1^*dy_2^*d\rho(\mathbf{x}_1)d\rho(\mathbf{x}_2)\\
&\quad - \iint \int_{0}^{\infty} \int_{y_2^*+\mathbf{x}_{12}'\boldsymbol{\theta}}^{y_2^*+\mathbf{x}_{12}'\boldsymbol{\vartheta}} \mathbf{x}_{12}\mathbf{x}_{12}'\tilde{f}(y_1^*,\mathbf{x}_1,\mathbf{w}_1)\tilde{f}(y_2^*,\mathbf{x}_2,\mathbf{w}_2)\mathbb{1}\{0<\mathbf{x}_{12}'(\boldsymbol{\vartheta}-\boldsymbol{\theta})\}dy_1^*dy_2^*d\rho(\mathbf{x}_1)d\rho(\mathbf{x}_2)\\
&\quad - \iint \int_{0}^{\infty} \int_{y_2^*+\mathbf{x}_{12}'\boldsymbol{\vartheta}}^{y_2^*+\mathbf{x}_{12}'\boldsymbol{\theta}} \mathbf{x}_{12}\mathbf{x}_{12}'\tilde{f}(y_1^*,\mathbf{x}_1,\mathbf{w}_1)\tilde{f}(y_2^*,\mathbf{x}_2,\mathbf{w}_2)\mathbb{1}\{0>\mathbf{x}_{12}'(\boldsymbol{\vartheta}-\boldsymbol{\theta})\}dy_1^*dy_2^*d\rho(\mathbf{x}_1)d\rho(\mathbf{x}_2)\\
&\quad + \iint \int_{-\infty}^{\infty} \int_{0\vee(\max\{y_1^*,0\}-\min\{\mathbf{x}_{12}'\boldsymbol{\theta},\mathbf{x}_{12}'\boldsymbol{\vartheta}\})}^{\infty} \mathbf{x}_{12}\mathbf{x}_{12}'\tilde{f}(y_1^*,\mathbf{x}_1,\mathbf{w}_1)\tilde{f}(y_2^*,\mathbf{x}_2,\mathbf{w}_2)dy_2^*dy_1^*d\rho(\mathbf{x}_1)d\rho(\mathbf{x}_2).
\end{aligned}
$$

Then, the second display follows from the dominated convergence theorem.

Assumption 2(iii) Note

$$
\begin{aligned}
&(y_{12}-\mathbf{x}_{12}'\boldsymbol{\vartheta})\mathbb{1}\{y_{12}>\mathbf{x}_{12}'\boldsymbol{\vartheta}\}-(y_{12}-\mathbf{x}_{12}'\boldsymbol{\theta})\mathbb{1}\{y_{12}>\mathbf{x}_{12}'\boldsymbol{\theta}\}+\mathbf{x}_{12}'(\boldsymbol{\vartheta}-\boldsymbol{\theta})\mathbb{1}\{y_{12}>\mathbf{x}_{12}'\boldsymbol{\theta}\}\\
&=(y_{12}-\mathbf{x}_{12}'\boldsymbol{\vartheta})\big(\mathbb{1}\{0<y_{12}-\mathbf{x}_{12}'\boldsymbol{\vartheta}\le-\mathbf{x}_{12}'(\boldsymbol{\vartheta}-\boldsymbol{\theta})\}-\mathbb{1}\{-\mathbf{x}_{12}'(\boldsymbol{\vartheta}-\boldsymbol{\theta})<y_{12}-\mathbf{x}_{12}'\boldsymbol{\vartheta}\le 0\}\big)
\end{aligned}
$$

and

$$
\begin{aligned}
&(y_{12}-\mathbf{x}_{12}'\boldsymbol{\vartheta})\mathbb{1}\{y_{12}<\mathbf{x}_{12}'\boldsymbol{\vartheta}\}-(y_{12}-\mathbf{x}_{12}'\boldsymbol{\theta})\mathbb{1}\{y_{12}<\mathbf{x}_{12}'\boldsymbol{\theta}\}+\mathbf{x}_{12}'(\boldsymbol{\vartheta}-\boldsymbol{\theta})\mathbb{1}\{y_{12}<\mathbf{x}_{12}'\boldsymbol{\theta}\}\\
&=(y_{12}-\mathbf{x}_{12}'\boldsymbol{\vartheta})\big(\mathbb{1}\{-\mathbf{x}_{12}'(\boldsymbol{\vartheta}-\boldsymbol{\theta})\le y_{12}-\mathbf{x}_{12}'\boldsymbol{\vartheta}<0\}-\mathbb{1}\{0\le y_{12}-\mathbf{x}_{12}'\boldsymbol{\vartheta}<-\mathbf{x}_{12}'(\boldsymbol{\vartheta}-\boldsymbol{\theta})\}\big)
\end{aligned}
$$

These expressions imply

$$
\begin{aligned}
|e_1(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta},\mathbf{t},\tau)| &\le |\mathbf{x}_{12}'\mathbf{t}|\mathbb{1}\{|y_{12}-\mathbf{x}_{12}'(\boldsymbol{\theta}+\mathbf{t}\tau)|\le|\mathbf{x}_{12}'\mathbf{t}|\tau\}\big(\mathbb{1}\{y_1>0\}+\mathbb{1}\{y_2>0\}\big) \\
&\le \|\mathbf{t}\|(\|\mathbf{x}_1\|+\|\mathbf{x}_2\|)
\end{aligned}
\tag{5.3}
$$

and the first two displays hold.

Let $\nu(u,v,\mathbf{x},\mathbf{w}) = f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x},\mathbf{w})+v|\mathbf{w}) - f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x},\mathbf{w})|\mathbf{w})$. Following calculations in Honoré (1992), we can show

$$
\begin{aligned}
|e_2(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta},\mathbf{t},\tau)| &\le |\mathbf{x}_{12}'\mathbf{t}|^2\int_0^{\infty}\sup_{|v|\le|\mathbf{x}_{12}'\mathbf{t}|\tau}\big|\nu(u+\mathbf{x}_{12}'\boldsymbol{\theta},v,\mathbf{x}_1,\mathbf{w}_1)\big|f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x}_2,\mathbf{w}_2)|\mathbf{w}_2)du \\
&\quad + |\mathbf{x}_{12}'\mathbf{t}|^2\int_0^{\infty}\sup_{|v|\le|\mathbf{x}_{12}'\mathbf{t}|\tau}\big|\nu(u-\mathbf{x}_{12}'\boldsymbol{\theta},v,\mathbf{x}_2,\mathbf{w}_2)\big|f_{\varepsilon|\mathbf{w}}(u-\mu(\mathbf{x}_1,\mathbf{w}_1)|\mathbf{w}_1)du \\
&\quad + |\mathbf{x}_{12}'\mathbf{t}|^2\Big(\sup_{|v|\le|\mathbf{x}_{12}'\mathbf{t}|\tau}\big|\nu(\mathbf{x}_{12}'\boldsymbol{\theta},v,\mathbf{x}_1,\mathbf{w}_1)\big|+\sup_{|v|\le|\mathbf{x}_{12}'\mathbf{t}|\tau}\big|\nu(-\mathbf{x}_{12}'\boldsymbol{\theta},v,\mathbf{x}_2,\mathbf{w}_2)\big|\Big) \\
&\le C\|\mathbf{t}\|^2\big(\|\mathbf{x}_1\|^2+\|\mathbf{x}_2\|^2\big).
\end{aligned}
\tag{5.4}
$$

Then, the third display holds.

For the fourth display, note $|\eta(\mathbf{z}_1,\mathbf{z}_2;\boldsymbol{\theta})|\le C$ with probability one, and $\|\mathbf{H}_{\text{PLT}}(\mathbf{w}_1,\mathbf{w}_2;\boldsymbol{\theta},\mathbf{t})\|\le C(\mathbb{E}[\|\mathbf{x}_1\|^2|\mathbf{w}_1]+\mathbb{E}[\|\mathbf{x}_2\|^2|\mathbf{w}_2])$. Thus, the desired result holds.

Assumption 2(iv) The first three displays follow from (5.3), (5.4), and the dominated convergence theorem.

Let $\Delta f(u, \mathbf{x}, \mathbf{w}, \mathbf{v}) = |f_{\varepsilon|\mathbf{w}}(u - \mu(\mathbf{x}, \mathbf{w} + \mathbf{v})|\mathbf{w} + \mathbf{v}) - f_{\varepsilon|\mathbf{w}}(u - \mu(\mathbf{x}, \mathbf{w})|\mathbf{w})|$. For each $\mathbf{x}_1, \mathbf{x}_2$, with sufficiently small $\epsilon > 0$,

$$\begin{aligned}
&\sup_{\boldsymbol{\theta} \in \Theta_0^\epsilon} \left| \eta(\mathbf{x}_1, \mathbf{w}_1, \mathbf{x}_2, \mathbf{w}_2 + \mathbf{v}; \boldsymbol{\theta}) - \eta(\mathbf{x}_1, \mathbf{w}_1, \mathbf{x}_2, \mathbf{w}_2; \boldsymbol{\theta}) \right| \\
&\leq 2 \int_0^\infty \sup_{|v| \leq \delta} f_{\varepsilon|\mathbf{w}}(u + v - \mu(\mathbf{x}_2, \mathbf{w}_1)|\mathbf{w}_1) \Delta f(u, \mathbf{x}_2, \mathbf{w}_2, \mathbf{v}) du + C \int_{-\infty}^0 \Delta f(u, \mathbf{x}_2, \mathbf{w}_2, \mathbf{v}) du
\end{aligned}$$

where the terms after the inequality go to zero as $\mathbf{v} \to 0$ by the dominated convergence theorem. Then, the fourth display follows from another application of the dominated convergence theorem.

The fifth display and $\mathbf{H}(\mathbf{w}_1, \mathbf{w}_1; \boldsymbol{\theta}_0, \mathbf{t})$ not depending on $\mathbf{t}$ follow from continuity of $\boldsymbol{\theta} \mapsto \eta(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta})$ with probability one and $\eta(\mathbf{z}_1, \mathbf{z}_2; \boldsymbol{\theta}_0) = \eta(\mathbf{z}_2, \mathbf{z}_1; \boldsymbol{\theta}_0)$ on the event $\{\mathbf{x}_{12}' \boldsymbol{\theta}_0 = 0\}$. $\square$

6 Conclusion

This paper developed robust distribution theory and bootstrap-based inference for a broad class of convex pairwise difference estimators. First, we established a general Gaussian distributional approximation based on small bandwidth asymptotics and debiasing via the generalized jackknife. Second, we showed that the nonparametric bootstrap leads to conservative inference due to variance inflation when localization is high (bandwidth is small). Third, we proposed a new bootstrap-based inference method that is asymptotically valid, thereby offering more robust uncertainty quantification for pairwise difference estimators. Our theoretical work carefully preserved and leveraged convexity of the objective function, which led to improved sufficient high-level conditions. We illustrated our robust inference methods with three examples in the context of partially linear regression, Logit, and Tobit models.

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 (Cattaneo et al., 2025b).

References

Ahn, 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.

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*.

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.

——— (2025b): "U-Process M-Estimation: Distribution Theory and Bootstrap Inference," .

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 (2025c): "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," *The 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 in Econometric Models*, Cambridge University Press, 520–553.

Jochmans, K. (2013): "Pairwise-Comparison Estimation with Non-parametric Controls," *The 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, J. L. (1994): "Estimation of Semiparametric Models," in *Handbook of Econometrics, Volume IV*, ed. by R. F. Engle and D. L. McFadden, New York: 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 Science & Business Media.

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