KIER DISCUSSION PAPER SERIES

KYOTO INSTITUTE
OF
ECONOMIC RESEARCH

Discussion Paper No.1053

"Optimal Minimax Rates of Specification Testing
with Data-driven Bandwidth"

Kohtaro Hitomi, Masamune Iwasawa and Yoshihiko Nishiyama

January 2021

KYOTO UNIVERSITY

KYOTO, JAPAN

Optimal Minimax Rates of Specification Testing with Data-driven Bandwidth

Kohtaro Hitomi* Masamune Iwasawa† Yoshihiko Nishiyama‡

January 29, 2021

*Kyoto Institute of Technology, Matsugasaki, Sakyo-ku, Kyoto, 606-8585, JAPAN, hitomi@kit.ac.jp

†Corresponding author: Masamune Iwasawa, Department of Economics, Otaru University of Commerce, 3-5-21 Midori, Otaru, Hokkaido 047-8501, Japan, Tel:+81(0)134275324, masamune-iwasawa@res.otaru-uc.ac.jp

‡Institute of Economic Research, Kyoto University, Yoshidahonmachi, Sakyoku, Kyoto, 606-8501, JAPAN, nishiyama.yoshihiko.3u@kyoto-u.ac.jp

Abstract

This study investigates optimal minimax rates of specification testing for linear and non-linear instrumental variable regression models. The rate implies that the uniform power of tests reduces when the dimension of instruments is large. The test constructed by non-parametric kernel techniques can be rate optimal when bandwidths satisfy two order conditions that depend on the dimensions of instruments and the smoothness of alternatives. Since bandwidths are often chosen in a data-dependent way in empirical studies, the rate optimality of the test with data-driven bandwidths are investigated. Bandwidths selected by the least squares cross-validation can satisfy conditions for the rate optimality.

Keywords: optimal minimax rate; specification test; instrumental variable regression; non-parametric kernel method; bandwidth selection

JEL Classification: C12; C14

4804 Words

1 Introduction

In the context of specification tests for the functional form of regression models, the minimax approach can be used to investigate uniform power against a set of alternatives (Ingster, 1993). In this approach, a set of alternatives can be defined to approach the null model at a specific rate. The maximum rate at which a test can uniformly detect any alternatives in this set is called the optimal minimax rate. Although the investigation of uniform power provides a deeper understanding of specification testing, research in this area is limited.

The optimal minimax rates for regression models have been investigated by Guerre and Lavergne (2002). Recently, a test based on the commonly used non-parametric K-nearest neighbors technique was shown to be rate optimal (H. Li, Li, & Liu, 2016). Hitomi, Iwasawa, and Nishiyama (2020) showed that a test based on the distance between non-parametric and parametric variance estimators is rate optimal against a set of non-smooth alternatives. However, optimal minimax rates of specification testing for other models, such as instrumental variable (IV) regression models, and rate optimality of other types of tests, such as kernel-type tests, have not been investigated.[1]

This study investigates the optimal minimax rates of specification testing for IV regression models. We find that the optimal minimax rate is $n^{-2(s+k)/[l_z+4(s+k)]}$, where n is the sample size, $s+k$ represents the smoothness of alternatives, as explained later in detail, and l_z is the dimension of instrument z, when the set of alternatives is smooth such that $s+k \geq l_z/4$. The rate implies that the uniform power of tests reduces when the dimension of instruments is large.

We adapt the kernel-type test proposed by Zheng (1996) for IV regression models. This test is based on the non-parametric kernel estimator for the conditional mean of the error term given instruments. The proposed test weakly converges to the standard normal distribution under the null hypothesis and is rate optimal when bandwidth h for the kernel satisfies $h^{\min\{q_k,k+q_z\}/2} n^{\frac{s+k}{l_z+4(s+k)}} = O(1)$ and $h^{-1} n^{\frac{-2}{l_z+4(s+k)}} = O(1)$, where the density of instruments is q_z-times continuously differentiable and the q_kth-order kernel is used.

In practice, bandwidths are often chosen by using data-driven methods such

[1] In the adaptive framework, in which the smoothness of the classes of alternatives is unknown, Horowitz and Spokoiny (2001) showed that their test based on non-parametric kernel techniques is rate optimal. This point will be discussed later in detail.

as the least squares cross-validation. In the non-parametric kernel estimator for the regression function, when the original bandwidth is replaced with one selected by a data-driven method, asymptotic normality still holds if some additional assumptions are made (see Racine & Li, 2004). However, it is not clear whether the rate optimality of the kernel-type test remains the same when data-driven bandwidths are used. Thus, we investigate the rate optimality of the kernel-type test when data-driven bandwidths $\hat{h}$ are used instead of h. We find that some additional assumptions regarding the kernel lead to parallel conditions on bandwidths that, in turn, ensure that the test is rate optimal. Furthermore, we show that the conditions are satisfied by bandwidths selected by the cross-validation when $4(s+k) \leq l_z + 8$ and $\min\{q_k, k+q_z\}[l_z + 4(s+k)] \geq 2(l_z+4)(s+k)$. This implies that bandwidths selected by the least squares cross-validation method can ensure that the test is rate optimal, although the procedure is designed for estimation rather than testing. In this sense, this study complements the results from Gao and Gijbels (2008), in which a bandwidth selection method that maximizes the power against a Pitman-type local alternative is proposed.

Specification tests for IV regression models were first developed by Donald, Imbens, and Newey (2003) and Tripathi and Kitamura (2003).[2] Tripathi and Kitamura (2003) proposed a smoothed empirical likelihood ratio-based test. Following Härdle and Mammen (1993) and Aït-Sahalia, Bickel, and Stoker (2001), Holzmann (2008) proposed a test for IV regression models using the squared distance between the parametric model and its non-parametric kernel estimates. The test proposed by Horowitz (2006) takes a form resembling the ICM test. Gørgens

[2]For a recent review of the development of specification testing, see González-Manteiga and Crujeiras (2013).

and Würtz (2012) proposed another type of test based on a sequence of Lagrange multiplier (LM) statistics. However, optimal minimax rates for IV regression models in a framework comparable with Guerre and Lavergne (2002) have not been investigated to date.

Specification tests that use a non-parametric kernel estimator are considered by Härdle and Mammen (1993), Zheng (1996), and Horowitz and Spokoiny (2001), among others. Horowitz and Spokoiny (2001) proposed a test that is adaptive to the unknown smoothness of the set of alternatives and showed the rate optimality of the test in the adaptive framework. The authors consider a family of test statistics, say, $\{T_n(h), h \in H_n\}$, where H_n represents finite sets of bandwidth values, and the test statistic is defined by $T = \max_{h \in H_n} T_n(h)$. The adaptiveness and the rate optimality of their test result from its use of the set of bandwidths. The choice of the set is important also for empirical studies, since the larger the set is, more intensive the computation become. However, the bandwidth selection approaches commonly used in applied research, such as the least squares cross-validation, find a single bandwidth.[3] Thus, from the practical point of view, characteristics of the kernel-type test with a single bandwidth is of great interest. Nonetheless, the rate optimality of these tests is yet to be formally validated in the literature. To the best of our knowledge, this is the first study that considers the optimal minimax rate of the kernel-type test with data-driven bandwidths. It is notable that the rate optimality of tests with data-driven bandwidths is not trivial, even if the optimality of tests with a deterministic sequence of bandwidths has been investigated. The core contribution of this paper is to show that the test can be

[3]To the best of our knowledge, how to select an appropriate set of bandwidth values H_n is an open question.

rate optimal when it is evaluated with bandwidths selected using the least squares cross-validation method.

The remainder of this paper is organized as follows. Section 2 introduces the model and testing framework. Section 3 shows the optimal minimax rate for the IV regression model. Section 4 proposes a kernel smoothing test and exemplifies its rate optimality under deterministic and data-driven bandwidths. Section 5 reports simulation results that demonstrate the test's encouraging finite sample performance. Following Horowitz (2006), the size and power properties of the proposed test are compared with those of various existing tests. Section 6 concludes the paper and discusses future research avenues.

2 Framework

Let $(Y, X, Z) \in \mathbb{R} \times \mathbb{R}^{l_x} \times \mathbb{R}^{l_z}$ be random variables. We consider parametric models

$$Y = g(X, \theta) + u, \tag{1}$$

where $g(X, \theta)$ is a known function defined up to parameters $\theta \in \Theta$, Θ is a compact subset of $\mathbb{R}^{l_\theta}$ with $l_\theta \leq l_z$, and u is an error term. The hypotheses to be tested are

$$H_0 : E(u|Z) = 0.$$

The null hypothesis is equivalent to saying that there exists $\theta_0 \in \Theta$ that satisfies $E(Y|Z) = E[g(X, \theta_0)|Z]$ almost surely (a.s.). The null hypothesis considers regression models when $Z = X$, and instrumental regression models when Z includes a subset of X, along with some other exogenous variables.

We examine the asymptotic power properties of testing by employing the minimax approach of Ingster (1993), in which the alternative hypothesis is a set of functions belonging to a smoothness class. Let $\mathcal{M}_{L,s,k}$ be a class of functions defined on a compact set Ω, such that:

$$\mathcal{M}_{L,s,k} = \left\{ m : \sum_{j=0}^{k} \sup_{|\beta|=j} \sup_{x\in\Omega} \|D^{\beta}m(x)\| + \sup_{|\beta|=k} \sup_{x,y\in\Omega} \frac{\|D^{\beta}m(x) - D^{\beta}m(y)\|}{\|x-y\|^{s}} \leq L \right\},$$

This applies for some smoothness index $s \in (0, 1]$, a non-negative integer k, and a positive constant L, where $\|\cdot\|$ denotes the Euclidean norm. $D^{\beta}m(x)$ indicates $|\beta|$-times partial derivatives of $m(\cdot)$. Then, the alternative hypothesis is defined as follows:

$$H_{n,1} : \mathcal{M}(\rho_n) = \left\{ m(\cdot) \in \mathcal{M}_{L,s,k} : \inf_{\theta\in\Theta} E\left\{[m(Z) - E[g(X,\theta)|Z]]^2\right\} \geq \rho_n^2. \right\},$$

where $m(Z) \equiv E(Y|Z)$. The minimax approach finds the fastest rate at which ρ_n approaches 0, while assuring the uniform detection of alternatives in $\mathcal{M}(\rho_n)$. The alternatives considered in this study are parallel to those in Guerre and Lavergne (2002).

The following notations are used throughout the paper. The true parameter θ_0 of the parametric model is defined such that $m(Z) = E[g(X,\theta_0)|Z]$. We denote $\delta_{\theta}(Z) \equiv m(Z) - E[g(X,\theta)|Z]$ and $\omega \equiv Y - m(Z)$, where $E(\omega|Z) = 0$ by definition. The variance of u is denoted by $\sigma^2(z) \equiv E(u^2|Z = z)$. For any $\sqrt{n}$-consistent estimator $\hat{\theta}$ of θ, residuals of the parametric model are denoted by $\hat{u} = Y - g(X, \hat{\theta})$.

3 Optimal Minimax Rate

We list the assumptions to establish the optimal minimax rate for IV regression models.

Assumption 1. *$\{Y_i, X_i, Z_i\}_{i=1}^n$ are a random sample on $(Y, X, Z) \in \mathbb{R} \times \mathbb{R}^{l_x} \times \mathbb{R}^{l_z}$, where l_x and l_z are finite. $E(\omega^2|Z = z)$ is continuously differentiable and bounded away from zero. A positive constant $M < \infty$ exists such that $E(|\omega|^4|Z) < M$ almost surely.*

Assumption 2. *The density of Z, denoted by $f(\cdot) : \mathbb{R}^{l_z} \to \mathbb{R}$, has compact support (without loss of generality $[0,1]^{l_z}$), satisfies $0 < \underline{f} \leq f(z) \leq \overline{f} < \infty$ for any $z \in [0,1]^{l_z}$, and is q_z-times continuously differentiable on $(0,1)^{l_z}$, where $q_z > 1$.*

Assumption 3. *For each x, $g(x, \theta)$ is twice continuously differentiable with respect to θ.*

Assumption 4. *For each $\theta \in \Theta$, $E[g(X, \theta)^4]$ is bounded from above.*

Assumption 5. *$E\left[\sup_{\theta \in \Theta} \|\frac{\partial}{\partial \theta} g(X, \theta)\|^2\right]$ is bounded from above.*

Assumption 6. *$E\left[\sup_{\theta \in \Theta} \|\frac{\partial}{\partial \theta \partial \theta'} g(X, \theta)\|^4\right]$ is bounded from above.*

Assumption 7. *For each $\theta \in \Theta$, $E[g(X, \theta)^2|Z] < \infty$ a.s.*

Assumption 8. *For each $\theta \in \Theta$, $E[g(X, \theta)|Z = z] \in \mathcal{M}_{L_{\mathcal{M}}, s, k}$ for some s, k, and $L_{\mathcal{M}} \leq L$.*

Assumption 9. *For each $\theta \in \Theta$, $G_\theta \equiv \frac{\partial}{\partial \theta} E[g(X, \theta)|Z = z]$ is Lipschitz continuous with respect to z with support on Z and $E(G_\theta G_\theta')$ is non-singular.*

Assumption 10. *Under the null hypothesis, we have a* $\sqrt{n}$*-consistent estimator* $\hat{\theta}_n$ *of* θ_0.

Assumption 11. *(i) For each* $m(\cdot) \in \mathcal{M}_{L,s,k}$*, there exists a unique pseudo-true value* θ_m^* *with respect to* $\hat{\theta}$.

(ii) $\sqrt{n}(\hat{\theta} - \theta_m^*) = O_p(1)$ *uniformly with respect to* $m(\cdot) \in \mathcal{M}_{L,s,k}$

(iii) For each $m(\cdot) \in \mathcal{M}_{L,s,k}$ *and a bounded function* $h(\cdot,\cdot)$*, a positive constant* c *exists such that* $\|\theta_m^* - \theta_0\| \leq c \int |E[h(\cdot, z)\delta_{\theta_0}(\cdot)]| f(z)dz$.

Assumptions 1 to 9 are standard in the literature (Guerre & Lavergne, 2002). Exceptions are dominance conditions, that is, assumptions 5 and 6, which guarantee uniform convergence of $\frac{1}{n}\sum_{i=1}^n \left\|\frac{\partial}{\partial\theta} g(X_i,\theta)\right\|^2$ and $\frac{1}{n}\sum_{i=1}^n \left\|\frac{\partial}{\partial\theta\partial\theta'} g(X_i,\theta)\right\|^2$ together with Assumption 3. The dominance conditions do not exclude the possibility that $g(\cdot,\theta)$ is linear, while linear models with unbounded regressors (e.g., normally distributed regressors) are excluded in Guerre and Lavergne (2002). The dominance condition for the first derivative is a standard assumption required for the asymptotic normality of commonly used estimators, such as the generalized method of moments (GMM). Assumption 9 is a key assumption for the existence of a parameter that satisfies $\inf_{\theta\in\Theta} E\left[\delta_\theta(Z)^2\right]$.

Assumption 10 requires a $\sqrt{n}$-consistent parametric estimator $\hat{\theta}_n$ of θ_0 under the null hypothesis. Assumption 11 restricts the behavior of the estimator under the alternative hypothesis. We illustrate these assumptions with two examples. For notational simplicity, subscripts are omitted. That is, $\hat{\theta}_n = \hat{\theta}$ and $\theta_m^* = \theta^*$ in all equations where no confusion will arise from this simplification.

Example 1. *(GMM estimators) Note that the null model is defined in terms of* *conditional moment restrictions, while the objective function of the GMM estimator is based on a finite number of unconditional moment restrictions. If $g(\cdot)$ is linear in parameters and the dimensions of the parameter vector are equal to the dimensions of the instrument, the GMM (two-stage least squares) estimator based on a finite number of unconditional moment restrictions satisfies Assumption 10 under regularity conditions (Hansen, 1982). When $g(\cdot)$ is non-linear, however, the GMM estimator based on a finite number of unconditional moment restrictions may be inconsistent (Dominguez & Lobato, 2004). The existence of the unique pseudo-true value in Assumption 11 (i) implicitly demands the identification condition that, for each $m(\cdot) \in \mathcal{M}_{L,s,k}$, $Q_m(\theta^*_m) < Q_m(\theta)$ for all $\theta \in \Theta \backslash \theta^*_m$, where $Q_m(\theta)$ is the GMM objective function in the population. The uniformity in Assumption 11 (ii) is essential for rate optimality, and a similar condition is assumed in previous studies of rate optimal testing (Guerre & Lavergne, 2002, Horowitz, 2006). Further, the asymptotic behaviors of the GMM estimator in misspecified models depend on the weighting matrix. For example, Hall and Inoue (2003) showed that a fixed weighting matrix or a sequence of weighting matrices with $\sqrt{n}$-asymptotic normality is required for the $\sqrt{n}$ asymptotic normality of the GMM estimator. To investigate Assumption 11 (iii), let us consider the first-order condition of the minimization problem for the GMM estimator, which is $H'_{\theta^*_m} W E(Zu^*) = 0$, where $H_\theta \equiv E[Z \frac{\partial}{\partial \theta'} g(X, \theta)]$, W is a $l_z \times l_z$ weighting matrix, and $u^* \equiv Y - g(X, \theta^*_m)$. Applying the mean value theorem to the first-order condition yields*

$$\theta^*_m - \theta_0 = \left(H'_{\theta^*_m} W H_{\tilde{\theta}} \right)^{-1} H'_{\theta^*_m} W E \left[Z \delta_{\theta_0}(Z) \right], \tag{2}$$

where $\tilde{\theta}$ is a segment between θ^*_m and θ_0. Thus, given the existence of the inverse *of* $H'_{\theta_m^*} W H_{\tilde{\theta}}$, *we obtain* $\|\theta_m^* - \theta_0\| \leq c|E[h(Z,\cdot)\delta_{\theta_0}(Z)]|$, *where* $h(Z,\cdot) = Z$ *and* $c = \|(H'_{\theta_m^*} W H_{\tilde{\theta}})^{-1}\| \|H_{\theta_m^*}\| \|W\| < \infty$.

Example 2. *Estimators using a continuum of unconditional moment restrictions such as those defined in Carrasco and Florens (2000) and Dominguez and Lobato (2004) are known to be* $\sqrt{n}$*-consistent under the null hypothesis. Let us consider the estimator described by Dominguez and Lobato (2004). The pseudo-true value* θ_m^* *of the estimator is defined as the minimizer* $\theta \in \Theta$ *of* $\int E|[m(Z) - g(X,\theta)]\mathbb{1}\{Z \leq z\}|^2 f(z)dx$. *The first-order condition of the minimization problem is* $\int E\left\{[m(Z) - g(X,\theta_m^*)]\mathbb{1}\{Z \leq z\}\right\} H_{\theta_m^*}(z) f(z)dx = 0$, *where we define* $H_\theta(z) \equiv E[\frac{\partial}{\partial\theta} g(X,\theta)\mathbb{1}\{Z \leq z\}]$. *Applying the mean value theorem yields* $m(Z) - g(X,\theta_m^*) = m(Z) - g(X,\theta_0) - \frac{\partial}{\partial\theta'} g(X,\tilde{\theta})(\theta_m^* - \theta_0)$, *where* $\tilde{\theta}$ *is the segment between* θ_m^* *and* θ_0, *which implies*

$$(\theta_m^* - \theta_0) = \left[\int H_{\theta_m^*}(z) H_{\tilde{\theta}}(z)' f(z)dx\right]^{-1} \int E[\delta_{\theta_0}(Z)\mathbb{1}\{Z \leq z\}] H_{\theta_m^*}(z) f(z)dx. \tag{3}$$

Thus, given the existence of the inverse of $E[H_{\theta_m^*}(Z) H_{\tilde{\theta}}(Z)']$, *we obtain, for some constant* $c > 0$, $\|\theta_m^* - \theta_0\| \leq c \int |E[\delta_{\theta_0}(Z) h(Z,z)]| f(z)dx$, *where* $h(Z,z) = \mathbb{1}\{Z \leq z\}$, *since* $H_{\theta_m^*}(z)$ *is bounded by Assumption 5.*

The following Theorem shows the optimal minimax rate of specification testing for IV regression models.

Theorem 1. *(Optimal Minimax Rate) Suppose Assumptions from 1 to 11 hold. If* $s + k \geq l_z/4$, *the optimal minimax rate against* $H_{n,1}$ *is* $n^{-2(s+k)/[l_z+4(s+k)]}$.

To prove the optimal minimax rate, we first show that no test has more than trivial uniform power against $\mathcal{M}(\tilde{\rho}_n)$ for any $\tilde{\rho}_n$ that approaches zero faster than $n^{-2(s+k)/[l_z+4(s+k)]}$. This is called the lower bound. Then, we modify the test proposed by Guerre and Lavergne (2002) for IV regression models and show that the modified test has non-trivial uniform power against $\mathcal{M}(\rho_n)$, where $\rho_n = n^{-2(s+k)/[l_z+4(s+k)]}$. The proof is given in Appendix A.

Theorem 1 shows that the optimal minimax rate $n^{-2(s+k)/[l_z+4(s+k)]}$ depends on the dimension of instruments and the smoothness of the set of alternatives. The rate implies that the uniform power of tests reduces when the dimension of instruments is large.

Theorem 1 considers the case of smooth alternatives ($s + k \geq l_z/4$). When $s + k < l_z/4$, the lower bound is $n^{-1/4}$, as shown in Appendix A. However, the optimal minimax rate is unknown because no specification test is shown to have non-trivial uniform power against such irregular non-smooth alternatives when evaluated with $n^{-1/4}$. Guerre and Lavergne (2002) argued that, against such irregular alternatives, the optimal minimax rate may differ from $n^{-1/4}$ and may depend on the smoothness of alternative classes. Hitomi et al. (2020) showed the set of non-smooth functions against which the optimal minimax rate is $n^{-1/4}$. Their non-smooth alternative consists of bounded functions, and no smoothness restrictions are imposed on those derivatives.

4 Smoothing-type Test

We adapt the test proposed by Zheng (1996) for IV regression models. The test is based on the sample analogue of $E[uE(u|Z)f(Z)]$. We define

$$T_n = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right) \hat{u}_i \hat{u}_j,$$

where $K(\cdot)$ is a product kernel function (e.g., a Gaussian kernel) that satisfies the assumption below and h is the smoothing parameter (bandwidth).

Assumption 12. *We have a q_kth-order symmetric kernel $k(\cdot)$ with $q_k \geq 2$ that satisfies $\int k(u)du = 1$, $\int |k(u)|du < \infty$, $\sup_u |k(u)| < \infty$, and $|uk(u)| \to 0$ if $u \to \infty$. The product kernel is denoted by $K(\cdot) = k(\cdot)k(\cdot)\cdots k(\cdot)$.*

The asymptotic normality of the test statistic $nh^{l_z/2}T_n$ under H_0 is shown in Theorem 1 of Zheng (1996) under the regression set up. This result can be extended to the IV regression set up by making minor modifications to the proof. We restate the asymptotic normality results under the current set up as follows:

Proposition 1. *(Asymptotic Normality) Suppose Assumptions 1, 2, 3, 5, 6, 10, and 12, hold. If $h \to 0$ and $nh^{l_z} \to \infty$, under the null hypothesis, $nh^{l_z/2}T_n$ converges weakly to $N(0, \Sigma)$, where $\Sigma \equiv 2\int K(u)^2du \int [\sigma^2(z)]^2 f(z)^2dz$. The asymptotic variance Σ can be consistently estimated by*

$$\hat{\Sigma} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 \hat{u}_i^2 \hat{u}_j^2.$$

The test is one-sided. The null hypothesis is rejected when $\hat{\Sigma}^{-1/2}nh^{l_z/2}T_n \geq z_\alpha$, where z_α is the $1-\alpha$ quantile of the standard normal distribution.

The null hypothesis will be rejected if there is misspecification when instruments are valid. To see this, decompose the test using $\hat{u}_i = [Y_i - g(X_i, \theta^*)] + [g(X_i, \theta^*) - g(X_i, \hat{\theta})]$. The term that includes $[Y_i - g(X_i, \theta^*)][Y_j - g(X_j, \theta^*)]$ converges to the normal distribution under the null and diverges under the alternative. The remaining terms include $g(X_i, \theta^*) - g(X_i, \hat{\theta})$, which is asymptotically negligible under Assumption 10 or 11 with differentiability of $g(X_i, \cdot)$. The existence of valid instruments is implicitly assumed. As long as instruments are valid, the source of power comes from the L^2-distance between $m(Z_i)$ and $E[g(X_i, \theta^*)|Z_i]$ in the first term. When instruments are invalid, however, biased parameter estimates contaminate the source of power. In this case, the rejection of the null hypothesis may be caused by invalid instruments, misspecification, or both.

The following theorem shows that the test is rate optimal when $s + k > l_z/4$.

Theorem 2. *(Rate Optimality) Suppose Assumptions 1, 2, 3, 5, 7, 8, 11, and 12 hold. Let* $\rho_n = n^{-2(s+k)/[l_z+4(s+k)]}$, $s + k \geq l_z/4$, *and the bandwidth* h *satisfies* $nh^{l_z} \to \infty$, $h^{\min\{q_k, k+q_z\}/2} n^{\frac{s+k}{l_z+4(s+k)}} = O(1)$, *and* $h^{-1} n^{\frac{-2}{l_z+4(s+k)}} = O(1)$. *For any prescribed bound* $\beta \in (0, 1-\alpha)$, *a constant* κ *exists such that*

$$\sup_{m \in \mathcal{M}(\kappa\rho_n)} P(nh^{l_z/2}\hat{\Sigma}^{-1/2} T_n \leq z_\alpha) \leq \beta + O(1).$$

Theorem 2 shows the orders of bandwidths that ensure the rate optimality of the proposed test. Unfortunately, however, they do not disclose the value of h, and thus, in practice, the choice of the bandwidth may rely on data-driven methods. Note that data-driven bandwidths are random variables. It is not trivial whether Theorem 2 holds analogously for the test with a data-driven bandwidth $\hat{h}$.

Let $T_n(\hat{h})$ be a version of T_n, in which h is replaced by $\hat{h}$; that is,

$$T_n(\hat{h}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{\hat{h}^{l_z}} K\left(\frac{Z_j - Z_i}{\hat{h}}\right) \hat{u}_i \hat{u}_j. \tag{4}$$

In the same manner, let $\hat{\Sigma}(\hat{h})$ be a version of $\hat{\Sigma}$, in which h is replaced by $\hat{h}$.

The following proposition shows that the results of Theorem 3 hold analogously, even when h is replaced with a data-driven bandwidth $\hat{h}$. In the following proposition, the s-th derivative of the kernel k is denoted by $k^{(s)}$.

Theorem 3. *Suppose the Assumptions in Theorem 1, Proposition 1, and Theorem 2 hold. Let the kernel k be m-times differentiable. We assume that* $\tilde{K}^{(s)}\left(\frac{Z_j - Z_i}{h}\right) \equiv h^s \frac{\partial^s}{\partial h^s} K\left(\frac{Z_j - Z_i}{h}\right)$ *satisfies* $\int |\tilde{K}^{(s)}(u)| du < \infty$, $\sup_u |\tilde{K}^{(s)}(u)| < \infty$, *and* $|u\tilde{K}^{(s)}(u)| \to 0$ *if* $u \to \infty$ *for all* $s = 1, \ldots, m$. *In addition, suppose that* $\hat{h} = h_0 + o_p(h_0)$ *for some deterministic sequence* h_0 *that converges to zero and* $\hat{h}^{-l_z}(\hat{h}/h_0 - 1)^m = o_p(1)$. *Then, the test with a data-driven bandwidth* $\hat{\Sigma}(\hat{h})^{-1/2} n \hat{h}^{l_z/2} T_n(\hat{h})$ *is rate optimal when* h_0 *satisfies* $h_0^{\min\{q_k, k+q_z\}/2} n^{\frac{s+k}{l_z+4(s+k)}} = O(1)$ *and* $h_0^{-1} n^{\frac{-2}{l_z+4(s+k)}} = O(1)$.

We show that the bandwidths selected by the least squares cross-validation can satisfy the conditions in Theorem 3. This method is one of the most widely used selection methods, in which one selects h that minimizes

$$CV(h) = \sum_{i=1}^{n} [Y_i - \hat{m}_{-i}(Z_i)]^2 w(Z_i), \tag{5}$$

where $\hat{m}_{-i}(Z_i) = \sum_{j \neq i} K\left(\frac{Z_j - Z_i}{h}\right) Y_j / \sum_{j \neq i} K\left(\frac{Z_j - Z_i}{h}\right)$ is the leave-one-out kernel estimator of $m(Z_i)$, and $0 \leq w(\cdot) \leq 1$ is a weight function. Let h_{cv} denote the value of h selected by cross-validation. It is well known that a unique, positive, and finite sequence h_0 exists such that $h_{cv} = h_0 + o_p(h_0)$, where $h_0 = O(n^{-1/(l_z+4)})$

(see Theorem 2.3 of Q. Li & Racine, 2007) and $h_{cv}/h_0 - 1 = O_p(n^{-\min\{l_z/2,2\}/(4+l_z)})$ (see Theorem 2.2 of Racine & Li, 2004). Then, the following corollary holds.

Corollary 1. *For a sufficiently smooth kernel such that $l_z < m\min\{l_z/2, 2\}$ and $k \geq 2$, the bandwidth chosen by cross-validation satisfies $h_{cv}^{-l_z}(h_{cv}/h_0 - 1)^m = o_p(1)$. Moreover, the test statistic evaluated with the bandwidth chosen by the cross-validation method $\hat{\Sigma}(h_{cv})^{-1/2} n\hat{h}^{l_z/2} T_n(h_{cv})$ is rate optimal when $4(s+k) \leq l_z + 8$ and $\min\{q_k, k+q_z\}[l_z + 4(s+k)] \geq 2(l_z+4)(s+k)$.*

Corollary 1 shows that the test evaluated with h_{cv} has rate optimal uniform power under these two conditions. Let us consider a higher-order kernel such that $\min\{q_k, k+q_z\} = k + q_z$. Then, the second condition holds when the density of Z_i is sufficiently smooth, such that $q_z \geq 3-k$ for $l_z = \{1, 2\}$, $q_z \geq 4-k$ for $l_z = \{3, 4\}$, and so on. When the first condition is satisfied, the second condition is satisfied for any l_z, when $q_z \geq 10$ (see Appendix for the derivation of the sufficient condition). The first condition $4(s+k) \leq l_z + 8$ implies that rate optimality is achieved only against the set of alternatives that are not too smooth. Intuitively, this condition arises because the optimal minimax rates depend on the smoothness of alternatives (whereby the rate is faster for smoother alternatives), while the convergence rate of $h_{cv} = h_0 + o_p(n^{-1/(l_z+4)})$ does not depend on this smoothness. This condition substantially restricts the cases in which $T_n(h_{cv})$ is rate optimal. In practice, however, bandwidths selected by the cross-validation method can perform well in terms of size and power, as shown in the simulation study below.

5 Simulation

The simulation aims to investigate and compare the size and power of several tests. We adapt the simulation set up of Horowitz (2006) so that the results are comparable with existing specification tests for instrumental variable regression models, including ICM-type tests (Bierens, 1982 and Bierens & Ploberger, 1997) and Horowitz (2006) and the exponential tilting test of Donald et al. (2003).

We test the null hypotheses that

$$g(x) = \beta_0 + \beta_1 x, \tag{6}$$

and

$$g(x) = \beta_0 + \beta_1 x + \beta_2 x^2. \tag{7}$$

The true models are (7) if (6) is H_0 and

$$g(x) = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3, \tag{8}$$

if (6) or (7) is H_0.

Data are generated by

$$\begin{aligned} X &= \Phi\left(\rho v_1 + (1-\rho^2)^{1/2} v_2\right), \\ Z &= \Phi(v_1), \\ u &= 0.2\Phi\left(\eta v_2 + (1-\eta^2)^{1/2} v_3\right), \end{aligned}$$

where $\Phi(\cdot)$ is the standard normal distribution function. v_1, v_2, and v_3 are drawn randomly from $\mathrm{N}(0,1)$.

Outcomes are generated by $y = g(x) + u$. In all experiments, $\beta_0 = 0$ and $\beta_1 = 0.5$. When (7) is the correct model, $\beta_2 = -0.5$. When (8) is the correct model, $\beta_2 = -1$, $\beta_3 = 1$ if (6) is H_0, and $\beta_3 = 2$ if (7) is H_0.

There are two parameters, ρ and η, for which the values vary among experiments. The parameter ρ balances the strengths of endogeneity and instrumental relevance. η modulates the exogenous component in u_i. We consider three sets of data generating processes (DGPs), called DGP 1, DGP 2, and DGP 3: DGP 1: $\rho = 0.8$ and $\eta = 0.1$; DGP 2: $\rho = 0.8$ and $\eta = 0.5$; DGP 3: $\rho = 0.7$ and $\eta = 0.1$.

In this experiment, X is endogenous and is instrumented by Z. The instruments to estimate (6) and (7) are $(1, Z)$ and $(1, Z, Z^2)$, respectively.

The kernel is Gaussian $k(v) = (2\pi)^{-1/2}\exp(-v^2/2)$. Bandwidths are selected by the least squares cross-validation, denoted by h_{cv}. We also report results obtained using the optimal bandwidth, denoted by h_{opt}, that minimizes the leading term of the cross-validation objective function.[4] Note that using the optimal bandwidth is infeasible in practice.

Critical values are obtained based on either the standard normal distribution or using the empirical distribution from $B = 1000$ simulation runs, where the test statistic in each simulation is computed using bootstrap observations, as per Gao and Gijbels (2008).[5] The sample size is $n = 500$ and the nominal level is 0.05. Size and power are obtained by $M = 1000$ simulation runs in each experiment.

[4]Although we know the DGP, the explicit form of the true IV regression function is not straightforward. Thus, optimal bandwidths are calculated using a random sample of size 150000.

[5]A bootstrap sample is $\{X_i, Z_i, Y_i^b\}_{i=1}^n$, where Y_i^b is generated by $Y_i^b = \hat{Y}_i + \hat{\sigma}_u^2 e_i^*$, in which $\hat{Y}_i$ are predicted values, $\hat{\sigma}_u$ is the residual standard error from the IV estimator under the null hypothesis, and $\{e_i^*\}_{i=1}^n$ is a sequence of random samples drawn from the standard normal distribution.

[Table 1 near here.]

Table 1 shows the simulation results. When H_0 is true, the test tends to under-reject the null hypothesis when critical values are obtained from the standard normal distribution. The under-rejections are severe for all set-ups, which may come from the well-known results that asymptotic approximations of IV estimators for linear models can be poor. In contrast, the test with h_{cv} tends to over-reject the null hypothesis when critical values are obtained by bootstrapping. The size distortions reduce when the sample size is increased to $n = 1000$, as shown in the supplemental material. The size is around the nominal level when optimal bandwidths are employed.

The power of T_n with h_{cv} is close to 1 when the null is (6). The power of testing (7) against (8) is remarkably low when DGP 3 is applied. Table 1 of Horowitz (2006) shows that the power of existing tests is low when DGP 3 is employed for all cases. Since the powers of T_n is close to 1, even under DGP 3, when the null is (6), the kernel-type test can be considered to complement other existing tests.

6 Conclusion

This study shows that the optimal minimax rate for linear and non-linear IV regression models is $n^{-2(s+k)/[l_z+4(s+k)]}$ when $s+k \geq l_z/4$, implying that rate optimal results in Guerre and Lavergne (2002) hold for more general IV regression frameworks, including linear models. The test $nh^{l_z/2}\hat{\Sigma}^{-1/2}T_n$ based on non-parametric kernel techniques is rate optimal when a deterministic sequence of bandwidths satisfy $h^{\min\{q_k,k+q_z\}/2}n^{\frac{s+k}{l_z+4(s+k)}} = O(1)$ and $h^{-1}n^{\frac{-2}{l_z+4(s+k)}} = O(1)$. Moreover, if the test is evaluated with a data-driven bandwidth $\hat{h}$ that can be described by

$\hat{h} = h_0 + o_p(h_0)$ for some deterministic sequence h_0, it is also rate optimal when h_0 satisfies the conditions above. Commonly applied bandwidth selection procedures such as the least squares cross-validation method can satisfy these conditions. A simulation study further validates that the proposed test can complement existing tests.

A possible future research direction is to consider the optimal minimax rate for specification testing against non-smooth alternatives ($s + k \leq l_z/4$). Against such alternatives, Guerre and Lavergne (2002) showed that the optimal minimax rate is $n^{-1/4}$ if the structure of the error variance conditional on regressors is known. Without this additional structure, however, it is unknown if any test exists that has non-trivial uniform power against non-smooth alternatives. Using a different set of non-smooth alternatives, Hitomi et al. (2020) showed that the optimal minimax rate is $n^{-1/4}$, and a test based on the difference between the non-parametric and parametric variance estimators is rate optimal, even when the structure of the error variance is unknown. However, research in this area is limited.

Additionally, the task of developing bandwidth selection procedures that maximize the uniform power of specification testing is left for future research. The power-maximizing selection procedure of Gao and Gijbels (2008) is based on a sequence of local alternatives that approach the null model as the sample size increases. A selection procedure based on maximizing the uniform power of testing is unknown.

Acknowledgments: This work was supported by JSPS KAKENHI Grant Numbers 17K03656, 19H01473, 19K23186, 20K01589, and Joint Usage and Research Project of Institute of Economic Research, Kyoto University. We would like to thank Yoichi Arai, Songnian Chen, Andrew Chesher, Jesus Gonzalo, Emmanuel Guerre, Hidehiko Ichimura, Hiroyuki Kasahara, Kengo Kato, Shakeeb Khan, Toru Kitagawa, Myoung-Jae Lee, Arthur Lewbel, Qingfeng Liu, Vadim Marmer, Yasumasa Matsuda, Tomoya Matsumoto, Ryo Okui, Peter Robinson, Naoya Sueishi, Shinya Tanaka, Takuya Ura, Yoshihiro Yajima, and the participants of the several meetings and conferences for their useful comments.

References

Aït-Sahalia, Y., Bickel, P. J., & Stoker, T. M. (2001). Goodness-of-fit tests for kernel regression with an application to option implied volatilities. *Journal of Econometrics*, *105(2)*, 363–412.

Bierens, H. J. (1982). Consistent model specification tests. *Journal of Econometrics*, *20(1)*, 105–134.

Bierens, H. J., & Ploberger, W. (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica*, *65(5)*, 1129–1151.

Carrasco, M., & Florens, J.-P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory*, *16(6)*, 797–834.

Dominguez, M. A., & Lobato, I. N. (2004). Consistent estimation of models defined by conditional moment restrictions. *Econometrica*, *72(5)*, 1601–1615.

Donald, S. G., Imbens, G. W., & Newey, W. K. (2003). Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, *117(1)*, 55–93.

Gao, J., & Gijbels, I. (2008). Bandwidth selection in nonparametric kernel testing. *Journal of the American Statistical Association*, *103*(484), 1584–1594.

González-Manteiga, W., & Crujeiras, R. M. (2013). An updated review of goodness-of-fit tests for regression models. *Test*, *22*, 361–411.

Gørgens, T., & Würtz, A. (2012). Testing a parametric function against a non-parametric alternative in IV and GMM settings. *Econometrics Journal*, *15(3)*, 462–489.

Guerre, E., & Lavergne, P. (2002). Optimal minimax rates for nonparametric specification testing in regression models. *Econometric Theory*, *18(5)*, 1139–1171.

Hall, A. R., & Inoue, A. (2003). The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics*, *114(2)*, 361–394.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, *50(4)*, 1029–1054.

Härdle, W., & Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistics*, *21(4)*, 1926–1947.

Hitomi, K., Iwasawa, M., & Nishiyama, Y. (2020). Optimal minimax rates against non-smooth alternatives. *KIER Discussion Paper Series No.1051*.

Holzmann, H. (2008). Testing parametric models in the presence of instrumental variables. *Statistics and Probability Letters*, *78(6)*, 629–636.

Horowitz, J. L. (2006). Testing a parametric model against a nonparametric alternative with identification through instrumental variables. *Econometrica*, *74(2)*, 521–538.

Horowitz, J. L., & Spokoiny, V. G. (2001). An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative. *Econometrica*, *69(3)*, 599–631.

Ingster, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Mathematical Methods of Statistics*, *2(2)*, 85–114.

Li, H., Li, Q., & Liu, R. (2016). Consistent model specification tests based on k-nearest-neighbor estimation method. *Journal of Econometrics*, *194(1)*, 187–202.

Li, Q., & Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice.* Princeton University Press.

Racine, J., & Li, Q. (2004). Nonparametric estimation of regression functions with both categorical and continuous data. *Journal of Econometrics*, *119*(1), 99–130.

Tripathi, G., & Kitamura, Y. (2003). Testing conditional moment restrictions. *Annals of Statistics*, *31(6)*, 2059–2095.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, *75(2)*, 263–289.

APPENDIX A: Proofs

A-1 Proof of Theorem 1

To prove Theorem 1, three propositions that complete the proof are given below. Proposition A1 shows the lower bound. We modify the test proposed by Guerre and Lavergne (2002) for IV regression models, denoted by T_n^{GL}. Asymptotic normality of T_n^{GL} under H_0 is given in Proposition A2. Finally, Proposition A3 shows that T_n^{GL} has non-trivial uniform power against $H_{n,1}$ evaluated with $n^{-2(s+k)/[l_z+4(s+k)]}$. Proofs of the propositions are given in the supplemental material.

Proposition A1. (Lower Bound) Suppose Assumptions 1, 2, 3, 5, 8, and 9 hold. Let $\tilde{\rho}_n = n^{-2(s+k)/[l_z+4(s+k)]}$ if $s+k \geq l_z/4$, $\tilde{\rho}_n = n^{-1/4}$ if $s+k < l_z/4$. If each ω_i is $N(0,1)$ conditionally upon Z_i, for any test t_n with $\sup_{m\in H_0} P(t_n > z_\alpha) \leq \alpha + o(1)$,

$$\sup_{m\in\mathcal{M}(\rho_n)} P(t_n \leq z_\alpha) \geq 1-\alpha+o(1), \quad \text{whenever} \quad \rho_n = o(\tilde{\rho}_n).$$

Let $I_k = \prod_{j=1}^{l_z}[k_j h_n, (k_j+1)h_n)$ be dyadic cubes that partition the support of instruments Z_i into $K_n^{l_z}$ cubes, where K_n is an integer, $h_n \equiv 1/K_n$ is the bandwidth that determines the number of cubes, and the index $k = (k_1, \ldots, k_{l_z})' \in \mathcal{K} \subset \mathbb{N}^{l_z}$ satisfies $0 \leq k_j \leq K_n - 1$ for $j = 1, \ldots, l_z$.

Following Guerre and Lavergne (2002), a test statistic is based on the average of the estimated parametric residuals $\hat{u}_i$ in each cube:

$$T_n^{GL} = \frac{1}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \hat{u}_i\hat{u}_j,$$

where $N_k \equiv \sum_{i=1}^n \mathbb{1}\{Z_i \in I_k\}$ is the number of observations of instruments in I_k. The estimator of the variance of T_n^{GL} is

$$\hat{v}_n^2 = \frac{1}{K_n^{l_z}} \sum_{k \in \mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i, Z_j\} \in I_k, i \neq j} \hat{u}_i^2 \hat{u}_j^2,$$

where the second summation is taken over i and $j \neq i$ that satisfies $\{Z_i, Z_j\} \in I_k$, that is, $\sum_{Z_i \in I_k} X_i = \sum_{i=1}^n \mathbb{1}\{Z_i \in I_k\} X_i$.

Proposition A2. (Asymptotic Normality) Suppose Assumptions 1, 2, 3, 5, 6, and 10, hold, $K_n \to \infty$, and $n/(K_n^{l_z} \log K_n^{l_z}) \to \infty$. Under the null hypothesis, the test $T_n^{GL}/\hat{v}_n$ converges to $N(0,1)$ weakly.

Proposition A3. (Rate Optimality of T_n^{GL}) Suppose Assumptions 1, 2, 3, 4, 5, 6, 7, 8, and 11 hold. Let $\rho_n = n^{-2(s+k)/[l_z+4(s+k)]}$, $s + k \geq l_z/4$, and $K_n = h_n^{-1} = (\lambda \rho_n^{1/(s+k)})^{-1}$ for some constant $\lambda > 0$. For any prescribed bound $\beta \in (0, 1-\alpha)$, a constant κ exists such that

$$\sup_{m \in \mathcal{M}(\kappa \rho_n)} P(\hat{v}_n^{-1} T_n^{GL} \leq z_\alpha) \leq \beta + o(1).$$

A-2 Proof of Proposition 1

Proof of Proposition 1. Using $\hat{u}_i = Y_i - g(X_i, \hat{\theta}) = g(X_i, \theta) - g(X_i, \hat{\theta}) + u_i$, the test statistic can be decomposed as follows:

$$\begin{aligned} & nh^{l_z/2} T_n \\ &= \frac{nh^{l_z/2}}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right) u_i u_j \end{aligned}$$

$$+ 2\frac{nh^{l_z/2}}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{h^{l_z}}K\left(\frac{Z_j - Z_i}{h}\right)u_i[g(X_j,\theta_0) - g(X_j,\hat{\theta})]$$

$$+ \frac{nh^{l_z/2}}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{h^{l_z}}K\left(\frac{Z_j - Z_i}{h}\right)[g(X_i,\theta_0) - g(X_i,\hat{\theta})][g(X_j,\theta_0) - g(X_j,\hat{\theta})]$$

$$\equiv T_1 + T_2 + T_3.$$

Under the null hypothesis, T_1 converges to the normal distribution, $T_2 = o_p(1)$, and $T_3 = o_p(1)$, as shown in Lemmas 1, 2, and 3, respectively. The proof for the asymptotic normality of T_1 is consistent with that for Lemma 3.3a of Zheng (1996). T_2 and T_3 include both covariates and instruments, which make the proof different from that for the regression set-up in Zheng (1996). Proofs are given in the supplemental material.

Lemma 1. *Under Assumptions 1, 2, and 12,* $T_1 \xrightarrow{d} N(0, 2\mathcal{K}(0)E\{[\sigma^2(Z)]^2 f(Z)\})$, *where* $\mathcal{K}(0)$ *denotes the convolution product.*

Lemma 2. *Under Assumptions 1, 2, 3, 6, 10, and 12, we have* $T_2 = o_p(1)$.

Lemma 3. *Under Assumptions 1, 2, 3, 5, 10, and 12, we have* $T_3 = o_p(1)$.

Lemma 4 shows that $\hat{\Sigma}$ is a consistent estimator for Σ under the null hypothesis. A proof of Lemma 4 is given in the supplemental material.

Lemma 4. *Under Assumptions 1, 2, 3, 5, 10, and 12, we have* $\hat{\Sigma} = \Sigma + o_p(1)$.

□

A-3 Proof of Theorem 2

Proof of Theorem 2. Proofs of all lemmas used in this proof are given in the supplemental material. Under $H_{n,1}$, we have

$$
\begin{aligned}
&nh^{l_z/2}T_n \\
&= \frac{nh^{l_z/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right) [Y_i - g(X_i, \theta^*)][Y_j - g(X_j, \theta^*)] \\
&+ \frac{2nh^{l_z/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right) [Y_i - g(X_i, \theta^*)][g(X_j, \theta^*) - g(X_j, \hat{\theta})] \\
&+ \frac{nh^{l_z/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right) [g(X_i, \theta^*) - g(X_i, \hat{\theta})][g(X_j, \theta^*) - g(X_j, \hat{\theta})] \\
&\equiv A_1 + A_2 + A_3. \qquad \text{(A.1)}
\end{aligned}
$$

The convergence of $(\hat{\theta} - \theta^*) = O_p(n^{-1/2})$ in Assumption 11 and other assumptions leads to $A_3 = o_p(1)$. The following lemma holds for A_2.

Lemma 5. *Suppose Assumptions 2, 3, 5, 8, 11, and 12 hold. Then,* $A_2 + A_3 = O_p(\sqrt{nh^{l_z}})\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2} + O_p(1)$.

The probability limit of $\hat{\Sigma}$ under $H_{n,1}$ can be shown as follows.

Lemma 6. *Suppose Assumptions 1, 2, 3, 5, 7, 11, and 12 hold. Let* $\sigma_{\theta^*}^2(Z_i) \equiv E(u_i^{*2}|Z_i)$, *where* $u_i^* = Y_i - g(X_i, \theta^*)$. *Then, under* $H_{n,1}$, *we obtain* $\hat{\Sigma} = \bar{\Sigma} + o_p(1)$, *where* $\bar{\Sigma} = 2\int K(u)^2 du E\{[\sigma_{\theta^*}^2(Z_i)]^2 f(Z_i)\}$ *is uniformly bounded in* $m \in \mathcal{M}(\kappa\rho_n)$.

These results imply, for arbitrary small ϵ, a constant $C > 0$ and z'_α exist such that

$$\sup_{m\in\mathcal{M}(\kappa\rho_n)} P(nh^{l_z/2}\overline{\Sigma}^{-1/2}T_n \le z_\alpha)$$
$$= \sup_{m\in\mathcal{M}(\kappa\rho_n)} P(A_1 \le \overline{\Sigma}^{1/2} z_\alpha - A_2 - A_3)$$
$$\le \sup_{m\in\mathcal{M}(\kappa\rho_n)} P(A_1 \le z'_\alpha + \sqrt{nh^{l_z}}\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2}C) + \epsilon.$$

When $E(A_1) - z'_\alpha - \sqrt{nh^{l_z}}\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2}C > 0$, Chebyshev's inequality yields

$$P(A_1 \le z'_\alpha + \sqrt{nh^{l_z}}\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2}C) \le \frac{var(A_1)}{[E(A_1) - z'_\alpha - \sqrt{nh^{l_z}}\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2}C]^2}. \tag{A.2}$$

Thus, it suffices to show that the following inequalities hold uniformly in $m \in \mathcal{M}(\kappa\rho_n)$.

$$E(A_1) - z'_\alpha - \sqrt{nh^{l_z}}\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2}C > 0, \tag{A.3}$$
$$\frac{var(A_1)}{[E(A_1) - z'_\alpha - \sqrt{nh^{l_z}}\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2}C]^2} \le \beta. \tag{A.4}$$

First, we show (A.3). To this end, we decompose A_1 as follows:

$$A_1 = \frac{nh^{l_z/2}}{n(n-1)}\sum_{i=1}^n\sum_{j\ne i}\frac{1}{h^{l_z}}K\left(\frac{Z_j-Z_i}{h}\right)[m(Z_i)-g(X_i,\theta^*)][m(Z_j)-g(X_j,\theta^*)]$$
$$+ \frac{nh^{l_z/2}}{n(n-1)}\sum_{i=1}^n\sum_{j\ne i}\frac{1}{h^{l_z}}K\left(\frac{Z_j-Z_i}{h}\right)[m(Z_i)-g(X_i,\theta^*)]\omega_j$$
$$+ \frac{nh^{l_z/2}}{n(n-1)}\sum_{i=1}^n\sum_{j\ne i}\frac{1}{h^{l_z}}K\left(\frac{Z_j-Z_i}{h}\right)[m(Z_j)-g(X_j,\theta^*)]\omega_i$$

$$+\frac{nh^{l_z/2}}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{h^{l_z}}K\left(\frac{Z_j-Z_i}{h}\right)\omega_i\omega_j$$

$$\equiv A_{1,1}+A_{1,2}+A_{1,3}+A_{1,4}.$$

Obviously, we have $E(A_{1,2})=E(A_{1,3})=E(A_{1,4})=0$. Let $q\equiv\min\{q_k,k+q_z\}$. A change of variables under Assumptions 2, 8, and 12 yields[6]

$$\begin{aligned}E(A_{1,1})&=nh^{-l_z/2}E\left[K\left(\frac{Z_j-Z_i}{h}\right)\delta_{\theta^*}(Z_i)\delta_{\theta^*}(Z_j)\right]\\&=nh^{l_z/2}E[\delta_{\theta^*}(Z_i)^2f(Z_i)]+O(nh^{l_z/2+q})E[\delta_{\theta^*}(Z_i)].\end{aligned}\tag{A.5}$$

Then, using the fact that $E[\delta^2_{\theta^*}(Z_i)]\geq\inf_{\theta\in\Theta}E[\delta^2_{\theta}(Z_i)]\geq\rho_n^2$, we obtain

$$\begin{aligned}&\frac{E(A_1)-z'_\alpha-\sqrt{nh^{l_z}}\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2}C}{nh^{l_z/2}E[\delta_{\theta^*}(Z_i)^2]}\\&\geq\frac{nh^{l_z/2}E[\delta_{\theta^*}(Z_i)^2]\underline{f}+O(nh^{l_z/2+q})E[\delta_{\theta^*}(Z_i)]-z'_\alpha-\sqrt{nh^{l_z}}\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2}C}{nh^{l_z/2}E[\delta_{\theta^*}(Z_i)^2]}\\&\geq\underline{f}-\frac{1}{\kappa}O\left(h^qn^{\frac{2(s+k)}{l_z+4(s+k)}}\right)-\frac{1}{\kappa^2}O\left(h^{-l_z/2}n^{\frac{-l_z}{l_z+4(s+k)}}\right)-o(1),\end{aligned}$$

where $\rho_n^2=n^{\frac{-4(s+k)}{l_z+4(s+k)}}$. When we chose h that satisfies both $h^{q/2}n^{\frac{(s+k)}{l_z+4(s+k)}}=O(1)$ and $h^{-1}n^{\frac{-2}{l_z+4(s+k)}}=O(1)$, the lower bound is increasing in κ and positive when κ and n are large enough, which implies equation (A.3).

Next, we show equation (A.4). A_1 is a second-order U-statistic:

$$\begin{aligned}A_1&=nh^{-l_z/2}\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j<i}K\left(\frac{Z_j-Z_i}{h}\right)[Y_i-g(X_i,\theta^*)][Y_j-g(X_j,\theta^*)]\\&\equiv nh^{-l_z/2}\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j<i}H_n(W_i,W_j),\end{aligned}$$

[6]For the derivation, see Lemma ?? in the supplemental material.

where $W_i = \{Y_i, X_i, Z_i\}$. Let us define $g(W_i) = Q(W_i) - Q$, where $Q(W_i) \equiv E[H_n(W_i, W_j)|W_i]$ and $Q \equiv E[H_n(W_i, W_j)]$ and $\eta(W_i, W_j) = H_n(W_i, W_j) - g(W_i) - g(W_j) - Q$, where $E[g(W_i)] = 0$, $E[\eta(W_i, W_j)] = 0$, and $E[\eta(W_i, W_j)\eta(W_i, W_k)] = 0$ for $j \neq k$. Then, the Hoeffding decomposition of second-order U-statistics and some calculations yield

$$\begin{aligned} var(A_1) &= n^2 h^{-l_z} var\left(\frac{2}{n}\sum_{i=1}^{n} g(W_i)\right) + n^2 h^{-l_z} var\left(\frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j<i}\eta(W_i, W_j)\right) \\ &= O(nh^{-l_z})E[g(W_i)^2] + O(h^{-l_z})\left(E[H_n(W_i, W_j)^2] - Q^2 - E[Q(W_i)^2]\right). \end{aligned} \tag{A.6}$$

Equation (A.5) implies for any $i \neq j$, $= O(h^{l_z})E[\delta_{\theta^*}(Z_i)^2] + O(h^{l_z+q})E[\delta_{\theta^*}(Z_i)]$. Since $Q(W_i) = E[H_n(W_i, W_j)|W_i] = [Y_i - g(X_i, \theta^*)]E\{K(\frac{Z_j - Z_i}{h})[Y_j - g(X_j, \theta^*)]|W_i\} = [Y_i - g(X_i, \theta^*)]E\{K(\frac{Z_j - Z_i}{h})\delta_{\theta^*}(Z_j)|W_i\}$, we have

$$\begin{aligned} E[Q(W_i)^2] &= E\left([Y_i - g(X_i, \theta^*)]^2\left\{\int K\left(\frac{z - Z_i}{h}\right)\delta_{\theta^*}(z)f(z)dz\right\}^2\right) \\ &= h^{2l_z} E\left(\sigma^2_{\theta^*}(Z_i)\left\{\delta_{\theta^*}(Z_i)f(Z_i) + O(h^q)\right\}^2\right) \\ &= O(h^{2l_z}E[\delta_{\theta^*}(Z_i)^2]) + O(h^{2l_z+2q}) + O(h^{2l_z+q})E[\delta_{\theta^*}(Z_i)]), \end{aligned}$$

where $\sigma^2_{\theta^*}(Z_i)$ is bounded almost surely by Assumptions 1 and 7 under $H_{n,1}$ Then,

$$\begin{aligned} E[g(W_i)^2] &= E[Q(W_i)^2 - 2Q(W_i)Q + Q^2] = E[Q(W_i)^2] - Q^2 \\ &\leq O(h^{2l_z}E[\delta_{\theta^*}(Z_i)^2]) + O(h^{2l_z+2q}) + O(h^{2l_z+q})E[\delta_{\theta^*}(Z_i)]). \end{aligned}$$

A change of variables under Assumptions 1, 2, 7, and 12 yields

$$E[H_n(W_1, W_2)^2] = \int K\left(\frac{z_2 - z_1}{h}\right)^2 \sigma_{\theta^*}^2(z_1)\sigma_{\theta^*}^2(z_2) f(z_1) f(z_2) dz_1 dz_2 = O(h^{l_z}).$$

These results together with equation (A.6) yield

$$\begin{aligned} var(A_1) \leq\ & O(nh^{l_z})E[\delta_{\theta^*}(Z_i)^2] + O(nh^{l_z+2q}) + O(nh^{l_z+q})E[\delta_{\theta^*}(Z_i)] + O(1) \\ & + O(h^{l_z})E[\delta_{\theta^*}(Z_i)^2]^2 + O(h^{l_z+2q})E[\delta_{\theta^*}(Z_i)]^2, \end{aligned}$$

which implies

$$\frac{var(A_1)}{n^2h^{l_z}E[\delta_{\theta^*}(Z_i)^2]^2} \leq \frac{1}{\kappa^4} O\left(n^{\frac{-l_z+4(s+k)}{l_z+4(s+k)}} h^{2q}\right) + \frac{1}{\kappa^4} O\left(n^{\frac{-2l_z}{l_z+4(s+k)}} h^{-l_z}\right).$$

The upper bound is a decreasing function of κ, when h is chosen such that $n^{\frac{-l_z/4+(s+k)}{l_z+4(s+k)}} h^{q/2} = O(1)$ and $n^{\frac{-2}{l_z+4(s+k)}} h^{-1} = O(1)$.

Therefore, equations (A.3) and (A.4) hold if the bandwidth value satisfies the following conditions:

$$n^{\frac{s+k}{l_z+4(s+k)}} h^{q/2} = O(1) \tag{A.7}$$

$$n^{\frac{-2}{l_z+4(s+k)}} h^{-1} = O(1). \tag{A.8}$$

The source of power, represented by the first term of the right hand side of equation (A.5), requires that $nh^{l_z/2}E[\delta_{\theta^*}(Z_i)^2] \approx nh^{l_z/2}\rho_n^2 = n^{l_z/[l_z+4(s+k)]}h^{l_z/2}$ does not shrink, which constrains the bandwidth to converge to zero at a rate slower than $n^{-2/[l_z+4(s+k)]}$. This requirement is reflected by condition (A.8). Thus, for example, we can choose the bandwidth that is $h = cn^{-2/[l_z+4(s+k)]}$ for some constant

$c > 0$, which satisfies condition (A.8). This choice of bandwidth also satisfies condition (A.7) when $n^{\frac{(s+k)-q}{l_z+4(s+k)}} = n^{\frac{(s+k)-\min\{q_k, k+q_z\}}{l_z+4(s+k)}} = O(1)$, which holds when $(s+k) \leq \min\{q_k, k+q_z\}$, which is equivalent to $(s+k) \leq q_k$ since $s < q_z$.

□

A-4 Proof of Theorem 3

Proof. Proofs of all lemmas used in this proof are given in the supplemental material. Theorem 2 shows the rate optimality of $nh^{l_z/2}\hat{\Sigma}^{-1/2}T_n$, in which h is treated as a deterministic sequence. Since data driven bandwidths are random variables, it is not trivial whether Theorem 2 holds analogously. Thus, we first show the rate optimality of testing in which h is replaced by a data-driven bandwidth $\hat{h}$. To this end, we decompose the test statistic as follows.

Lemma 7. *Suppose Assumptions 1, 5, 7, 11, and 12 hold. Let $\hat{h}$ be data-driven bandwidth such that $\hat{h} = h_0 + o_p(h_0)$ for some deterministic sequence h_0 that converges to zero and $\hat{h}^{-l_z}(\hat{h}/h_0 - 1)^m = o_p(1)$. We assume that the kernel k be m-times differentiable and $\tilde{K}^{(s)}\left(\frac{Z_j - Z_i}{h}\right) \equiv h^s \frac{\partial^s}{\partial h^s} K\left(\frac{Z_j - Z_i}{h}\right)$ satisfies $\int |\tilde{K}^{(s)}(u)|du < \infty$, $\sup_u |\tilde{K}^{(s)}(u)| < \infty$, and $|u\tilde{K}^{(s)}(u)| \to 0$ if $u \to \infty$ for all $s = 1, \ldots, m$. Then, the test statistic can be decomposed as follows.*

$$T_n(\hat{h}) = \left(\frac{h_0}{\hat{h}}\right)^{l_z} T_n(h_0) + o_p\left(T_n(h_0)\right). \tag{A.9}$$

Lemma 7 and equation (A.1) imply that

$$n\hat{h}^{l_z/2}T_n(\hat{h}) = \tilde{A}_1 + \left(h_0/\hat{h}\right)^{l_z}(A_2 + A_3) + \tilde{A}, \tag{A.10}$$

where $\tilde{A}_1 = \left(h_0/\hat{h}\right)^{l_z} A_1 + o_p(1)A_1$ and $\tilde{A} = o_p(A_2 + A_3)$.

The following lemma shows asymptotic behavior of the variance of test statistic.

Lemma 8. *Suppose Assumptions 1, 2, 3, 5, 7, 11, and 12 hold let $\hat{h}$ be data-driven bandwidth such that $\hat{h} = h_0 + o_p(h_0)$ for some deterministic sequence h_0 that converges to zero and $\hat{h}^{-l_z}(\hat{h}/h_0 - 1)^m = o_p(1)$. We assume that the kernel k be m-times differentiable and $\tilde{K}^{(s)}\left(\frac{Z_j - Z_i}{h}\right) \equiv h^s \frac{\partial^s}{\partial h^s} K\left(\frac{Z_j - Z_i}{h}\right)$ satisfies $\int |\tilde{K}^{(s)}(u)|du < \infty$, $\sup_u |\tilde{K}^{(s)}(u)| < \infty$, and $|u\tilde{K}^{(s)}(u)| \to 0$ if $u \to \infty$ for all $s = 1, \ldots, m$. Then, we have*

$$\hat{\Sigma}(\hat{h}) = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{\hat{h}^{l_z}} K\left(\frac{Z_j - Z_i}{\hat{h}}\right)^2 \hat{u}_i^2 \hat{u}_j^2 = \bar{\Sigma} + o_p(1).$$

These results and $(h_0/\hat{h})^{l_z}(A_2 + A_3) = O_p(\sqrt{nh_0^{l_z}})\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2}$ by Lemma 5 imply that, for arbitrary small ϵ, a constant $C > 0$ and z'_α exist such that

$$\begin{aligned}
&\sup_{m \in \mathcal{M}(\kappa\rho_n)} P(n\hat{h}^{l_z/2}[\Sigma(\hat{h})]^{-1/2} T_n(\hat{h}) \leq z_\alpha) \\
&\quad = \sup_{m \in \mathcal{M}(\kappa\rho_n)} P(\tilde{A}_1 \leq [\Sigma(\hat{h})]^{1/2} z_\alpha - \left(h_0/\hat{h}\right)^{l_z} (A_2 + A_3) - \tilde{A}) \\
&\quad \leq \sup_{m \in \mathcal{M}(\kappa\rho_n)} P(\tilde{A}_1 \leq z'_\alpha + \sqrt{nh^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2} C) + \epsilon.
\end{aligned}$$

Since $\tilde{A}_1 = [(h_0/\hat{h})^{l_z} + o_p(1)]A_1 = [1 + o_p(1)]A_1$, for arbitrary small ϵ, a constant $c > 1$ exists such that

$$P(\tilde{A}_1 \leq z'_\alpha + \sqrt{nh_0^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2} C) \leq P(A_1 \leq c[z'_\alpha + \sqrt{nh_0^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2} C]) + \epsilon,$$

when n is large enough. The right hand side of the above equation is equivalent to the left hand side of equation (A.2) up to the constant c. Thus, the optimal minimax rate of the test with stochastic bandwidth $\hat{h}$ can be derived analogously to that given in Theorem 2, which implies that the test is rate optimal when h_0 satisfies $nh_0^{l_z} \to \infty$ and the following conditions:

$$h_0^{\min\{q_k, k+q_z\}/2} n^{\frac{s+k}{l_z+4(s+k)}} = O(1) \tag{A.11}$$

$$h_0^{-1} n^{\frac{-2}{l_z+4(s+k)}} = O(1). \tag{A.12}$$

□

A-5 Proof of Corollary 1

Proof. It is well known that a unique, positive, and finite sequence h_0 exists such that $h_{cv} = h_0 + o_p(h_0)$, where $h_0 = O(n^{-1/(l_z+4)})$ (see Theorem 2.3 of Q. Li & Racine, 2007) and $h_{cv}/h_0 - 1 = O_p(n^{-\min\{l_z/2,2\}/(4+l_z)})$ (see Theorem 2.2 of Racine & Li, 2004). Thus, $h_{cv}^{-l_z}(h_{cv}/h_0 - 1)^m = O_p\left(n^{\frac{l_z - m\min\{l_z/2,2\}}{l_z+4}}\right)$, which converges zero in probability when $l_z \leq m\min\{l_z/2, 2\}$.

Recall that $h_0 = O\left(n^{\frac{-1}{l_z+4}}\right)$. Then, first,

$$h_0^{-1} n^{\frac{-2}{l_z+4(s+k)}} = O\left(n^{\frac{-l_z+4(s+k-2)}{[l_z+4(s+k)](l_z+4)}}\right), \tag{A.13}$$

which is $O(1)$, if $4(s+k) \leq l_z + 8$. Second,

$$h_0^{\frac{\min\{q_k, k+q_z\}}{2}} n^{\frac{s+k}{l_z+4(s+k)}} = O\left(n^{\frac{-\min\{q_k, k+q_z\}[l_z+4(s+k)]/2+(l_z+4)(s+k)}{[l_z+4(s+k)](l_z+4)}}\right), \tag{A.14}$$

which is $O(1)$, if $\min\{q_k, k+q_z\}[l_z + 4(s+k)] \geq 2(l_z+4)(s+k)$.

Suppose that we use a higher-order kernel such that $\min\{q_k, k+q_z\} = k+q_z$. Then, (A.14) is $O(1)$ if $(k+q_z)[l_z+4(s+k)] \geq 2(l_z+4)(s+k)$, which holds if the density of Z_i is smooth enough (q_z is large enough).

For example, if $q_z \geq 10$, (A.14) is $O(1)$ for any l_z such that $l_z \leq 4(s+k) \leq l_z+8$. To see this, we replace $4(s+k)$ in the left hand side of $(k+q_z)[l_z+4(s+k)] \geq 2(l_z+4)(s+k)$ with l_z. This yields a sufficient condition, which is $(k+q_z)l_z \geq (l_z+4)(s+k)$. Simple calculation and using $4(s+k) \leq l_z+8$ yields $q_z \geq 8/l_z+s+1$, which holds as long as $q_z \geq 8/l_z+2$. Substituting $l_z = 1$ yields $q_z \geq 10$.

□

Table 1

Table 1: Size and power of T_n with $n = 500$.

					Bootstrap		Normal	
H_0	H_1		ρ	η	h_{cv}	h_{opt}	h_{cv}	h_{opt}
H_0 is true								
(6)		DGP 1	0.8	0.1	0.059	0.046	0.020	0.029
		DGP 2	0.8	0.5	0.062	0.042	0.025	0.029
		DGP 3	0.7	0.1	0.058	0.050	0.018	0.028
(7)		DGP 1	0.8	0.1	0.067	0.050	0.016	0.018
		DGP 2	0.8	0.5	0.058	0.040	0.015	0.023
		DGP 3	0.7	0.1	0.076	0.044	0.018	0.019
H_0 is false								
(6)	(7)	DGP 1	0.8	0.1	1.000	0.934	1.000	0.902
		DGP 2	0.8	0.5	1.000	0.905	0.999	0.861
		DGP 3	0.7	0.1	0.990	0.498	0.949	0.370
(6)	(8)	DGP 1	0.8	0.1	0.999	0.842	0.999	0.781
		DGP 2	0.8	0.5	0.999	0.797	0.998	0.726
		DGP 3	0.7	0.1	0.972	0.192	0.944	0.147
(7)	(8)	DGP 1	0.8	0.1	0.903	0.609	0.741	0.527
		DGP 2	0.8	0.5	0.886	0.506	0.717	0.437
		DGP 3	0.7	0.1	0.434	0.204	0.177	0.144

Note: Critical values are obtained from bootstrapping (columns labeled by Bootstrap) and the normal distribution (columns labeled by Normal).

Supplemental material for the paper entitled
"Optimal Minimax Rates of Specification Testing with Data-driven Bandwidth "

Supplemental Material: Proof of Propositions

S-1 Proof of Proposition A1

The proof of Proposition A1 goes along with that of Theorem 1 of Guerre and Lavergne (2002).

Proof of Proposition A1. Let $\phi(\cdot)$ be a map from $\mathbb{R}^{l_z}$ to $\mathbb{R}$ with support $[0,p]^{l_z}$ for $p > 0$ that is infinitely differentiable, $\phi(z) < \infty$ for any $z \in [0,p]^{l_z}$, $\int \phi(z)dz = 0$, and $\phi(\cdot) \in \mathcal{M}_{(L-L_{\mathcal{M}})/2,s,k}$. Guerre and Lavergne (2002) give an example that satisfies these conditions.

We define the dyadic cubes that partition $[0,1]^{l_z}$ into $K_n(p)^{l_z} \equiv [1/(ph_n)]^{l_z}$ non-overlapping cubes, where intersections of any two cubes are empty and $1/(ph_n)$ is assumed to be an integer.[7] To define the cubes, let $\mathcal{K}_n(p)$ denotes a collection of all possible distinct values for $\kappa \equiv (\kappa_1, \dots, \kappa_{l_z})'$ such that $\mathcal{K}_n(p) = \{\kappa \in \mathbb{Z}^{l_z} : 0 \leq \kappa_j \leq 1/(ph_n) - 1, j = 1, 2, \dots, l_z\}$, which indicates that $\mathcal{K}_n(p)$ contains $K_n(p)^{l_z}$ elements. For $\kappa \in \mathcal{K}_n(p)$, we define $I_{\kappa,p} = \prod_{j=1}^{l_z} [p\kappa_j h_n, p(\kappa_j + 1)h_n)$. Then, $\bigcup_{\kappa \in \mathcal{K}_n(p)} I_{\kappa,p} = [0,1]^{l_z}$ and $I_{\kappa,p} \cap I_{jp} = \emptyset$ for all $\kappa, j \in \mathcal{K}_n(p)$ when $\kappa \neq j$. The number of partitions are determined by h_n, and we define that $h_n = (\lambda \rho_n)^{1/(s+k)}$ for some constant $\lambda > 0$.

For $\kappa \in \mathcal{K}_n(p)$, let $\phi_\kappa(\cdot) : \mathbb{R}^{l_z} \to \mathbb{R}$ be a function such that $\phi_\kappa(z) = h_n^{-l_z/2} \phi\left(\frac{z - p\kappa h_n}{h_n}\right)$. Then, $\phi_\kappa(z)$ takes non-zero value only when $z \in I_{\kappa,p}$. Thus, the functions $\phi_\kappa(\cdot)$'s are orthogonal with disjoint supports $I_{\kappa,p}$, namely, $\phi_\kappa(z)\phi_{\kappa'}(z) = 0$ as long as $\kappa \neq \kappa'$. For

[7] When $1/(ph_n)$ is not an integer, define $K_n(p)$ to be the maximum integer smaller than $1/(ph_n)$.

any sequence $\{B_\kappa\}_{\kappa\in\mathcal{K}_n(p)}$ with $|B_\kappa| = 1$, we define

$$m_n(\cdot) = E[g(X,\theta_0)|\cdot] + \delta_n(\cdot), \qquad \delta_n(\cdot) = \lambda\rho_n h_n^{l_z/2}\sum_{\kappa\in\mathcal{K}_n(p)} B_\kappa\phi_\kappa(\cdot)$$

Let $\check{\theta}_{m_n}$ satisfies $\inf_{\theta\in\Theta} E[\delta_\theta(Z_i)^2] = E[\delta_{\check{\theta}_{m_n}}(Z_i)^2]$. We show that a positive constant C and a bounded function $h(\cdot)$ exist such that $\|\check{\theta}_{m_n} - \theta_0\| \le C|E[h(Z_i)\delta_n(Z_i)]|$, so that $\check{\theta}_{m_n}$ satisfies Assumption 11 (iii) for any n. Under Assumption 3, the definition of $\check{\theta}_{m_n}$ yields

$$\begin{aligned} 0 = E\left\{G_{\check{\theta}_{m_n}}[\delta_{\check{\theta}_{m_n}}(Z_i)]\right\} &= E\left(G_{\check{\theta}_{m_n}}\{m_n(Z_i) - E[g(X_i,\check{\theta}_{m_n})|Z_i]\}\right)\\ &= E\left(G_{\check{\theta}_{m_n}}\{\delta_n(Z_i) + E[g(X_i,\theta_0)|Z_i] - E[g(X_i,\check{\theta}_{m_n})|Z_i]\}\right), \end{aligned}$$

where $G_{\check{\theta}_{m_n}} = \frac{\partial}{\partial\theta}E[g(X_i,\theta)|Z_i]|_{\theta=\check{\theta}_{m_n}}$. Taylor expansion yields

$$\begin{aligned} E\left[G_{\check{\theta}_{m_n}}\delta_n(Z_i)\right] &= E\left(G_{\check{\theta}_{m_n}}\{E[g(X_i,\check{\theta}_{m_n})|Z_i] - E[g(X_i,\theta_0)|Z_i]\}\right)\\ &= E\left(G_{\check{\theta}_{m_n}}G'_{\tilde{\theta}_n}\right)(\check{\theta}_{m_n} - \theta_0) \end{aligned}$$

for some $\tilde{\theta}_n \in \Theta$. Since $\check{\theta}_{m_n} \to \theta_0$ as $n\to\infty$, the dominated convergence theorem under Assumptions 2, 3, and 9 yields $lim_{n\to\infty}E(G_{\check{\theta}_{m_n}}G'_{\tilde{\theta}_n}) = E(G_{\theta_0}G'_{\theta_0})$. We obtain

$$(\check{\theta}_{m_n} - \theta_0) = \left[E\left(G_{\theta_0}G'_{\theta_0}\right) + o(1)\right]^{-1} E\left[G_{\check{\theta}_{m_n}}\delta_n(Z)\right].$$

Since $E(G_{\theta_0}G'_{\theta_0})$ is invertible by Assumption 9, a constant C exists such that $\|\check{\theta}_{m_n} - \theta_0\| < C|E[G_{\check{\theta}_{m_n}}\delta_n(Z_i)]|$, where $G_{\check{\theta}_{m_n}}$ is bounded for any n by Assumptions 2 and 9.

Lemma S.9. *Under Assumptions 1, 2, 3, 5, 8, and 9, $E[m_n(Z_i)^4]$ is bounded and $m_n(Z_i)$ is in $\mathcal{M}(\rho_n)$ when λ and n are large enough.*

Proof. it suffices to show that (i) $m_n(Z_i) \in \mathcal{M}_{L,s,k}$ and (ii) $\inf_{\theta\in\Theta} E[\delta_\theta(Z_i)^2] \ge \rho_n^2$.

(i) Since $E[g(X_i,\theta_0)|Z_i] \in \mathcal{M}_{L_{\mathcal{M}},s,k}$ by Assumption 8, it suffices to show that $\delta_n(Z_i)$ is in $\mathcal{M}_{L-L_{\mathcal{M}},s,k}$. For any $z \in I_{\kappa',p}$, we have

$$|D^k\delta_n(z)| = \lambda\rho_n h_n^{-k}\left|\sum_{\kappa\in\mathcal{K}_n(p)} B_\kappa \phi^{(k)}\left(\frac{z-p\kappa h_n}{h_n}\right)\right| \;= \lambda\rho_n h_n^{-k}\left|B_{\kappa'}\phi^{(k)}\left(\frac{z-p\kappa' h_n}{h_n}\right)\right|,$$

where $\phi^{(k)}(Z_i)$ is k-times derivative of $\phi(Z_i)$. If z and y are in a same bin $I_{\kappa',p}$,

$$\begin{aligned}|D^k\delta_n(z) - D^k\delta_n(y)| &= \lambda\rho_n h_n^{-k}\left|B_{\kappa'}\phi^{(k)}\left(\frac{z-p\kappa' h_n}{h_n}\right) - B_{\kappa'}\phi^{(k)}\left(\frac{y-p\kappa' h_n}{h_n}\right)\right| \\ &\le \lambda\rho_n h_n^{-k}\left|\phi^{(k)}\left(\frac{z-p\kappa' h_n}{h_n}\right) - \phi^{(k)}\left(\frac{y-p\kappa' h_n}{h_n}\right)\right| \\ &\le \lambda\rho_n h_n^{-k}\frac{L-L_{\mathcal{M}}}{2}\left\|\frac{z-y}{h_n}\right\|^s = \frac{L-L_{\mathcal{M}}}{2}\|z-y\|^s,\end{aligned}$$

because $\phi \in \mathcal{M}_{(L-L_{\mathcal{M}})/2,s,k}$ and $h_n = (\lambda\rho_n)^{1/(s+k)}$. If $z \in I_{\kappa_z,p}$ and $y \in I_{\kappa_y,p}$ for $\kappa_z \neq \kappa_y$,

$$\begin{aligned}|D^k\delta_n(z) - D^k\delta_n(y)| &= \lambda\rho_n h_n^{-k}\left|B_{\kappa_z}\phi^{(k)}\left(\frac{z-p\kappa_z h_n}{h_n}\right) - B_{\kappa_y}\phi^{(k)}\left(\frac{y-p\kappa_y h_n}{h_n}\right)\right| \\ &\le \lambda\rho_n h_n^{-k}\left|B_{\kappa_z}\phi^{(k)}\left(\frac{z-p\kappa_z h_n}{h_n}\right) - B_{\kappa_z}\phi^{(k)}\left(\frac{y-p\kappa_z h_n}{h_n}\right)\right| \\ &+ \lambda\rho_n h_n^{-k}\left|B_{\kappa_y}\phi^{(k)}\left(\frac{z-p\kappa_y h_n}{h_n}\right) - B_{\kappa_y}\phi^{(k)}\left(\frac{y-p\kappa_y h_n}{h_n}\right)\right| \\ &\le (L-L_{\mathcal{M}})\|z-y\|^s,\end{aligned}$$

where we use the fact that $\phi_{\kappa_y}(z) = 0$ when $z \in I_{\kappa_z,p}$. Therefore, $\delta_n(Z_i) \in \mathcal{M}_{L-L_{\mathcal{M}},s,k}$ for any n and λ.

(ii) We have $\inf_{\theta\in\Theta} E[\delta_\theta(Z_i)^2] = E[\delta_{\check{\theta}_{m_n}}(Z_i)^2]$. Then, Minkowski's inequality yields

$$\begin{aligned}\{E[\delta_{\check{\theta}_{m_n}}(Z_i)^2]\}^{1/2} &= [E(\{\delta_n(Z_i) + E[g(X,\theta_0)|Z_i] - E[g(X_i,\check{\theta}_{m_n})|Z_i]\}^2)]^{1/2} \\ &\ge \{E[\delta_n(Z_i)^2]\}^{1/2} - [E(\{E[g(X_i,\theta_0)|Z_i] - E[g(X_i,\check{\theta}_{m_n})|Z_i]\}^2)]^{1/2}\end{aligned}$$

$$
\begin{aligned}
&= \{E[\delta_n(Z_i)^2]\}^{1/2} - \left\{ E\left(E\left[\left. \frac{\partial g(X_i,\theta)}{\partial \theta'}\right|_{\theta=\tilde{\theta}} (\check{\theta}_{m_n} - \theta_0) \right| Z_i \right]^2 \right) \right\}^{1/2} \\
&\geq \{E[\delta_n(Z_i)^2]\}^{1/2} - \left(E\left\{ \left\| \left. \frac{\partial g(X_i,\theta)}{\partial \theta'}\right|_{\theta=\tilde{\theta}} \right\|^2 \right\} \right)^{1/2} \|\check{\theta}_{m_n} - \theta_0\| \\
&= \{E[\delta_n(Z_i)^2]\}^{1/2} - O(1)\|\check{\theta}_{m_n} - \theta_0\|, \qquad \text{(S.1)}
\end{aligned}
$$

where Assumption 3 guarantees the mean value theorem for $g(X_i, \cdot)$ for an interior point $\tilde{\theta}$ between θ_0 and $\check{\theta}_{m_n}$, and the last equality holds by Assumption 5. The first term in the right hand side of equation (S.1) is $E[\delta_n(Z_i)^2] = \lambda^2\rho_n^2 p^{-l_z} f(0) \int \phi(u)^2 du + o(1) = \lambda^2\rho_n^2 p^{-l_z} C^2 + o(1)$, for some positive constant C, where the density f is bounded by Assumption 2 and $\int \phi(u)^2 du$ is bounded by its definition. From Assumption 11 (iii), a positive constant C' exists such that $\|\check{\theta}_{m_n} - \theta_0\| = C'|E[h(Z_i)\delta_n(Z_i)]|$, where

$$
\begin{aligned}
|E[h(Z_i)\delta_n(Z_i)]| &= \left| \lambda\rho_n \sum_{\kappa\in\mathcal{K}_n(p)} B_\kappa E\left[h(Z_i)\phi\left(\frac{Z_i - p\kappa h_n}{h_n} \right) \right] \right| \\
&\leq \lambda\rho_n K_n(p)^{l_z} \left| \int h(z)\phi\left(\frac{z - p\kappa h_n}{h_n} \right) f(z)dz \right| \\
&= \lambda\rho_n K_n(p)^{l_z} \left| h_n^{l_z} f(0)h(0) \int \phi(u)du + o(h_n^{l_z}) \right| \\
&= \lambda\rho_n p^{-l_z} o(1). \qquad \text{(S.2)}
\end{aligned}
$$

The last equality holds because $\int \phi(z)dz = 0$. Thus, (S.1) and (S.2) implies that

$$
\{E[\delta_n(Z_i)^2]\}^{1/2} - C'\|\theta_0 - \check{\theta}_{m_n}\| \geq \lambda\rho_n p^{-l_z/2}[C - p^{-l_z/2}o(1)],
$$

which is bounded from below by ρ_n when λ and n are large enough.

□

In what follows we construct a Bayesian a priori measure by using the result of Lemma

S.9 and show even the optimal Bayesian test that has the smallest errors of testing does not have non-trivial power. Replacing the minimax problem by a Bayesian problem is standard arguments to show the lower bound of testing power (see, for example, Ingster, 1993; Spokoiny, 1996; Lepski & Spokoiny, 1999; Lepski & Tsybakov, 2000; Guerre & Lavergne, 2002; Abramovich, Feis, Italia, & Theofanis, 2009; Ingster & Sapatinas, 2009). To prove Proposition A1, it suffices to show that

$$\sup_{m\in\mathcal{M}(\tilde{\rho}_n)} P(t_n \leq z_\alpha) + \sup_{m\in H_0} P(t_n > z_\alpha) \geq 1 + o(1). \tag{S.3}$$

To give a lower bound of the left hand side of equation (S.3), we consider a Bayesian a priori measure over H_0 and $H_{n,1}$ by regarding $m(\cdot)$ as a random variable defined on $H_0 \cup H_{n,1}$.

First, let Π_0 be the priori distribution defined on H_0 that has Dirac mass:

$$\Pi_0\{m(\cdot) = E[g(X,\theta_0)|\cdot]\} = 1.$$

Second, let B_κ be an i.i.d. Rademacher random variable independent of the observations with $P(B_\kappa = 1) = P(B_\kappa = -1) = 1/2$. For a sequence $\{b_\kappa \in \{-1,1\}\}_{\kappa\in\mathcal{K}_n(p)}$, let $\Pi_{n,1}$ be the priori distribution defined on $H_{n,1}$:

$$\Pi_{n,1}\left[m(\cdot) = E[g(X,\theta_0)|\cdot] + \lambda\rho_n h_n^{l_z/2}\sum_{\kappa\in\mathcal{K}_n(p)} b_\kappa\phi_\kappa(\cdot)\right] = \prod_{\kappa\in\mathcal{K}_n(p)} P\left(B_\kappa = b_\kappa\right),$$

where Lemma S.9 guarantees $\Pi_{n,1}$ to be an a priori measure over $H_{n,1}$. Then, $\Pi_n = \Pi_0 + \Pi_{n,1}$ is an a priori Bayesian measure over $H_0 \cup H_{n,1}$.

This gives the lower bound

$$\sup_{m\in\mathcal{M}(\tilde{\rho}_n)} P(t_n \leq z_\alpha) + \sup_{m\in H_0} P(t_n > z_\alpha) \geq \int P(t_n \leq z_\alpha) d\Pi_{n,1} + \int P(t_n > z_\alpha) d\Pi_0. \tag{S.4}$$

The right hand side of the above equation is the Bayes error of the test t_n that is the sum of type I and type II errors of testing. It is known that the optimal Bayesian test based on the likelihood ratio has the smallest error, which we now introduce.

Let $\mathcal{Y}$ and $\mathcal{Z}$ be the set of observations Y and Z, respectively, where the joint distribution of Y and Z (specifically, the conditional mean of Y given Z) is described by $m(\cdot)$, which suggests that the relation between Y and Z depends on $m(\cdot)$. Then, we denote by $p_m(\mathcal{Y}, \mathcal{Z})$ the joint density of $\mathcal{Y}$ and $\mathcal{Z}$. Average densities under the null and alternative hypotheses are $p_0(\mathcal{Y}, \mathcal{Z}) \equiv \int p_m(\mathcal{Y}, \mathcal{Z}) d\Pi_0$ and $p_{n,1}(\mathcal{Y}, \mathcal{Z}) \equiv \int p_m(\mathcal{Y}, \mathcal{Z}) d\Pi_{n,1}$, respectively. Let L_n denotes the likelihood ratio of the optimal Bayesian test, which is

$$L_n = \frac{p_{n,1}(\mathcal{Y}, \mathcal{Z})}{p_0(\mathcal{Y}, \mathcal{Z})} = \frac{\int p_m(\mathcal{Y}|\mathcal{Z}) d\Pi_{n,1}}{\int p_m(\mathcal{Y}|\mathcal{Z}) d\Pi_0} \equiv \frac{p_{n,1}(\mathcal{Y}|\mathcal{Z})}{p_0(\mathcal{Y}|\mathcal{Z})}.$$

By using the The Bayesian error of the optimal Bayes test (see, Theorem 13.3.1 of Lehmann & Romano, 2005, p.528), Guerre and Lavergne (2002) show that (S.3) holds if

$$\int L_n^2 p_0(\mathcal{Y}|\mathcal{Z}) d\mathcal{Y} \equiv E_0(L_n^2|\mathcal{Z}) \xrightarrow{p} 1, \tag{S.5}$$

where E_0 is the expectation under p_0.

By assumption, each ω_i is standard normal conditionally upon Z_i, where $\omega_i = Y_i - m(Z_i)$. Under Π_0, the conditional density of $\mathcal{Y}$ given $\mathcal{Z}$ is normal with mean $E[g(X_i, \theta_0)|Z_i]$. Since we have n observations,

$$p_0(\mathcal{Y}|\mathcal{Z}) = (2\pi)^{-n/2} \int \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \omega_i^2\right) d\Pi_0 = (2\pi)^{-n/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \omega_{i,0}^2\right),$$

where $\omega_{i,0} \equiv Y_i - E[g(X_i, \theta_0)|Z_i]$. Since $\omega_i = Y_i - m(Z_i) = Y_i - m_n(Z_i)$ almost surely under $H_{1,n}$, we yield

$p_{n,1}(\mathcal{Y}|\mathcal{Z})$

$$= (2\pi)^{-n/2} \int \exp\left(-\frac{1}{2} \sum_{i=1}^{n} [Y_i - m_n(Z_i)]^2 \right) d\Pi_{n,1}(m)$$

$$= (2\pi)^{-n/2} \int \exp\left(-\frac{1}{2} \sum_{i=1}^{n} [Y_i - E[g(X_i, \theta_0|Z_i)] + E[g(X_i, \theta_0|Z_i)] - m_n(Z_i)]^2 \right) d\Pi_{n,1}(m)$$

$$= (2\pi)^{-n/2} \int \exp\left(-\frac{1}{2} \sum_{i=1}^{n} [\omega_{i,0} - \delta_n(Z_i)]^2 \right) d\Pi_{n,1}(m)$$

$$= (2\pi)^{-n/2} \int \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \omega_{i,0}^2 + \sum_{i=1}^{n} \omega_{i,0}\delta_n(Z_i) - \frac{1}{2} \sum_{i=1}^{n} \delta_n(Z_i)^2 \right) d\Pi_{n,1}(m)$$

$$= p_0(\mathcal{Y}|\mathcal{Z}) \int \exp\left(\sum_{i=1}^{n} \omega_{i,0}\delta_n(Z_i) - \frac{1}{2} \sum_{i=1}^{n} \delta_n(Z_i)^2 \right) d\Pi_{n,1}(m).$$

Recall that

$$\sum_{i=1}^{n} \omega_{i,0}\delta_n(Z_i) = \lambda \rho_n h_n^{l_z/2} \sum_{\kappa \in \mathcal{K}_n(p)} \sum_{i=1}^{n} \omega_{i,0} B_\kappa \phi_\kappa(Z_i),$$

and

$$\sum_{i=1}^{n} \delta_n(Z_i)^2 = \lambda^2 \rho_n^2 h_n^{l_z} \sum_{i=1}^{n} \left[\sum_{\kappa \in \mathcal{K}_n(p)} B_\kappa \phi_\kappa(Z_i) \right]^2 = \lambda^2 \rho_n^2 h_n^{l_z} \sum_{i=1}^{n} \sum_{\kappa \in \mathcal{K}_n(p)} \phi_\kappa(Z_i)^2.$$

Thus,

$$L_n$$

$$= \frac{p_{n,1}(\mathcal{Y}, \mathcal{Z})}{p_0(\mathcal{Y}, \mathcal{Z})}$$

$$= \int \exp\left(\sum_{i=1}^{n} \omega_{i,0}\delta_n(Z_i) - \frac{1}{2} \sum_{i=1}^{n} \delta_n(Z_i)^2 \right) d\Pi_{n,1}(m)$$

$$= \int \exp\left(\sum_{i=1}^{n} \omega_{i,0}\delta_n(Z_i) \right) \exp\left(-\frac{1}{2} \sum_{i=1}^{n} \delta_n(Z_i)^2 \right) d\Pi_{n,1}(m)$$

$$= \int \exp\left(\lambda \rho_n h_n^{l_z/2} \sum_{\kappa \in \mathcal{K}_n(p)} \sum_{i=1}^{n} \omega_{i,0} B_\kappa \phi_\kappa(Z_i) \right) \exp\left(-\frac{1}{2} \lambda^2 \rho_n^2 h_n^{l_z} \sum_{i=1}^{n} \sum_{\kappa \in \mathcal{K}_n(p)} \phi_\kappa(Z_i)^2 \right) d\Pi_{n,1}(m)$$

$$= \exp\left(-\frac{\lambda^2\rho_n^2 h_n^{l_z}}{2}\sum_{i=1}^{n}\sum_{\kappa\in\mathcal{K}_n(p)}\phi_\kappa(Z_i)^2\right)$$

$$\times \prod_{\kappa\in\mathcal{K}_n(p)}\frac{1}{2}\left[\exp\left(\lambda\rho_n h_n^{l_z/2}\sum_{i=1}^{n}\omega_{i,0}\phi_\kappa(Z_i)\right)+\exp\left(-\lambda\rho_n h_n^{l_z/2}\sum_{i=1}^{n}\omega_{i,0}\phi_\kappa(Z_i)\right)\right].$$

Thus,

$$L_n^2$$

$$= \exp\left(-\lambda^2\rho_n^2 h_n^{l_z}\sum_{i=1}^{n}\sum_{\kappa\in\mathcal{K}_n(p)}\phi_\kappa(Z_i)^2\right)$$

$$\times \prod_{\kappa\in\mathcal{K}_n(p)}\frac{1}{4}\left[\exp\left(2\lambda\rho_n h_n^{l_z/2}\sum_{i=1}^{n}\omega_{i,0}\phi_\kappa(Z_i)\right)+2+\exp\left(-2\lambda\rho_n h_n^{l_z/2}\sum_{i=1}^{n}\omega_{i,0}\phi_\kappa(Z_i)\right)\right].$$

Conditionally on $\mathcal{Z}$, $\{2\lambda\rho_n h_n^{l_z/2}\omega_{i,0}\phi_\kappa(Z_i)\}_{i=1}^n$ is independent centered Gaussian for all $\kappa\in\mathcal{K}_n(p)$ with conditional variance given by $4\lambda^2\rho_n^2 h_n^{l_z}\phi_\kappa(Z_i)^2$. Since $E[\exp(u)] = \exp(\sigma^2/2)$ for any random variable u that follows centered gaussian with variance σ^2, we get

$$E_0(L_n^2|\mathcal{Z})$$

$$= \exp\left(-\lambda^2\rho_n^2 h_n^{l_z}\sum_{i=1}^{n}\sum_{\kappa\in\mathcal{K}_n(p)}\phi_\kappa(Z_i)^2\right)$$

$$\times \prod_{\kappa\in\mathcal{K}_n(p)}\frac{1}{4}\left[\exp\left(2\lambda^2\rho_n^2 h_n^{l_z}\sum_{i=1}^{n}\phi_\kappa(Z_i)^2\right)+2+\exp\left(2\lambda^2\rho_n^2 h_n^{l_z}\sum_{i=1}^{n}\phi_\kappa(Z_i)^2\right)\right]$$

$$= \exp\left(-\lambda^2\rho_n^2 h_n^{l_z}\sum_{i=1}^{n}\sum_{\kappa\in\mathcal{K}_n(p)}\phi_\kappa(Z_i)^2\right)\prod_{\kappa\in\mathcal{K}_n(p)}\frac{1}{2}\left[\exp\left(2\lambda^2\rho_n^2 h_n^{l_z}\sum_{i=1}^{n}\phi_\kappa(Z_i)^2\right)+1\right]$$

$$= \prod_{\kappa\in\mathcal{K}_n(p)}\exp\left(-\lambda^2\rho_n^2 h_n^{l_z}\sum_{i=1}^{n}\phi_\kappa(Z_i)^2\right)\frac{1}{2}\left[\exp\left(2\lambda^2\rho_n^2 h_n^{l_z}\sum_{i=1}^{n}\phi_\kappa(Z_i)^2\right)+1\right]$$

$$= \prod_{\kappa\in\mathcal{K}_n(p)}\frac{1}{2}\left[\exp\left(\lambda^2\rho_n^2 h_n^{l_z}\sum_{i=1}^{n}\phi_\kappa(Z_i)^2\right)+\exp\left(-\lambda^2\rho_n^2 h_n^{l_z}\sum_{i=1}^{n}\phi_\kappa(Z_i)^2\right)\right]$$

$$= \prod_{\kappa \in \mathcal{K}_n(p)} \cosh\left(\lambda^2 \rho_n^2 h_n^{l_z} \sum_{i=1}^n \phi_\kappa(Z_i)^2\right),$$

where $\cosh(\cdot)$ is the hyperbolic cosine function. By using $1 \leq \cosh(z) \leq \exp(z^2)$, we obtain,[8]

$$1 \leq E_0(L_n^2|\mathcal{Z}) \leq \exp\left(\sum_{\kappa \in \mathcal{K}_n(p)} \left[\lambda^2 \rho_n^2 h_n^{l_z} \sum_{i=1}^n \phi_\kappa(Z_i)^2\right]^2\right).$$

Then, (S.5) holds if

$$\sum_{\kappa \in \mathcal{K}_n(p)} \left[\rho_n^2 h_n^{l_z} \sum_{i=1}^n \phi_\kappa(Z_i)^2\right]^2 \xrightarrow{p} 0.$$

We see this by considering the expectation of this positive random variable. We obtain

$$\begin{aligned}
E\left\{\sum_{\kappa \in \mathcal{K}_n(p)} \left[\rho_n^2 h_n^{l_z} \sum_{i=1}^n \phi_\kappa(Z_i)^2\right]^2\right\} &= \rho_n^4 h_n^{2l_z} \sum_{\kappa \in \mathcal{K}_n(p)} E\left\{\left[\sum_{i=1}^n \phi_\kappa(Z_i)^2\right]^2\right\} \\
&= \rho_n^4 h_n^{2l_z} \sum_{\kappa \in \mathcal{K}_n(p)} E\left\{\sum_{i_1=1}^n \sum_{i_2=1}^n \phi_\kappa(Z_{i_1})^2 \phi_\kappa(Z_{i_2})^2\right\} \\
&= \rho_n^4 h_n^{2l_z} \sum_{\kappa \in \mathcal{K}_n(p)} E\left\{\sum_{i=1}^n \phi_\kappa(Z_i)^4 + \sum_{i_1,i_2=1}^n \sum_{i_1 \neq i_2} \phi_\kappa(Z_{i_1})^2 \phi_\kappa(Z_{i_2})^2\right\} \\
&= \rho_n^4 h_n^{2l_z} \sum_{\kappa \in \mathcal{K}_n(p)} \left\{nE[\phi_\kappa(Z)^4] + n(n-1)E[\phi_\kappa(Z)^2]^2\right\}.
\end{aligned}$$

Since $f < \overline{f}$ by Assumption 2, we have $E[\phi_\kappa(Z)^4] = h_n^{-2l_z} \int \phi\left[(z - p\kappa h_n)/h_n\right]^4 f(z)dz \leq h_n^{-l_z} \overline{f} \int \phi(u)^4 du = O(h_n^{-l_z})$, and $E[\phi_\kappa(Z)^2] = h_n^{-l_z} \int \phi\left[(z - p\kappa h_n)/h_n\right]^2 f(z)dz \leq \overline{f} \int \phi(u)^2 du = O(1)$. Since $K_n(p) = 1/(ph_n) = 1/(p(\lambda\rho_n)^{1/(s+k)}) = O(\rho_n^{-1/(s+k)})$ and $h_n = O(\rho_n^{1/(s+k)})$,

$$\begin{aligned}
E\left\{\sum_{\kappa \in \mathcal{K}_n(p)} \left[\rho_n^2 h_n^{l_z} \sum_{i=1}^n \phi_\kappa(Z_i)^2\right]^2\right\} &\leq \sum_{\kappa \in \mathcal{K}_n(p)} \left\{nO(\rho_n^4 h_n^{l_z}) + n(n-1)O(\rho_n^4 h_n^{2l_z})\right\} \\
&= O(\rho_n^{-l_z/(s+k)}) \left\{nO(\rho_n^4 \rho_n^{l_z/(s+k)}) + n(n-1)O(\rho_n^4 \rho_n^{2l_z/(s+k)})\right\}
\end{aligned}$$

[8] $\cosh(x) = 2^{-1}[\exp(x) + \exp(-x)]$. On the one hand Maclaurin expansion yields $\cosh(x) = 1 + x^2/2! + x^4/4! + \ldots$. On the other hand, Maclaurin expansion of $\exp(x^2/2)$ yields $\exp(x^2/2) = 1 + x^2/2! + 2x^4/4! + \ldots$. Therefore, we yield $\cosh(x) \leq \exp(x^2/2) \leq \exp(x^2)$.

$$= O(\rho_n^{-l_z/(s+k)}) \left\{ nO(\rho_n^{[4(s+k)+l_z]/(s+k)}) + n(n-1)O(\rho_n^{[4(s+k)+2l_z]/(s+k)}) \right\}$$
$$= \left\{ nO(\rho_n^4) + n(n-1)O(\rho_n^{[4(s+k)+l_z]/(s+k)}) \right\}.$$

Then, we consider the following cases:

(i) $s+k \geq l_z/4$: since $\rho_n = o(\tilde{\rho}_n) = o(n^{-2(s+k)/[l_z+4(s+k)]})$, it holds that $n\rho_n^4 = o(n^{[l_z-4(s+k)]/[l_z+4(s+k)]}) = o(1)$ and $n^2 O(\rho_n^{[4(s+k)+l_z]/(s+k)}) = n^2 o(n^{-2}) = o(1)$.

(ii) $s+k < l_z/4$: since $\rho_n = o(\tilde{\rho}_n) = o(n^{-1/4})$, we have $n\rho_n^4 = no(n^{-1})) = o(1)$ and $n^2 O(\rho_n^{[4(s+k)+l_z]/(s+k)}) = n^2 o(n^{-[4(s+k)+l_z]/4(s+k)}) = o(n^{[4(s+k)-l_z]/4(s+k)}) = o(1)$.

□

S-2 Proof of Proposition A2

Proof of Proposition A2. We first show asymptotic behavior of $\hat{v}_n$ under the null hypothesis. We define $v_n^2 = \frac{1}{K_n^{l_z}} \sum_{k \in \mathcal{K}} \frac{(N_k-1)\mathbb{1}\{N_k>1\}}{N_k} [E(\omega_i^2 | Z_i \in I_k)]^2$.

Lemma S.10. *Under Assumptions 1, 2, 3, 5, 6, and 10, v_n^2 is bounded from above, stochastically bounded from below uniformly in $m \in \mathcal{M}_{L,s,k}$, and satisfies $\hat{v}_n^2 - v_n^2 = o_p(1)$.*

Proof. From Assumption 2, we have $P(Z_i \in I_k) = \int_{Z_i \in I_k} f(z) dz \leq \overline{f} h_n^{l_z}$. In the same way, we obtain $P(Z_i \in I_k) \geq \underline{f} h_n^{l_z}$. Then, we obtain

$$v_n^2 \leq K_n^{-l_z} \sum_{k \in \mathcal{K}} [E(\omega_i^2 | Z_i \in I_k)]^2 \leq \frac{1}{\underline{f}} \sum_{k \in \mathcal{K}} P(Z_i \in I_k) [E(\omega_i^2 | Z_i \in I_k)]^2$$
$$\leq \frac{1}{\underline{f}} \sum_{k \in \mathcal{K}} P(Z_i \in I_k) E(\omega_i^4 | Z_i \in I_k) = \underline{f}^{-1} E(\omega_i^4),$$

where the right hand side is bounded by a constant by Assumption 1.

Note that $(N_k - 1)\mathbb{1}\{N_k > 1\}/N_k = (1 - 1/N_k)\mathbb{1}\{N_k \geq 2\} \geq 1/2$. It can be shown that $P(\min_{k \in \mathcal{K}} \mathbb{1}\{N_k > 1\} = 1) \to 1$ when $n/(K_n^{l_z} \log K_n^{l_z}) \to \infty$ under Assumptions 1 and 2 (see, Lemma 4 in Guerre & Lavergne, 2002). Thus, it holds that

$v_n^2 \geq \frac{1}{2K_n^{l_z}} \sum_{k\in\mathcal{K}} [E(\omega_i^2|Z_i \in I_k)]^2 \geq \frac{1}{2\bar{f}} P(Z_i \in I_k) \sum_{k\in\mathcal{K}} [E(\omega_i^2|Z_i \in I_k)]^2 \geq \frac{1}{2\bar{f}} P(Z_i \in I_k) \sum_{k\in\mathcal{K}} P(Z_i \in I_k)^2 [E(\omega_i^2|Z_i \in I_k)]^2 = \frac{1}{2\bar{f}} P(Z_i \in I_k)[E(\omega_i^2)]^2 > 0$ with probability one.

Now, $\hat{v}_n^2$ is decomposed as follows

$$
\begin{aligned}
&\hat{v}_n^2 \\
&= \frac{1}{K_n^{l_z}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i, Z_j\}\in I_k, i\neq j} \hat{u}_i^2 \hat{u}_j^2 \\
&= \frac{1}{K_n^{l_z}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i, Z_j\}\in I_k, i\neq j} [g(X_i,\theta_0) - g(X_i,\hat{\theta}) + \omega_i]^2 [g(X_j,\theta_0) - g(X_j,\hat{\theta}) + \omega_j]^2 \\
&= \frac{1}{K_n^{l_z}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i, Z_j\}\in I_k, i\neq j} \Big\{ [g(X_i,\theta_0) - g(X_i,\hat{\theta})]^2 [g(X_j,\theta_0) - g(X_j,\hat{\theta})]^2 \\
&+4\omega_i [g(X_i,\theta_0) - g(X_i,\hat{\theta})][g(X_j,\theta_0) - g(X_j,\hat{\theta})]^2 + 2\omega_i^2 [g(X_j,\theta_0) - g(X_j,\hat{\theta})]^2 \\
&+4\omega_i^2 \omega_j [g(X_j,\theta_0) - g(X_j,\hat{\theta})] + 4\omega_i\omega_j [g(X_i,\theta_0) - g(X_i,\hat{\theta})][g(X_j,\theta_0) - g(X_j,\hat{\theta})] + \omega_i^2\omega_j^2 \Big\} \\
&\equiv \frac{1}{K_n^{l_z}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i, Z_j\}\in I_k, i\neq j} \omega_i^2 \omega_j^2 \\
&+ \frac{4}{K_n^{l_z}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i, Z_j\}\in I_k, i\neq j} \omega_i^2 \omega_j [g(X_j,\theta_0) - g(X_j,\hat{\theta})] + R_n \\
&\equiv \bar{v}_n^2 + D_n + R_n,
\end{aligned}
$$

where R_n represents smaller terms due to $g(X_i,\theta_0) - g(X_i,\hat{\theta}) = O_p(1/\sqrt{n})$ by Assumptions 3, 5, and 10 under the null hypothesis.

First, we show that $D_n = o_p(1)$. Decompose $D_n = \sqrt{n}(\theta_0 - \hat{\theta})'\bar{D}_n + \sqrt{n}(\theta_0 - \hat{\theta})'\tilde{D}_n\sqrt{n}(\theta_0 - \hat{\theta})$, where

$$
\begin{aligned}
\bar{D}_n &= \frac{4}{K_n^{l_z}\sqrt{n}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i, Z_j\}\in I_k, i\neq j} \omega_i^2 \omega_j \frac{\partial g(X_j,\theta)}{\partial \theta} \\
\tilde{D}_n &= \frac{4}{K_n^{l_z} n} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i, Z_j\}\in I_k, i\neq j} \omega_i^2 \omega_j \frac{\partial g(X_j,\tilde{\theta})}{\partial \theta \partial \theta'},
\end{aligned}
$$

for some $\tilde{\theta}$ between θ_0 and $\hat{\theta}$. For some positive constant C,

$$
\begin{aligned}
&E(|\bar{D}_n|) \\
&\leq \frac{4}{K_n^{l_z}\sqrt{n}} \sum_{k\in\mathcal{K}} E\left[\frac{\mathbb{1}\{N_k>1\}}{N_k^2} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \omega_i^2|\omega_j| \left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\|\right] \\
&= \frac{4}{K_n^{l_z}\sqrt{n}} \sum_{k\in\mathcal{K}} E\left[\frac{\mathbb{1}\{N_k>1\}}{N_k^2} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} E(\omega_i^2|Z_i) E\left[|\omega_j| \left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\| \middle| Z_j\right]\right] \\
&= \frac{4}{K_n^{l_z}\sqrt{n}} \sum_{k\in\mathcal{K}} E\left\{\frac{\mathbb{1}\{N_k>1\}}{N_k^2} N_k(N_k-1) E(\omega_i^2|Z_i\in I_k) E\left[|\omega_j| \left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\| \middle| Z_j\in I_k\right]\right\} \\
&\leq \frac{4C}{K_n^{l_z}\sqrt{n}} \sum_{k\in\mathcal{K}} E\left(\mathbb{1}\{N_k>1\}\right) E\left[|\omega_j| \left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\| \middle| Z_j\in I_k\right] \\
&\leq \frac{4C}{K_n^{l_z}\sqrt{n}\underline{f}h_n^{l_z}} \sum_{k\in\mathcal{K}} P(Z_j\in I_k) E\left[|\omega_j| \left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\| \middle| Z_j\in I_k\right] \\
&\leq \frac{4C}{\sqrt{n}\underline{f}} E\left[|\omega_j| \left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\|\right] = O(1/\sqrt{n}),
\end{aligned}
$$

where $E(\omega_i^2|Z_i\in I_k) = E(\omega_i^2\mathbb{1}\{Z_i\in I_k\})$ is bounded by Assumption 1 and $E[|\omega_j|\|\partial g(X_j,\tilde{\theta})/\partial\theta\|]$ is bounded by Assumptions 1 and 5. Furthermore,

$$
\begin{aligned}
&E(|\tilde{D}_n|) \\
&\leq \frac{4}{K_n^{l_z}n} \sum_{k\in\mathcal{K}} E\left[\frac{\mathbb{1}\{N_k>1\}}{N_k^2} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \omega_i^2|\omega_j| \left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|\right] \\
&\leq \frac{4}{K_n^{l_z}n} \sum_{k\in\mathcal{K}} E\left(\left(\frac{\mathbb{1}\{N_k>1\}}{N_k^2} \sum_{Z_i\in I_k} \omega_i^4\right)^{1/2} \left[\frac{\mathbb{1}\{N_k>1\}}{N_k^2} \sum_{Z_i\in I_k} \left\{\sum_{Z_j\in I_k, j\neq i} |\omega_j| \left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|\right\}^2\right]^{1/2}\right) \\
&\leq \frac{4}{K_n^{l_z}n} \sum_{k\in\mathcal{K}} \left[E\left(\frac{\mathbb{1}\{N_k>1\}}{N_k^2} \sum_{Z_i\in I_k} \omega_i^4\right)\right]^{1/2} \left\{E\left[\frac{\mathbb{1}\{N_k>1\}}{N_k^2} \sum_{Z_i\in I_k} \left(\sum_{Z_j\in I_k, j\neq i} |\omega_j| \left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|\right)^2\right]\right\}^{1/2} \\
&= \frac{4}{K_n^{l_z}n} \sum_{k\in\mathcal{K}} \left[E\left(\frac{\mathbb{1}\{N_k>1\}}{N_k}\right)\right]^{1/2} \left[E(\omega_i^4|Z_i\in I_k)\right]^{1/2} \left\{E\left[\frac{\mathbb{1}\{N_k>1\}}{N_k^2} \sum_{Z_i\in I_k} \left(\sum_{Z_j\in I_k, j\neq i} |\omega_j|^2 \left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|^2\right.\right.\right. \\
&\qquad \left.\left.\left. + \sum_{Z_j\in I_k, j\neq i} \sum_{Z_l\in I_k, l\neq j} |\omega_j||\omega_l| \left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\| \left\|\frac{\partial g(X_l,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|\right)\right]\right\}^{1/2} \\
&\leq \frac{C}{K_n^{l_z}n} \sum_{k\in\mathcal{K}} \left\{E\left[\frac{(N_k-1)\mathbb{1}\{N_k>1\}}{N_k}\right] E\left[|\omega_j|^2 \left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|^2 \middle| Z_j\in I_k\right]\right.
\end{aligned}
$$

$$
\begin{aligned}
&\quad + E\left[\frac{(N_k-1)(N_k-2)\mathbb{1}\{N_k>1\}}{N_k}\right]\left[E\left(|\omega_j|\left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|\middle| Z_j\in I_k\right)\right]^2\Bigg\}^{1/2} \\
&\le \frac{C}{K_n^{l_z}n}\Bigg\{\sum_{k\in\mathcal{K}}E\left[\frac{(N_k-1)\mathbb{1}\{N_k>1\}}{N_k}\right]E\left[|\omega_j|^2\left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|^2\middle| Z_j\in I_k\right] \\
&\quad + \sum_{k\in\mathcal{K}}E\left[\frac{(N_k-1)(N_k-2)\mathbb{1}\{N_k>1\}}{N_k}\right]\left[E\left(|\omega_j|\left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|\middle| Z_j\in I_k\right)\right]^2\Bigg\}^{1/2} \\
&\le \frac{C}{K_n^{l_z}n}\Bigg\{\frac{1}{\underline{f}h_n^{l_z}}\sum_{k\in\mathcal{K}}P(Z_j\in I_k)E\left[|\omega_j|^2\left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|^2\middle| Z_j\in I_k\right] \\
&\quad + \sum_{k\in\mathcal{K}}E\left[N_k\mathbb{1}\{N_k>1\}\right]\left[E\left(|\omega_j|\left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|\middle| Z_j\in I_k\right)\right]^2\Bigg\}^{1/2} \\
&\le \frac{C}{K_n^{l_z}n}\Bigg\{\frac{1}{\underline{f}h_n^{l_z}}\left\{E\left(|\omega_j|^4\right)E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial g(X_j,\theta)}{\partial\theta\partial\theta'}\right\|^4\right]\right\}^{1/2} \\
&\quad + n\left[\sum_{k\in\mathcal{K}}P(Z_j\in I_k)E\left(|\omega_j|\left\|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right\|\middle| Z_j\in I_k\right)\right]^2\Bigg\}^{1/2} \\
&\le \frac{C}{K_n^{l_z}n}\left\{\frac{O(1)}{\underline{f}h_n^{l_z}}+nE\left(|\omega_j|^2\right)E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial g(X_j,\theta)}{\partial\theta\partial\theta'}\right\|^2\right]\right\}^{1/2} \\
&= \left\{O\left(\frac{1}{K_n^{l_z}n^2}\right)+O\left(\frac{1}{K_n^{2l_z}n}\right)\right\}^{1/2} = o(1),
\end{aligned}
$$

where $E(\mathbb{1}\{N_k>1\}/N_k)\le 1$ and $E(\omega_i^4|Z_i\in I_k)$, $E(|\omega_j|^4)$, and $E[\sup_{\theta\in\Theta}\|\partial g(X_j,\theta)/\partial\theta\partial\theta'\|^4]$ are bounded by Assumptions 1 and 6. Further, we use the fact that, given $\mathbb{1}\{N_k>1\}=1$, $(N_k-1)/N_k<1$ and $(N_k-1)(N_k-2)/N_k<N_k$. Thus, $D_n=o_p(1)$.

Next, we show $\bar{v}_n^2-v_n^2=o_p(1)$. Since we have i.i.d. observation,

$$
\begin{aligned}
[E(\omega_i^2|Z_i\in I_k)]^2 &= [E(\omega_i^2\mathbb{1}\{Z_i\in I_k\})]^2 \\
&= E(\omega_i^2\mathbb{1}\{Z_i\in I_k\})E(\omega_j^2\mathbb{1}\{Z_j\in I_k\}) \\
&= \frac{1}{N_k(N_k-1)}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}E(\omega_i^2\mathbb{1}\{Z_i\in I_k\})E(\omega_j^2\mathbb{1}\{Z_j\in I_k\}).
\end{aligned}
$$

Thus,

$$
E[|\bar{v}_n^2-v_n^2|^2]
$$

$$
\begin{aligned}
&= E\left\{\left|\frac{1}{K_n^{l_z}}\sum_{k\in\mathcal{K}}\frac{\mathbb{1}\{N_k>1\}}{N_k^2}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}\{\omega_i^2\omega_j^2 - E(\omega_i^2|Z_i\in I_k)E(\omega_j^2|Z_j\in I_k)\}\right|^2\right\} \\
&\leq \frac{1}{K_n^{2l_z}}\sum_{k\in\mathcal{K}} E\left\{\frac{\mathbb{1}\{N_k>1\}}{N_k^4}\left|\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}\{\omega_i^2\omega_j^2 - E(\omega_i^2|Z_i\in I_k)E(\omega_j^2|Z_j\in I_k)\}\right|^2\right\} \\
&= \frac{1}{K_n^{2l_z}}\sum_{k\in\mathcal{K}} E\left\{\frac{\mathbb{1}\{N_k>1\}}{N_k^4}\sum_{\substack{\{Z_{i_1},Z_{j_1}\}\in I_k\\ i_1\neq j_1}}\sum_{\substack{\{Z_{i_2},Z_{j_2}\}\in I_k\\ i_2\neq j_2}}\omega_{i_1}^2\omega_{j_1}^2\omega_{i_2}^2\omega_{j_2}^2\right\} \\
&- \frac{2}{K_n^{2l_z}}\sum_{k\in\mathcal{K}} E\left\{\frac{\mathbb{1}\{N_k>1\}}{N_k^4}\sum_{\substack{\{Z_{i_1},Z_{j_1}\}\in I_k\\ i_1\neq j_1}}\sum_{\substack{\{Z_{i_2},Z_{j_2}\}\in I_k\\ i_2\neq j_2}}\omega_{i_1}^2\omega_{j_1}^2 E(\omega_{i_2}^2|Z_{i_2}\in I_k)E(\omega_{j_2}^2|Z_{j_2}\in I_k)\right\} \\
&+ \frac{1}{K_n^{2l_z}}\sum_{k\in\mathcal{K}} E\left\{\frac{\mathbb{1}\{N_k>1\}}{N_k^4}\sum_{\substack{\{Z_{i_1},Z_{j_1}\}\in I_k\\ i_1\neq j_1}}\sum_{\substack{\{Z_{i_2},Z_{j_2}\}\in I_k\\ i_2\neq j_2}} E(\omega_{i_1}^2|Z_{i_1}\in I_k)E(\omega_{j_1}^2|Z_{j_1}\in I_k)\right. \\
&\qquad \left. E(\omega_{i_2}^2|Z_{i_2}\in I_k)E(\omega_{j_2}^2|Z_{j_2}\in I_k)\right\} \\
&= \frac{1}{K_n^{2l_z}}\sum_{k\in\mathcal{K}} E\left\{\frac{\mathbb{1}\{N_k>1\}}{N_k^4}\sum_{\substack{\{Z_{i_1},Z_{j_1}\}\in I_k\\ i_1\neq j_1}}\sum_{\substack{\{Z_{i_2},Z_{j_2}\}\in I_k\\ i_2\neq j_2}}\omega_{i_1}^2\omega_{j_1}^2\omega_{i_2}^2\omega_{j_2}^2\right\} \\
&- \frac{2}{K_n^{2l_z}}\sum_{k\in\mathcal{K}} E\left\{\frac{\mathbb{1}\{N_k>1\}}{N_k^2}[E(\omega_i^2|Z_i\in I_k)]^2\sum_{\substack{\{Z_{i_1},Z_{j_1}\}\in I_k\\ i_1\neq j_1}}\omega_{i_1}^2\omega_{j_1}^2\right\} \\
&+ \frac{1}{K_n^{2l_z}}\sum_{k\in\mathcal{K}} E\left\{\frac{\mathbb{1}\{N_k>1\}}{N_k^4}\left\{\sum_{\substack{\{Z_i,Z_j\}\in I_k\\ i_1\neq j_1}} E(\omega_i^2|Z_i\in I_k)E(\omega_j^2|Z_j\in I_k)\right\}^2\right\} \\
&\leq O(K_n^{-l_z}) + O(K_n^{-l_z})E\left\{E(\omega_{i_1}^2|Z_{i_1}\in I_k)E(\omega_{j_1}^2|Z_{j_1}\in I_k)\right\} + O(K_n^{-l_z}) \\
&= O(h^{l_z}) = o(1),
\end{aligned} \tag{S.6}
$$

where $E(\omega_i^4|Z_i \in I_k) = E(\omega_i^4\mathbb{1}\{Z_i \in I_k\}) = E[E(\omega_i^4|Z_i)\mathbb{1}\{Z_i \in I_k\}]$ is bounded by Assumption 1. This together with $D_n = o_p(1)$ yields $\hat{v}_n^2 - v_n^2 = o_p(1)$. □

The test statistic can be decomposed as follows;

$$
\begin{aligned}
&T_n^{GL}\\
&= \frac{1}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \hat{u}_i\hat{u}_j\\
&= \frac{1}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} [g(X_i,\theta_0)-g(X_i,\hat{\theta})+\omega_i][g(X_j,\theta_0)-g(X_j,\hat{\theta})+\omega_j]\\
&= \frac{1}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} [g(X_i,\theta_0)-g(X_i,\hat{\theta})][g(X_j,\theta_0)-g(X_j,\hat{\theta})]\\
&+ 2\frac{1}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} [g(X_i,\theta_0)-g(X_i,\hat{\theta})]\omega_j\\
&+ \frac{1}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \omega_i\omega_j\\
&\equiv T_1^{GL}+T_2^{GL}+T_3^{GL}.
\end{aligned}
$$

It is straightforward to show that $T_1^{GL}=o_p(1)$ by using $g(X_j,\theta_0)-g(X_i,\hat{\theta})=O_p(1/\sqrt{n})$ that holds by Assumptions 3, 5, and 10 under the null hypothesis. $T_2^{GL}=o_p(1)$ can be shown analogously to Lemma S.16 under Assumptions 1, 3, 5, 6, and 10.[9] We show the asymptotic heavier of T_3^{GL}.

Lemma S.11. *Under Assumptions 1, 2, 3, 5, and 10,* $T_3^{GL}/\hat{v}_n \xrightarrow{d} N(0,1)$.

Proof. We follow and extend the proof of Theorem 2 in Guerre and Lavergne (2002) into IV setup. Let $J_1,\ldots,J_n$ be any rearrangement of the indices $i=1,\ldots,n$ such that $X_{J_i}\in I_k$ if and only if (iff) $\sum_{l<k}N_l<J_i\leq\sum_{l\leq k}N_l$. Let $\mathcal{F}_{n,k}\equiv\{Y_1,\ldots,Y_{J_i},N_{\mathcal{K}}:\sum_{l<k}N_l<J_i\leq\sum_{l\leq k}N_l, k\in\mathcal{K}\}$ be a increasing sequence, where $N_{\mathcal{K}}\equiv\{N_k:k\in\mathcal{K}\}$ and $T_{3,n,k}^{GL}\equiv\frac{1}{\sqrt{2}K_n^{l_z/2}}\sum_{k'\leq k}\frac{\mathbb{1}\{N_{k'}>1\}}{N_{k'}}\sum_{\{Z_i,Z_j\}\in I_{k'},i\neq j}\omega_i\omega_j$. Then, $\{T_{3,n,k}^{GL},\mathcal{F}_{n,k},k\in\mathcal{K},n\geq1\}$ is zero-mean square integrable martingale array. Note that $T_{3,n,k}^{GL}-T_{3,n,k-1}^{GL}=$

[9] Note that we consider asymptotic behavior of T_2^{GL} under the null hypothesis, while Lemma S.16 is under the alternative. Thus, A_1 term, that comes form misspecification, in the proof of Lemma S.16 does not appear here. Only to show is A_2 with $\overline{\eta}_i$ replaced with ω_i to be $o_p(1)$.

$(2K_n^{l_z})^{-1/2}\frac{\mathbb{1}\{N_k>1\}}{N_k}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}\omega_i\omega_j \equiv (2K_n^{l_z})^{-1/2}w_k$ denotes the martingale differences. To prove $T_3^{GL}/\hat{v}_n \xrightarrow{d} N(0,1)$, it suffices to show that

$$\hat{v}_n^{-2}\sum_{k\in\mathcal{K}} E[(2K_n^{l_z})^{-1}w_k^2\mathbb{1}\{|(2K_n^{l_z})^{-1/2}w_k| > \epsilon\hat{v}_n\}|\mathcal{F}_{n,k-1}] \xrightarrow{p} 0 \quad \text{for all} \quad \epsilon > 0, \tag{S.7}$$

and

$$\hat{v}_n^{-2}\sum_{k\in\mathcal{K}} E[(2K_n^{l_z})^{-1}w_k^2|\mathcal{F}_{n,k-1}] \xrightarrow{p} 1, \tag{S.8}$$

by Corollary 3.1 in P. Hall and Heyde (1980, p.58). Square of the left hand side of (S.7) is bounded from above by

$$\hat{v}_n^{-4}\sum_{k\in\mathcal{K}} E\left[(2K_n^{l_z})^{-2}w_k^4|\mathcal{F}_{n,k-1}\right] = \frac{1}{\hat{v}_n^4 4K_n^{2l_z}}\sum_{k\in\mathcal{K}} E\left[w_k^4|\mathcal{F}_{n,k-1}\right] = O_p(K_n^{-l_z}),$$

because $\hat{v}_n^2 \xrightarrow{p} v_n^2$ from Lemma S.10 and for some constant C, we have

$$\begin{aligned}
& E\left[w_k^4|\mathcal{F}_{n,k-1}\right] \\
&= \frac{\mathbb{1}\{N_k>1\}}{N_k^4} E\left[\left(\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}\omega_i\omega_j\right)^4\middle|\mathcal{F}_{n,k-1}\right] \\
&= \frac{\mathbb{1}\{N_k>1\}}{N_k^4} E\left[\left(\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}\omega_i^2\omega_j^2\right)^2\middle|\mathcal{F}_{n,k-1}\right] \\
&= \frac{\mathbb{1}\{N_k>1\}}{N_k^4} \sum_{\{Z_{i_1},Z_{j_1}\}\in I_k, i_1\neq j_1}\ \sum_{\{Z_{i_2},Z_{j_2}\}\in I_k, i_2\neq j_2} E\left[\omega_{i_1}^2\omega_{j_1}^2\omega_{i_2}^2\omega_{j_2}^2\middle|\mathcal{F}_{n,k-1}\right] \\
&\leq \frac{\mathbb{1}\{N_k>1\}}{N_k^4} \sum_{\{Z_{i_1},Z_{j_1}\}\in I_k, i_1\neq j_1}\ \sum_{\{Z_{i_2},Z_{j_2}\}\in I_k, i_2\neq j_2} E\left(\omega_{i_1}^4\middle|\mathcal{F}_{n,k-1}\right)^{1/2} E\left(\omega_{j_1}^4\middle|\mathcal{F}_{n,k-1}\right)^{1/2} \\
&\quad \times E\left(\omega_{i_2}^4\middle|\mathcal{F}_{n,k-1}\right)^{1/2} E\left(\omega_{j_2}^4\middle|\mathcal{F}_{n,k-1}\right)^{1/2} \\
&\leq \mathbb{1}\{N_k>1\}\left[E\left(\omega_{i_1}^4\middle|\mathcal{F}_{n,k-1}\right)\right]^2 < C,
\end{aligned}$$

where the second equality comes from the orthogonality between ω_i and Z_i. Thus, equation (S.7) holds. Equation (S.8) is implied by

$$\begin{aligned}\sum_{k\in\mathcal{K}} E[(2K_n^{l_z})^{-1} w_k^2|\mathcal{F}_{n,k-1}] &= \frac{1}{K_n^{l_z}}\sum_{k\in\mathcal{K}}\frac{\mathbb{1}\{N_k>1\}}{N_k^2}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j} E[\omega_i^2\omega_j^2|N_{\mathcal{K}}] \\ &= \frac{1}{K_n^{l_z}}\sum_{k\in\mathcal{K}}\frac{(N_k-1)\mathbb{1}\{N_k>1\}}{N_k}[E(\omega_i^2|Z_i\in I_k)]^2 \\ &= v_n^2\end{aligned}$$

and Lemma S.10. □

□

S-3 Proof of Proposition A3

Proof of Proposition A3. We first show asymptotic behavior of $\hat{v}_n$ under the alternative hypothesis. We define $v_n^{*2} = \frac{1}{K_n^{l_z}}\sum_{k\in\mathcal{K}}\frac{(N_k-1)\mathbb{1}\{N_k>1\}}{N_k}[E(u_i^{*2}|Z_i \in I_k)]^2$, where $u_i^* \equiv Y_i - g(X_i,\theta^*)$.

Lemma S.12. *Under Assumptions 1, 2, 3, 4, 5, 7, 8, and 11, v_n^{*2} is bounded from above, stochastically bounded from below uniformly in $m \in \mathcal{M}_{L,s,k}$, and satisfies $\hat{v}_n^2 - v_n^{*2} = o_p(1)$.*

Proof. Since $E(u_i^{*4}) \le E(Y_i^4) + 2E(Y_i^2 g(X_i,\theta^*)^2) + E[g(X_i,\theta^*)^4] < \infty$ by Assumptions 1 and 4, it can be show similar to Lemma S.10 that $E(u_i^{*2}|Z_i \in I_k)$ is bounded from above uniformly in $m \in \mathcal{M}_{L,s,k}$ under the alternative by replacing ω_i in Lemma S.10 with u_i^*.

Similar to Lemma S.10, by applying Lemma 4 in Guerre and Lavergne (2002), we can show that v_n^{*2} is stochastically bounded from below uniformly in $m \in \mathcal{M}_{L,s,k}$.

Now, $\hat{v}_n^2$ can be decomposed as follows

$\hat{v}_n^2$

$$
\begin{aligned}
&= \frac{1}{K_n^{l_z}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \hat{u}_i^2 \hat{u}_j^2 \\
&= \frac{1}{K_n^{l_z}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} [g(X_i,\theta^*) - g(X_i,\hat{\theta}) + u_i^*]^2 [g(X_j,\theta^*) - g(X_j,\hat{\theta}) + u_j^*]^2 \\
&= \frac{1}{K_n^{l_z}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k^2} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \Big\{ [g(X_i,\theta^*) - g(X_i,\hat{\theta})]^2 [g(X_j,\theta^*) - g(X_j,\hat{\theta})]^2 \\
&+ 4u_i^* [g(X_i,\theta^*) - g(X_i,\hat{\theta})][g(X_j,\theta^*) - g(X_j,\hat{\theta})]^2 + 2u_i^{*2} [g(X_j,\theta^*) - g(X_j,\hat{\theta})]^2 \\
&+ 4u_i^* u_j^{*2} [g(X_j,\theta^*) - g(X_j,\hat{\theta})] + 4u_i^* u_j^* [g(X_i,\theta^*) - g(X_i,\hat{\theta})][g(X_j,\theta^*) - g(X_j,\hat{\theta})] + u_i^{*2} u_j^{*2} \Big\},
\end{aligned}
$$

where $g(X_i,\theta^*) - g(X_i,\hat{\theta}) = O_p(1/\sqrt{n})$ by Assumptions 3, 5, and 11. Similar to the proof of Lemma S.10, the dominated term in $\hat{v}_n^2$ is $\bar{v}_n^{*2} \equiv \frac{1}{K_n^{l_z}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k>1\}}{N_k^2} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} u_i^{*2} u_j^{*2}$. The convergence of $\bar{v}_n^{*2}$ is resulted by $E(|\bar{v}_n^{*2} - v_n^{*2}|^2) = o(1)$, whose proof goes along with equation (S.6) in Lemma S.10 and replacing ω with u^*. □

Under the alternative hypothesis,

$$
\begin{aligned}
T_n^{GL} &= \frac{1}{\sqrt{2} K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \hat{u}_i \hat{u}_j \\
&= \frac{1}{\sqrt{2} K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} [Y_i - g(X_i,\theta^*) + g(X_i,\theta^*) - g(X_i,\hat{\theta})] \\
&\quad \times [Y_j - g(X_j,\theta^*) + g(X_j,\theta^*) - g(X_j,\hat{\theta})] \\
&= \frac{1}{\sqrt{2} K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} [Y_i - g(X_i,\theta^*)][Y_j - g(X_j,\theta^*)] \\
&+ \frac{2}{\sqrt{2} K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k > 1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} [Y_i - g(X_i,\theta^*)][g(X_j,\theta^*) - g(X_j,\hat{\theta})] + B \\
&\equiv T_n^{GL*} + A + B,
\end{aligned}
$$

where $A = O(\sqrt{n h_n^{l_z}}) \{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2}$ by Lemma S.16 in the supplemental material for lemmas and B denotes smaller terms that is $O_p(1)$, which is shown by using $|g(X_i,\theta^*) -$

$g(X_i, \hat{\theta})| = O_p(n^{-1/2})$ by Assumptions 3, 5, and 11. Then, we obtain

$$P(T_n^{GL} \leq z_\alpha \hat{v}_n) = P(T_n^{GL} \leq z_\alpha v_n^*) + o(1) \leq P(T_n^{GL*} \leq C + \bar{C}\sqrt{nh_n^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2}) + o(1)$$

for some positive positive constants C and $\bar{C}$. Further,

$$\begin{aligned} &P\left(-[T_n^{GL*} - E(T_n^{GL*})] \geq E(T_n^{GL*}) - C - \bar{C}\sqrt{nh_n^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2}\right) + o(1) \\ &\leq \frac{var\left(T_n^{GL*}\right)}{[E(T_n^{GL*}) - C - \bar{C}\sqrt{nh_n^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2}]^2}, \end{aligned}$$

if $E(T_n^{GL*}) - C - \bar{C}\sqrt{nh_n^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2} > 0$. Then, it is sufficient to show that κ can be chosen so that

$$E(T_n^{GL*}) - C - \bar{C}\sqrt{nh_n^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2} > 0 \tag{S.9}$$

$$\frac{var\left(T_n^{GL*}\right)}{[E(T_n^{GL*}) - C - \bar{C}\sqrt{nh_n^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2}]^2} \leq \beta, \tag{S.10}$$

uniformly in $m \in \mathcal{M}(\kappa\rho_n)$. Now,

$$\begin{aligned} T_n^{GL*} &= \frac{1}{\sqrt{2}K_n^{l_z/2}}\sum_{k\in\mathcal{K}}\frac{\mathbb{1}\{N_k > 1\}}{N_k}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}[m(Z_i) - g(X_i,\theta^*) + \omega_i][m(Z_j) - g(X_j,\theta^*) + \omega_j] \\ &= \frac{1}{\sqrt{2}K_n^{l_z/2}}\sum_{k\in\mathcal{K}}\frac{\mathbb{1}\{N_k > 1\}}{N_k}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}[m(Z_i) - g(X_i,\theta^*)][m(Z_j) - g(X_j,\theta^*)] \\ &+ \frac{2}{\sqrt{2}K_n^{l_z/2}}\sum_{k\in\mathcal{K}}\frac{\mathbb{1}\{N_k > 1\}}{N_k}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}\omega_i[m(Z_j) - g(X_j,\theta^*)] \\ &+ \frac{1}{\sqrt{2}K_n^{l_z/2}}\sum_{k\in\mathcal{K}}\frac{\mathbb{1}\{N_k > 1\}}{N_k}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}\omega_i\omega_j \\ &\equiv T_{n,1}^{GL*} + T_{n,2}^{GL*} + T_{n,3}^{GL*}. \end{aligned}$$

It is obvious that $E(T_{n,2}^{GL*}) = 0$ and $E(T_{n,3}^{GL*}) = 0$. Then,

$$\begin{aligned} E(T_n^{GL*}) &= \frac{1}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} E\left[\frac{\mathbb{1}\{N_k > 1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \delta_{\theta^*}(Z_i)\delta_{\theta^*}(Z_j) \right] \\ &= \frac{1}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} E\left[(N_k - 1)\mathbb{1}\{N_k > 1\}\right] E[\delta_{\theta^*}(Z_i)|Z_i \in I_k]^2 \\ &\geq \frac{C_1 n h_n^{l_z/2}}{\sqrt{2}} \{E[\delta_{\theta^*}^2(Z_i)]^{1/2} - h_n^{s+k}\}^2, \end{aligned}$$

where the last inequality holds for large n and some constant $C_1 > 0$ under Assumption 2 by using Proposition 7 in Guerre and Lavergne (2002). By using this, we yield

$$\begin{aligned} &\frac{E(T_n^{GL*}) - C - \bar{C}\sqrt{nh_n^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2}}{nh_n^{lz/2}E[\delta_{\theta^*}^2(Z_i)]} \\ &\geq \frac{C_1(n-1)h_n^{l_z/2}\{E[\delta_{\theta^*}^2(Z_i)]^{1/2} - h_n^{s+k}\}^2 - C - \bar{C}\sqrt{nh_n^{l_z}}\{E[\delta_{\theta^*}^2(Z_i)]\}^{1/2}}{nh_n^{lz/2}E[\delta_{\theta^*}^2(Z_i)]} \\ &\geq \frac{C_1(n-1)}{n}\left\{1 - \frac{h_n^{s+k}}{\kappa\rho_n}\right\}^2 - \frac{C}{nh_n^{lz/2}\kappa^2\rho_n^2} - \frac{\bar{C}}{\sqrt{n}\kappa\rho_n} \\ &= \frac{C_1(n-1)}{n}\left\{1 - \frac{\lambda^{s+k}}{\kappa}\right\}^2 - \frac{C}{\kappa^2\lambda^{lz/2}} - O\left(n^{\frac{-l_z/2}{l_z+4(s+k)}}\right), \end{aligned}$$

which is increasing in κ and positive for κ large enough.

Next, let $\mu_k \equiv \frac{\mathbb{1}\{N_k>1\}}{N_k}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}[Y_i - g(X_i,\theta^*)][Y_j - g(X_j,\theta^*)]$ Then, we can write $T_n^{GL*} \equiv \frac{1}{\sqrt{2}K_n^{l_z/2}}\sum_{k\in\mathcal{K}}\mu_k$, where μ_k's are uncorrelated given $N_{\mathcal{K}} \equiv \{N_k, k \in \mathcal{K}\}$. Note that $(N_k - 1)\mathbb{1}\{N_k > 1\} = N_k - 1 + \mathbb{1}\{N_k = 0\}$, since $N_k - 1 = (N_k - 1)\mathbb{1}\{N_k > 1\} + (N_k - 1)\mathbb{1}\{N_k \leq 1\} = (N_k - 1)\mathbb{1}\{N_k > 1\} - \mathbb{1}\{N_k = 0\}$. We have

$$\begin{aligned} E(\mu_k|N_{\mathcal{K}}) &= \frac{\mathbb{1}\{N_k > 1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} E\left[\delta_{\theta^*}(Z_i)\delta_{\theta^*}(Z_j)|N_{\mathcal{K}}\right] \\ &= (N_k - 1)\mathbb{1}\{N_k > 1\}\{E\left[\delta_{\theta^*}(Z_i)|Z_i \in I_k\right]\}^2 \end{aligned}$$

and

$$
\begin{aligned}
E(\mu_k^2|N_{\mathcal{K}}) &= \frac{\mathbb{1}\{N_k>1\}}{N_k^2}E\left(\left.\left\{\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}[Y_i-g(X_i,\theta^*)][Y_j-g(X_j,\theta^*)]\right\}^2\right|N_{\mathcal{K}}\right)\\
&= \frac{(N_k-1)(N_k-2)(N_k-3)}{N_k}\mathbb{1}\{N_k>1\}\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^4\\
&+ \frac{4(N_k-1)(N_k-2)}{N_k}\mathbb{1}\{N_k>1\}\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^2E\left[u_i^{*2}|Z_i\in I_k\right]\\
&+ \frac{2(N_k-1)}{N_k}\mathbb{1}\{N_k>1\}[E\left(u_i^{*2}|Z_i\in I_k\right)]^2,
\end{aligned}
$$

which implies

$$
\begin{aligned}
var\left(\mu_k|N_{\mathcal{K}}\right) &= E(\mu_k^2|N_{\mathcal{K}})-[E(\mu_k|N_{\mathcal{K}})]^2\\
&= \frac{-2(N_k-1)(2N_k-3)}{N_k}\mathbb{1}\{N_k>1\}\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^4\\
&+ \frac{4(N_k-1)(N_k-2)}{N_k}\mathbb{1}\{N_k>1\}\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^2E\left[u_i^{*2}|Z_i\in I_k\right]\\
&+ \frac{2(N_k-1)}{N_k}\mathbb{1}\{N_k>1\}[E\left(u_i^{*2}|Z_i\in I_k\right)]^2.
\end{aligned}
$$

From the law of total variance, we obtain

$$
\begin{aligned}
var\left(T_n^{GL*}\right) &= \frac{1}{2K_n^{l_z}}var\left(\sum_{k\in\mathcal{K}}\mu_k\right) = \frac{1}{2K_n^{l_z}}E\left[var\left(\left.\sum_{k\in\mathcal{K}}\mu_k\right|N_{\mathcal{K}}\right)\right]+\frac{1}{2K_n^{l_z}}var\left[E\left(\left.\sum_{k\in\mathcal{K}}\mu_k\right|N_{\mathcal{K}}\right)\right]\\
&= \frac{1}{2K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left[var\left(\mu_k|N_{\mathcal{K}}\right)\right]+\frac{1}{2K_n^{l_z}}var\left[\sum_{k\in\mathcal{K}}E\left(\mu_k|N_{\mathcal{K}}\right)\right].
\end{aligned}
$$

Note that, given $\mathbb{1}\{N_k>1\}$, we have $-(N_k-1)(2N_k-3)/N_k<N_k$ and $4(N_k-1)(N_k-2)/N_k<N_k$. Since $E(N_k\mathbb{1}\{N_k>1\})=E(N_k)-E(N_k\mathbb{1}\{N_k\leq 1\})\leq E(N_k)$, we obtain

$$
\begin{aligned}
&\frac{1}{2K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left[var\left(\mu_k|N_{\mathcal{K}}\right)\right]\\
&= \frac{1}{2K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left(\frac{-2(N_k-1)(2N_k-3)}{N_k}\mathbb{1}\{N_k>1\}\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^4\right)
\end{aligned}
$$

$$+\frac{1}{2K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left(\frac{4(N_k-1)(N_k-2)}{N_k}\mathbb{1}\{N_k>1\}\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^2E\left[u_i^{*2}|Z_i\in I_k\right]\right)$$
$$+\frac{1}{2K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left(\frac{2(N_k-1)}{N_k}\mathbb{1}\{N_k>1\}[E\left(u_i^{*2}|Z_i\in I_k\right)]^2\right)$$
$$\leq\frac{2}{K_n^{l_z}}\sum_{k\in\mathcal{K}}E(N_k\mathbb{1}\{N_k>1\})\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^2\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^2$$
$$+\frac{2}{K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left(N_k\mathbb{1}\{N_k>1\}\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^2E\left[u_i^{*2}|Z_i\in I_k\right]\right)+E\left(v_n^{*2}\right)$$
$$\leq\frac{2n}{K_n^{l_z}}\sum_{k\in\mathcal{K}}P(Z_i\in I_k)E\left[\delta_{\theta^*}^2(Z_i)|Z_i\in I_k\right]E\left[\delta_{\theta^*}^2(Z_i)|Z_i\in I_k\right]$$
$$+\frac{2O(n)}{K_n^{l_z}}\sum_{k\in\mathcal{K}}P(Z_i\in I_k)E\left[\delta_{\theta^*}^2(Z_i)|Z_i\in I_k\right]+E\left(v_n^{*2}\right)$$
$$\leq\frac{2n}{K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left[\delta_{\theta^*}^2(Z_i)\right]^2+O(1)nh_n^{l_z}E\left[\delta_{\theta^*}^2(Z_i)\right]+E\left(v_n^{*2}\right)$$
$$\leq O(n)\{E\left[\delta_{\theta^*}^2(Z_i)\right]\}^2+O(nh_n^{l_z})E\left[\delta_{\theta^*}^2(Z_i)\right]+E\left(v_n^{*2}\right),$$

where $E\left[u_i^{*2}|Z_i\in I_k\right]$ is uniformly bounded by Assumptions 1 and 7 under the alternative hypothesis and $v_n^{*2}\equiv\frac{1}{K_n^{l_z}}\sum_{k\in\mathcal{K}}\frac{(N_k-1)\mathbb{1}\{N_k>1\}}{N_k}[E(u_i^{*2}|Z_i\in I_k)]^2$ is the limit of $\hat{v}_n$ under the alternative.

$$\frac{1}{2K_n^{l_z}}var\left[\sum_{k\in\mathcal{K}}E\left(\mu_k|N_{\mathcal{K}}\right)\right]$$
$$=\frac{1}{2K_n^{l_z}}var\left(\sum_{k\in\mathcal{K}}(N_k-1)\mathbb{1}\{N_k>1\}\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^2\right)$$
$$=\frac{1}{2K_n^{l_z}}\sum_{k\in\mathcal{K}}\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^4var\left((N_k-1)\mathbb{1}\{N_k>1\}\right)$$
$$+\frac{1}{2K_n^{l_z}}\sum_{k\neq k'}\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^2\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_{k'}\right]\}^2cov\left((N_k-1)\mathbb{1}\{N_k>1\},(N_{k'}-1)\mathbb{1}\{N_{k'}>1\}\right)$$
$$\leq\frac{n}{K_n^{l_z}}\sum_{k\in\mathcal{K}}P(Z_i\in I_k)\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^4$$
$$+\frac{n}{K_n^{l_z}}\sum_{k\neq k'}P(Z_i\in I_k)P(Z_i\in I_{k'})\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_k\right]\}^2\{E\left[\delta_{\theta^*}(Z_i)|Z_i\in I_{k'}\right]\}^2$$

$$\leq O(nh_n^{l_z})\{E\left[\delta^2_{\theta^*}(Z_i)\right]\} + O(nh_n^{l_z})\{E\left[\delta^2_{\theta^*}(Z_i)\right]\}^2$$

where we use the results of Lemma 3 in Guerre and Lavergne (2002), that is, $var((N_k - 1)\mathbb{1}\{N_k > 1\}) \leq 2nP(Z_i \in I_k)$ and $cov\left((N_k - 1)\mathbb{1}\{N_k > 1\}, (N_{k'} - 1)\mathbb{1}\{N_{k'} > 1\}\right) \leq 2nP(Z_i \in I_k)P(Z_i \in I_{k'})$. The last inequality holds because $\delta^2_{\theta^*}(Z_i) = \{m(Z_i) - E[g(X_i, \theta)|Z_i]\}^2$ is bounded by Assumptions 2 and 8 under the alternative hypothesis. Thus, we obtain

$$var\left(T_n^{GL*}\right) \leq O(n)\{E\left[\delta^2_{\theta^*}(Z_i)\right]\}^2 + O(nh_n^{l_z})E\left[\delta^2_{\theta^*}(Z_i)\right] + E\left(v_n^{*2}\right),$$

which implies

$$\frac{var\left(T_n^{GL*}\right)}{n^2 h_n^{lz} E[\delta^2_{\theta^*}(Z_i)]^2} \leq \frac{O(1)}{nh_n^{lz}} + \frac{O(1)}{n\kappa^2\rho_n^2} + \frac{E\left(v_n^{*2}\right)}{n^2 h_n^{lz}\kappa^4\rho_n^4} = O\left(\frac{1}{nh_n^{lz}}\right) + O\left(n^{\frac{-l_z}{l_z+4(s+k)}}\right)\frac{1}{\kappa^2} + \frac{E\left(v_n^{*2}\right)}{\kappa^4}.$$

The upper bound is bounded and decreasing in κ. □

Supplemental Material: Proof of Lemmas

S-1 Proof of Lemma 1

First, we introduce the following lemma.

Lemma S.13. *Under Assumptions 2 and 12, we have*

$$\int K\left(\frac{x-y}{h}\right) f(x)f(y)dxdy = h^{l_z}\int K(u)\,du \int f(y)^2dy + O(h^{l_z+1})$$

Proof.

$$\begin{aligned}\int K\left(\frac{x-y}{h}\right) f(x)f(y)dxdy &= h^{l_z}\int K(u)\,f(y+uh)f(y)dudy \\ &= h^{l_z}\int K(u)\,[f(y)+hu'\Delta^{(1)}f(y)+R_n]f(y)dudy \\ &= h^{l_z}\int K(u)\,du\int f(y)^2dy + O(h^{l_z+1})\end{aligned}$$

□

Proof. T_1 can be written as a second order U-statistic form multiplied by $nh^{l_z/2}$:

$$T_1 = \frac{2nh^{l_z/2}}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}\frac{1}{h^{l_z}}K\left(\frac{Z_j-Z_i}{h}\right)u_iu_j = \frac{2n}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i}H(W_i,W_j),$$

where $W_i = \{Z_i, u_i\}$, $H(W_i,W_j) \equiv h^{-l_z/2}K\left(\frac{Z_j-Z_i}{h}\right)u_iu_j$ is symmetric by Assumption 12, centered, that is, $E[H(W_i,W_j)] = 0$, and degenerate, that is, $E[H(W_i,W_j)|Z_i,u_i] = E\{E[H(W_i,W_j)|Z_i,Z_ju_i]Z_i,u_i\} = 0$. The second moment of $H(W_i,W_j)$ is bounded because

$$\begin{aligned}E[H(W_i,W_j)^2] &= \frac{1}{h^{l_z}}E\left[K\left(\frac{Z_j-Z_i}{h}\right)^2\sigma^2(Z_i)\sigma^2(Z_j)\right] \\ &= \frac{1}{h^{l_z}}\int K\left(\frac{x-y}{h}\right)^2\sigma^2(x)\sigma^2(y)f(x)f(y)dxdy\end{aligned}$$

$$= \int K(u)^2 \sigma^2(y+uh)\sigma^2(y) f(y+uh) f(y) du dy$$
$$= \mathcal{K}(0) \int [\sigma^2(y)]^2 f(y)^2 dy + O(h),$$

where the convolution product $\mathcal{K}(0)$ is bounded by Assumption 12. The variance of the error term $\sigma^2(\cdot)$ and the density of instrument $f(\cdot)$ are bounded by Assumptions 1 and 2, respectively. The last equality is shown in Lemma S.13.

Let $G(W_1, W_2) = E[H(W_1, W_3)H(W_2, W_3)|W_1, W_2]$. According to the central limit theorem for degenerate U-statistic, i.e. Q. Li and Racine (2007),

$$T_1/\sqrt{2E[H(W_i, W_j)^2]} \xrightarrow{d} N(0,1),$$

if

$$\frac{E[G(W_i, W_j)^2] + n^{-1}E[H(W_i, W_j)^4]}{\{E[H(W_i, W_j)^2]\}^2} \to 0, \tag{S.1}$$

as $n \to \infty$. We show that equation (S.1) holds. First, note that

$$\begin{aligned}
G(W_1, W_2) &= E[H(W_1, W_3)H(W_2, W_3)|W_1, W_2] \\
&= h^{-l_z} E\left[K\left(\frac{Z_1 - Z_3}{h}\right) K\left(\frac{Z_2 - Z_3}{h}\right) u_1 u_2 u_3^2 \middle| Z_1, Z_2, u_1, u_2 \right] \\
&= h^{-l_z} u_1 u_2 E\left[K\left(\frac{Z_1 - Z_3}{h}\right) K\left(\frac{Z_2 - Z_3}{h}\right) \sigma^2(Z_3) \middle| Z_1, Z_2, u_1, u_2 \right] \\
&= h^{-l_z} u_1 u_2 \int K\left(\frac{Z_1 - z}{h}\right) K\left(\frac{Z_2 - z}{h}\right) \sigma^2(z) f(z) dz \\
&= u_1 u_2 \int K(u) K\left(\frac{Z_2 - Z_1}{h} + u\right) \sigma^2(Z_1 - uh) f(Z_1 - uh) du.
\end{aligned}$$

Then,

$$\begin{aligned}
&E[G(W_1, W_2)^2] \\
&= E\left\{ u_1^2 u_2^2 \left[\int K(u) K\left(\frac{Z_2 - Z_1}{h} + u\right) \sigma^2(Z_1 - uh) f(Z_1 - uh) du \right]^2 \right\}
\end{aligned}$$

$$
\begin{aligned}
&= \int \sigma^2(x)\sigma^2(y) \left[\int K(u) K\left(\frac{y-x}{h} + u\right) \sigma^2(x-uh) f(x-uh) du \right]^2 f(x) f(y) dx dy \\
&= h^{l_z} \int \sigma^2(x)\sigma^2(x+vh) \left[\int K(u) K\left(v+u\right) \sigma^2(x-uh) f(x-uh) du \right]^2 f(x) f(x+vh) dx dv \\
&= h^{l_z} \int \sigma^2(x)[\sigma^2(x) + O(h)] \left[\sigma^2(x) f(x) \int K(u) K\left(v+u\right) du + O(h) \right]^2 f(x)[f(x) + O(h)] dx dv \\
&= h^{l_z} \int [\sigma^2(x)]^4 f(x)^4 dx \int \left[\int K(u) K\left(v+u\right) du \right]^2 dv + o(h^{l_z}) \\
&= O(h^{l_z}). \qquad \text{(S.2)}
\end{aligned}
$$

Second, let us define $\sigma^4(Z_i) = E(\omega_i^4 | Z_i)$, which is bounded by Assumption 1. Then,

$$
\begin{aligned}
E[H(W_i, W_j)^4] &= \frac{1}{h^{2l_z}} E\left[K\left(\frac{Z_j - Z_i}{h}\right)^4 \sigma^4(Z_i)\sigma^4(Z_j) \right] \\
&\leq \frac{M^2}{h^{2l_z}} \int K\left(\frac{x-y}{h}\right)^4 f(x) f(y) dx dy \\
&= \frac{M^2}{h^{l_z}} \left[\int K(u)^4 du \int f(y)^4 dy + O(h) \right] = O(h^{-l_z}). \qquad \text{(S.3)}
\end{aligned}
$$

Third,

$$
\begin{aligned}
\{E[H(W_i, W_j)^2]\}^2 &= \frac{1}{h^{2l_z}} \left\{ E\left[K\left(\frac{Z_j - Z_i}{h}\right)^2 \sigma^2(Z_i)\sigma^2(Z_j) \right] \right\}^2 \\
&= \frac{1}{h^{2l_z}} \left[\int K\left(\frac{x-y}{h}\right)^2 f(x) f(y) \sigma^2(x)\sigma^2(y) dx dy \right]^2 \\
&= \left[\int K(u)^2 f(y+uh) f(y) \sigma^2(y+uh)\sigma^2(y) du dy \right]^2 \\
&= \left\{ \int K(u)^2 [f(y) + O(h)] f(y) [\sigma^2(y) + O(h)] \sigma^2(y) du dy \right\}^2 \\
&= \left\{ \int K(u)^2 du \int f(y)^2 [\sigma^2(y)]^2 dy + O(h) \right\}^2 \\
&= O(1). \qquad \text{(S.4)}
\end{aligned}
$$

Equation (S.2), (S.3), and (S.4) implies

$$\frac{E[G(W_i,W_j)^2]+n^{-1}E[H(W_i,W_j)^4]}{\{E[H(W_i,W_j)^2]\}^2}=\frac{O(h^{l_z})+O(n^{-1}h^{-l_z})}{O(1)}\to 0, \qquad \text{(S.5)}$$

as $n\to\infty$. Thus, equation (S.1) holds, and we obtain

$$T_1 \xrightarrow{d} N(0,2\mathcal{K}(0)E\{[\sigma^2(y)]^2 f(y)\}).$$

□

S-2 Proof of Lemma 2

Proof. Applying Taylor expansion to $g(X_j,\hat{\theta})$ around θ_0 under Assumption 3 yields

$$\begin{aligned}
T_2 &= \frac{2}{(n-1)h^{l_z/2}}\sum_{i=1}^{n}\sum_{j\neq i}K\left(\frac{Z_j-Z_i}{h}\right)u_i[g(X_j,\theta_0)-g(X_j,\hat{\theta})]\\
&= \frac{2(\hat{\theta}-\theta_0)'}{(n-1)h^{l_z/2}}\sum_{i=1}^{n}\sum_{j\neq i}K\left(\frac{Z_j-Z_i}{h}\right)u_i\frac{\partial}{\partial\theta}g(X_j,\theta_0)+R_n\\
&\equiv \frac{2\sqrt{n}(\hat{\theta}-\theta_0)'}{\sqrt{n}(n-1)h^{l_z/2}}\sum_{i=1}^{n}\sum_{j<i}\mu_{i,j}+\frac{2\sqrt{n}(\hat{\theta}-\theta_0)'}{\sqrt{n}(n-1)h^{l_z/2}}\sum_{i=1}^{n}\sum_{j>i}\mu_{i,j}+R_n,
\end{aligned}$$

where $\mu_{i,j}=K\left(\frac{Z_j-Z_i}{h}\right)u_i\frac{\partial}{\partial\theta}g(X_j,\theta_0)$ is a $l_z\times 1$ vector, R_n represents smaller terms under Assumptions 1, 2, 6, and 12, and $\sqrt{n}(\hat{\theta}-\theta)=O_p(1)$ by Assumption 10. It is useful that for $j\neq i$, $E(\mu_{i,j})=E[E(\mu_{i,j}|Z,X_{-i})]=\mathbf{0}$, where $\mathbf{0}$ denotes $l_z\times 1$ zero vector. By using this, we obtain that $\sum_{i=1}^{n}\sum_{j<i}E(\mu_{i,j})=\sum_{i=1}^{n}\sum_{j<i}E[E(\mu_{i,j}|Z_1,\ldots,Z_n,X_1,\ldots,X_{i-1})]=0$, and $\sum_{i=1}^{n}\sum_{j>i}E(\mu_{i,j})=0$. Thus, to show $T_2'=o_p(1)$, it suffices that variance for each elements of $\frac{2}{\sqrt{n}(n-1)h^{l_z/2}}\sum_{i=1}^{n}\sum_{j<i}\mu_{i,j}$ and $\frac{2}{\sqrt{n}(n-1)h^{l_z/2}}\sum_{i=1}^{n}\sum_{j>i}\mu_{i,j}$ is $o(1)$. To simplify the notation, we show the case for $l_z=1$. Let $G(z)\equiv E[\frac{\partial}{\partial\theta}g(x,\theta)|z]$ and

$G_2(z) \equiv E[|\frac{\partial}{\partial\theta}g(x,\theta)|^2|z]$. Then,

$$
\begin{aligned}
&var\left(\frac{2}{\sqrt{n}(n-1)h^{l_z/2}}\sum_{i=1}^{n}\sum_{j<i}\mu_{i,j}\right) = \frac{4}{n(n-1)^2h^{l_z}}E\left(\sum_{i_1=1}^{n}\sum_{j_1<i_1}\sum_{i_2=1}^{n}\sum_{j_2<i_2}\mu_{i_1,j_1}\mu_{i_2,j_2}\right)\\
&\quad = \frac{4}{n(n-1)^2h^{l_z}}\sum_{i}^{n}\sum_{j<i}E\left(\mu_{i,j}^2\right) + \frac{4}{n(n-1)^2h^{l_z}}\sum_{i}^{n}\sum_{j_1<i}\sum_{\substack{j_2<i\\ j_2\neq j_1}}E\left(\mu_{i,j_1}\mu_{i,j_2}\right)\\
&\quad = O(n^{-1}) + O(h^{l_z}),
\end{aligned}
$$

because for all i, $j_1 < i$, $j_2 < i$, and $j_1 \neq j_2$,

$$
\begin{aligned}
&E\left(\mu_{i,j_1}\mu_{i,j_2}\right)\\
&= E\left[K\left(\frac{Z_{j_1}-Z_i}{h}\right)K\left(\frac{Z_{j_2}-Z_i}{h}\right)\frac{\partial g(X_{j_1},\theta)}{\partial\theta}\frac{\partial g(X_{j_2},\theta)}{\partial\theta}\sigma^2(Z_i)\right]\\
&= E\left[\int K\left(\frac{Z_{j_1}-z}{h}\right)K\left(\frac{Z_{j_2}-z}{h}\right)\sigma^2(z)f(z)dz\frac{\partial g(X_{j_1},\theta)}{\partial\theta}\frac{\partial g(X_{j_2},\theta)}{\partial\theta}\right]\\
&= h^{l_z}E\left[\frac{\partial g(X_{j_1},\theta)}{\partial\theta}\frac{\partial g(X_{j_2},\theta)}{\partial\theta}\int K(u)K\left(\frac{Z_{j_2}-Z_{j_1}}{h}+u\right)\sigma^2(Z_{j_1}-uh)f(Z_{j_1}-uh)du\right]\\
&= h^{l_z}E\left[K^2\left(\frac{Z_{j_2}-Z_{j_1}}{h}\right)\sigma^2(Z_{j_1})f(Z_{j_1})\frac{\partial g(X_{j_1},\theta)}{\partial\theta}\frac{\partial g(X_{j_2},\theta)}{\partial\theta}\right] + R_n\\
&= h^{l_z}E\left[\sigma^2(Z_{j_1})f(Z_{j_1})\frac{\partial g(X_{j_1},\theta)}{\partial\theta}\int K^2\left(\frac{z_2-Z_{j_1}}{h}\right)\frac{\partial g(x_2,\theta)}{\partial\theta}f_{z,x}(z_2,x_2)dz_2dx_2\right] + R_n\\
&= h^{2l_z}E\left[\sigma^2(Z_{j_1})f(Z_{j_1})\frac{\partial g(X_{j_1},\theta)}{\partial\theta}\int K^2(u)\frac{\partial g(x_2,\theta)}{\partial\theta}f_{z,x}(Z_{j_1}+uh,x_2)dudx_2\right] + R_n\\
&= h^{2l_z}E\left\{\sigma^2(Z_{j_1})f(Z_{j_1})^2\frac{\partial g(X_{j_1},\theta)}{\partial\theta}\int K^2(u)duE\left[\frac{\partial g(X,\theta)}{\partial\theta}\right]\right\} + R_n\\
&= O(h^{2l_z})E\left[\sigma^2(Z_{j_1})f(Z_{j_1})^2\frac{\partial g(X_{j_1},\theta)}{\partial\theta}\right] + R_n\\
&= O(h^{2l_z}),
\end{aligned}
$$

where R_n represents smaller terms, $f_{z,x}(\cdot,\cdot)$ is joint density of Z and X, $E[\frac{\partial}{\partial\theta}g(X,\theta)]$ is bounded by Assumption 3, $K^2(u)$ is two times convolution product of kernel with $\int K^2(u)du < \infty$ by Assumption 12, $\sigma^2(Z)$ and $f(Z)$ are bounded by Assumptions 1 and

2, respectively, and for all i, $j = j_1 = j_2 < i$,

$$
\begin{aligned}
E\left(\mu_{i,j}^2\right) &= E\left\{K\left(\frac{Z_j - Z_i}{h}\right)^2 \sigma(Z_i)^2 \frac{\partial g(X_j,\theta_0)}{\partial\theta}\right\} \\
&= E\left\{\frac{\partial g(X_j,\theta_0)}{\partial\theta}\int K\left(\frac{Z_j - z}{h}\right)^2 \sigma(z)^2 f(z)dz\right\} \\
&= h^{l_z} E\left\{\frac{\partial g(X_j,\theta_0)}{\partial\theta}\int K(u)^2\sigma(Z_j - uh)^2 f(Z_j - uh)du\right\} \\
&= h^{l_z} E\left\{\frac{\partial g(X_j,\theta_0)}{\partial\theta} f(Z_j)\sigma(Z_j)^2 \int K(u)^2 du\right\} \\
&= O(h^{l_z}),
\end{aligned}
$$

where these derivations holds under Assumptions 1, 2, 3, and 12. In the same way, we obtain $var\left(\frac{2}{\sqrt{n}(n-1)h^{l_z/2}}\sum_{i=1}^n\sum_{j>i}\mu_{i,j}\right) = O(h^{l_z})$. □

S-3 Proof of Lemma 3

Proof. From Assumption 3, we obtain

$$
\begin{aligned}
T_3 &= \frac{nh^{l_z/2}}{n(n-1)}\sum_{i=1}^n\sum_{j\neq i}\frac{1}{h^{l_z}}K\left(\frac{Z_j - Z_i}{h}\right)(\theta_0 - \hat{\theta})' \left.\frac{\partial g(X_i,\theta)}{\partial\theta}\right|_{\theta=\tilde{\theta}_1} \left.\frac{\partial g(X_j,\theta)}{\partial\theta'}\right|_{\theta=\tilde{\theta}_2}(\theta_0 - \hat{\theta}) \\
&= \sqrt{n}(\theta_0 - \hat{\theta})'\frac{1}{n(n-1)h^{l_z/2}}\sum_{i=1}^n\sum_{j\neq i} T_3'\sqrt{n}(\theta_0 - \hat{\theta})
\end{aligned}
$$

where $T_3' \equiv K\left(\frac{Z_j - Z_i}{h}\right) \left.\frac{\partial g(X_i,\theta)}{\partial\theta}\right|_{\theta=\tilde{\theta}_1} \left.\frac{\partial g(X_j,\theta)}{\partial\theta'}\right|_{\theta=\tilde{\theta}_2}$. We have

$$
\begin{aligned}
E(\|T_3'\|) &\le \left\{E\left[K\left(\frac{Z_j - Z_i}{h}\right)^2\right]\right\}^{1/2}\left\{E\left[\left\|\left.\frac{\partial g(X_i,\theta)}{\partial\theta}\right|_{\theta=\tilde{\theta}_1}\right\|^2\left\|\left.\frac{\partial g(X_j,\theta)}{\partial\theta'}\right|_{\theta=\tilde{\theta}_2}\right\|^2\right]\right\}^{1/2} \\
&\le \left[\int K\left(\frac{z_2 - z_1}{h}\right)^2 f(z_1)f(z_2)dz_1dz_2\right]^{1/2} E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial g(X_i,\theta)}{\partial\theta}\right\|^2\right] \\
&= \left[h^{l_z}\int K(u)^2 f(z_1)f(z_1 + uh)dz_1du\right]^{1/2} O(1) = O(h^{l_z/2}),
\end{aligned}
$$

where these derivations hold by Assumptions 1, 2, 5, and 12. Since $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$ by Assumption 10, we obtain $T_3 = O_p(h^{l_z/2}) = o_p(1)$.

□

S-4 Proof of Lemma 4

First, we introduce the following lemma.

Lemma S.14. *Let $k_{i,j}$ and S_i be any function such that $\sup_{i,j} |k_{i,j}| < \infty$ and $\sum_i^n |S_i| = O_p(n)$. Then,*

$$\sum_i^n \sum_j^n k_{i,j} S_i S_j = O_p(n^2).$$

Proof. $\sum_i^n \sum_j^n k_{i,j} S_i S_j \leq \sup_{i,j} |k_{i,j}| \sum_i^n |S_i| \sum_j^n |S_j| = O_p(n^2)$. □

Lemma S.14 seems straightforward but is useful. The kernel function $K(\cdot)$ satisfies the condition for $k_{i,j}$ by Assumption 12. The condition for S_i is satisfied by $\frac{\partial}{\partial \theta_l} g(X_i, \theta)\Big|_{\theta=\tilde{\theta}}$, $\left[\frac{\partial}{\partial \theta_l} g(X_i, \theta)\Big|_{\theta=\tilde{\theta}}\right]^2$, and $\frac{\partial}{\partial \theta_l} g(X_i, \theta)\Big|_{\theta=\tilde{\theta}} u_i$ for any $l = \{1, \ldots, l_z\}$ and any $\tilde{\theta}$ between θ_0 and $\hat{\theta}$ by Assumptions 1, 3, 5, and 10, as well as u_i^2 by Assumption 1.

Proof. $\hat{\Sigma}$ can be decomposed as follows:

$$\begin{aligned}
\hat{\Sigma} &= \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 \hat{u}_i^2 \hat{u}_j^2. \\
&= \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 [g(X_i, \theta_0) - g(X_i, \hat{\theta}) + u_i]^2 [g(X_j, \theta_0) - g(X_j, \hat{\theta}) + u_j]^2 \\
&= \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 u_i^2 u_j^2 \\
&+ \frac{2}{n(n-1)} \sum_{i=1}^n \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 [g(X_i, \theta_0) - g(X_i, \hat{\theta})] u_i u_j^2 + R_n, \\
&\equiv 2\hat{\Sigma}_1 + \hat{\Sigma}_2 + R_n,
\end{aligned}$$

where R_n represents smaller terms that converges to zero in probability, which can be shown by using Lemma S.14 under Assumptions 1, 3, 5, 10, and 12. First, we show that $\hat{\Sigma}_2 = O_p(n^{-1/2})$.

$$\begin{aligned}\hat{\Sigma}_2 &= \sqrt{n}(\theta_0 - \hat{\theta})' \frac{2}{\sqrt{n}n(n-1)h^{l_z}} \sum_{i=1}^{n} \sum_{j \neq i} K\left(\frac{Z_j - Z_i}{h}\right)^2 \frac{\partial g(X_i, \theta)}{\partial \theta}\bigg|_{\theta=\tilde{\theta}} u_i u_j^2 \\ &\equiv \sqrt{n}(\theta_0 - \hat{\theta})' \frac{2}{\sqrt{n}n(n-1)h^{l_z}} \sum_{i=1}^{n} \sum_{j \neq i} S_{i,j},\end{aligned}$$

where $S_{i,j}$ is a $l_z \times 1$ vector. Since $\sqrt{n}(\hat{\theta} - \theta_0) = O_p(1)$ by Assumption 10, it suffices to show that each element of $n^{-5/2}h^{-l_z} \sum_{i=1}^{n} \sum_{j \neq i} S_{i,j}$ is $o_p(1)$, which holds becuase

$$\begin{aligned}E(\|S_{i,j}\|) &\leq E\left[K\left(\frac{Z_j - Z_i}{h}\right)^2 \left\| \frac{\partial g(X_i, \theta)}{\partial \theta}\bigg|_{\theta=\tilde{\theta}} u_i \right\| \sigma(Z_j)^2\right] \\ &= E\left[\left\| \frac{\partial g(X_i, \theta)}{\partial \theta}\bigg|_{\theta=\tilde{\theta}} u_i \right\| \int K\left(\frac{z - Z_i}{h}\right)^2 \sigma(z)^2 f(z) dz\right] \\ &= h^{l_z} E\left[\left\| \frac{\partial g(X_i, \theta)}{\partial \theta}\bigg|_{\theta=\tilde{\theta}} u_i \right\| \int K(u)^2 \sigma(Z_i + uh)^2 f(Z_i + uh) du\right] \\ &= h^{l_z} E\left[\left\| \frac{\partial g(X_i, \theta)}{\partial \theta}\bigg|_{\theta=\tilde{\theta}} u_i \right\| \sigma(Z_i)^2 f(Z_i) \int K(u)^2 du\right] \\ &\leq O(h^{l_z}) E\left[\sup_{\theta \in \Theta} \left\| \frac{\partial g(X_i, \theta)}{\partial \theta} \right\|^2\right]^{1/2} [\sigma(Z_i)^2]^{1/2} = O(h^{l_z}),\end{aligned}$$

where the last equality holds under Assumptions 1, 2, 5, and 12. Thus, $\hat{\Sigma}_2 = O_p(n^{-1/2})$.

Next, we show that $\hat{\Sigma}_1 = \Sigma + o_p(1)$. Note that $\hat{\Sigma}_1$ is a second-order U statistic, where $H_1(W_i, W_j) \equiv h^{-l_z} K\left(\frac{Z_j - Z_i}{h}\right)^2 u_i^2 u_j^2$ is symmetric by Assumption 12.

$$\begin{aligned}E[|H_1(W_i, W_j)|^2] &= \frac{1}{h^{2l_z}} \int K\left(\frac{z_2 - z_1}{h}\right)^4 [\sigma^2(z_1)]^2 [\sigma^2(z_2)]^2 f(z_1) f(z_2) dz_1 dz_2 \\ &= \frac{1}{h^{l_z}} \int K(u)^4 [\sigma^2(z_1)]^2 [\sigma^2(z_1 + uh)]^2 f(z_1) f(z_1 + uh) dz_1 du \\ &= \frac{1}{h^{l_z}} \int K(u)^4 du \int [\sigma^2(z_1)]^4 f(z_1)^2 dz_1 + O(h^{1-l_z}) \\ &= O(h^{-l_z}) = O\left(\frac{n}{nh^{l_z}}\right) = o(n), \qquad \text{(S.6)}\end{aligned}$$

where these derivations holds under Assumptions 1, 2, 5, and 12. Applying Lemma 3.1 of Zheng (1996) implies that $\hat{\Sigma}_1 = E(H_1(W_i, W_j)) + O_p(n^{-1})$, where

$$\begin{aligned} E[H_1(W_i, W_j)] &= \frac{1}{h^{l_z}} \int K\left(\frac{z_2 - z_1}{h}\right)^2 \sigma^2(z_1)\sigma^2(z_2) f(z_1) f(z_2) dz_1 dz_2 \\ &= \frac{1}{h^{l_z}} \int K\left(\frac{z_2 - z_1}{h}\right)^2 \sigma^2(z_1)\sigma^2(z_2) f(z_1) f(z_2) dz_1 dz_2 \\ &= \int K(u)^2 du \int [\sigma^2(z)]^2 f(z)^2 dz + O(h). \end{aligned}$$

□

S-5 Proof of Lemma 5

Proof. From Assumption 3, there is $\tilde{\theta}$ between θ^* and $\hat{\theta}$ such that

$$\begin{aligned} A_2 &= \frac{2nh^{l_z/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right) [Y_i - g(X_i, \theta^*)][g(X_j, \theta^*) - g(X_j, \hat{\theta})] \\ &= \frac{2(\hat{\theta} - \theta^*)'}{(n-1)h^{l_z/2}} \sum_{i=1}^{n} \sum_{j \neq i} K\left(\frac{Z_j - Z_i}{h}\right) u_i^* \frac{\partial g(X_j, \tilde{\theta})}{\partial \theta} \\ &\equiv (\hat{\theta} - \theta^*)' \tilde{A}_2 \end{aligned}$$

where $(\hat{\theta} - \theta^*) = O_p(n^{-1/2})$ by Assumption 11. Now, we have

$$\begin{aligned} &var(\tilde{A}_2) \\ &\leq E(\tilde{A}_2^2) \\ &= \frac{4}{(n-1)^2 h^{l_z}} \sum_{i=1}^{n} \sum_{j \neq i} \sum_{k=1}^{n} \sum_{l \neq k} E\left[K\left(\frac{Z_j - Z_i}{h}\right) K\left(\frac{Z_l - Z_k}{h}\right) u_i^* u_k^* \frac{\partial g(X_j, \tilde{\theta})}{\partial \theta} \frac{\partial g(X_l, \tilde{\theta})}{\partial \theta}\right] \\ &= \frac{4}{(n-1)^2 h^{l_z}} \sum_{i} \sum_{j \neq i} \sum_{\substack{k \neq i \\ k \neq j}} \sum_{\substack{l \neq k \\ l \neq i \\ l \neq j}} E\left[K\left(\frac{Z_j - Z_i}{h}\right) K\left(\frac{Z_l - Z_k}{h}\right) u_i^* u_k^* \frac{\partial g(X_j, \tilde{\theta})}{\partial \theta} \frac{\partial g(X_l, \tilde{\theta})}{\partial \theta}\right] \end{aligned}$$

$$
\begin{aligned}
&+\frac{8}{(n-1)^2h^{l_z}}\sum_i\sum_{j\neq i}\sum_{\substack{l\neq j\\ l\neq i}}E\left[K\left(\frac{Z_j-Z_i}{h}\right)K\left(\frac{Z_l-Z_j}{h}\right)u_i^*u_j^*\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta}\frac{\partial g(X_l,\tilde{\theta})}{\partial\theta}\right]\\
&+\frac{8}{(n-1)^2h^{l_z}}\sum_i\sum_{j\neq i}\sum_{\substack{k\neq i\\ k\neq j}}E\left[K\left(\frac{Z_j-Z_i}{h}\right)K\left(\frac{Z_i-Z_k}{h}\right)u_i^*u_k^*\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta}\frac{\partial g(X_i,\tilde{\theta})}{\partial\theta}\right]\\
&+\frac{4}{(n-1)^2h^{l_z}}\sum_i\sum_{j\neq i}E\left[K\left(\frac{Z_j-Z_i}{h}\right)^2u_i^{*2}\left|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta}\right|^2\right]\\
&+\frac{4}{(n-1)^2h^{l_z}}\sum_i\sum_{j\neq i}E\left[K\left(\frac{Z_j-Z_i}{h}\right)u_i^*\left|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta}\right|\right]^2\\
&=\frac{4n(n-2)(n-3)}{(n-1)h^{l_z}}E\left[K\left(\frac{Z_j-Z_i}{h}\right)K\left(\frac{Z_l-Z_k}{h}\right)\delta_{\theta^*}(Z_i)\delta_{\theta^*}(Z_k)\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta}\frac{\partial g(X_l,\tilde{\theta})}{\partial\theta}\right]+R_n,
\end{aligned}
\tag{S.7}
$$

where R_n represents small terms. Thus,

$$
\begin{aligned}
&|E(\tilde{A}_2^2)|\\
&\leq\frac{4n(n-2)(n-3)}{(n-1)h^{l_z}}E\left[\left|K\left(\frac{Z_j-Z_i}{h}\right)\right|\left|K\left(\frac{Z_l-Z_k}{h}\right)\right||\delta_{\theta^*}(Z_i)|\right.\\
&\qquad\qquad\left.|\delta_{\theta^*}(Z_k)|\sup_{\theta\in\Theta}\left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\|\sup_{\theta\in\Theta}\left\|\frac{\partial g(X_l,\theta)}{\partial\theta}\right\|\right]+R_n\\
&=\frac{4n(n-2)(n-3)}{(n-1)h^{l_z}}\left\{E\left[\left|K\left(\frac{Z_j-Z_i}{h}\right)\right||\delta_{\theta^*}(Z_i)|\sup_{\theta\in\Theta}\left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\|\right]\right\}^2+R_n.
\end{aligned}
\tag{S.8}
$$

From Assumptions 2, 5, 8, and 12, we obtain

$$
\begin{aligned}
&E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\|\int\left|K\left(\frac{Z_j-z}{h}\right)\right||\delta_{\theta^*}(z)|\,f(z)dz\right]\\
&=h^{l_z}E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\|\int|K(u)||\delta_{\theta^*}(Z_j-uh)|f(Z_j-uh)du\right]\\
&=h^{l_z}E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\||\delta_{\theta^*}(Z_j)|f(Z_j)\int|K(u)|du\right]+R_n\\
&\leq h^{l_z}\overline{f}\left\{E\left[\sup_{\theta\in\Theta}\left\|\frac{\partial g(X_j,\theta)}{\partial\theta}\right\|^2\right]\right\}^{1/2}\left\{E\left[\delta_{\theta^*}(Z_j)^2\right]\right\}^{1/2}+R_n
\end{aligned}
$$

$$= O(h^{l_z}) \left\{ E\left[\delta_{\theta^*}(Z_j)^2 \right] \right\}^{1/2}. \tag{S.9}$$

Thus, we obtain $E(\tilde{A}_2^2) = O(n^2 h^{l_z}) E[\delta^2_{\theta^*}(Z_i)]$. Thus, $\tilde{A}_2 = O_p(nh^{l_z/2})\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2}$, which implies $A_2 = O_p(\sqrt{nh^{l_z}})\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2}$.

Next, we show that $A_3 = o_p(A_2)$. From Assumption 3, $\tilde{\theta}_1$ and $\tilde{\theta}_2$ exist such that

$$\begin{aligned}
A_3 &= \frac{nh^{l_z/2}}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right) [g(X_i, \theta^*) - g(X_i, \hat{\theta})][g(X_j, \theta^*) - g(X_j, \hat{\theta})] \\
&= \frac{(\hat{\theta} - \theta^*)'}{(n-1)h^{l_z/2}} \sum_{i=1}^{n} \sum_{j \neq i} K\left(\frac{Z_j - Z_i}{h}\right) \frac{\partial g(X_i, \tilde{\theta}_1)}{\partial \theta} \frac{\partial g(X_j, \tilde{\theta}_2)}{\partial \theta'} (\hat{\theta} - \theta^*) \\
&\equiv (\hat{\theta} - \theta^*)' \tilde{A}_3 (\hat{\theta} - \theta^*).
\end{aligned} \tag{S.10}$$

We have

$$\begin{aligned}
E\|\tilde{A}_3\| &= nh^{-l_z/2} E\left\| K\left(\frac{Z_j - Z_i}{h}\right) \frac{\partial g(X_i, \tilde{\theta}_1)}{\partial \theta} \frac{\partial g(X_j, \tilde{\theta}_2)}{\partial \theta'} \right\| \\
&\leq nh^{-l_z/2} E\left[\left| K\left(\frac{Z_j - Z_i}{h}\right) \right|^2 \right]^{1/2} E\left[\left\| \frac{\partial g(X_i, \tilde{\theta})}{\partial \theta} \right\|^2 \left\| \frac{\partial g(X_j, \tilde{\theta})}{\partial \theta'} \right\|^2 \right]^{1/2} \\
&\leq O(nh^{-l_z/2}) \left[\int \left| K\left(\frac{z_1 - z_2}{h}\right) \right|^2 f(z_1) f(z_2) dz_1 dz_2 \right]^{1/2} \\
&= n \left[\int |K(u)|^2 f(z_2 + uh) f(z_2) du dz_2 \right]^{1/2} \\
&= O(n),
\end{aligned} \tag{S.11}$$

where $E\left[\sup_{\theta \in \Theta} \|\frac{\partial}{\partial \theta} g(X, \theta)\|^2\right]$ is bounded by Assumption 5 and the last equality holds by Assumptions 2 and 12. Since $(\hat{\theta} - \theta^*) = O_p(n^{-1/2})$ by Assumption 11, we obtain $A_3 = O_p(1)$. Thus, we have $A_2 + A_3 = O_p(\sqrt{nh^{l_z}})\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2} + O_p(1)$.

□

S-6 Proof of Lemma 6

Proof. It is obvious that $\bar{\Sigma}$ is bounded uniformly in $m \in \mathcal{M}(\kappa\rho_n)$ by Assumptions 1, 2, 7, and 12. Let $u_i^* \equiv Y_i - g(X_i, \theta^*)$.

$$
\begin{aligned}
\hat{\Sigma} &= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 \hat{u}_i^2 \hat{u}_j^2 \\
&= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 [Y_i - g(X_i, \hat{\theta})]^2 [Y_j - g(X_j, \hat{\theta})]^2 \\
&= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 [g(X_i, \theta^*) - g(X_i, \hat{\theta}) + u_i^*]^2 [g(X_j, \theta^*) - g(X_j, \hat{\theta}) + u_j^*]^2 \\
&= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 \Big\{ [g(X_i, \theta^*) - g(X_i, \hat{\theta})]^2 [g(X_j, \theta^*) - g(X_j, \hat{\theta})]^2 \\
&+ 4u_i^* [g(X_i, \theta^*) - g(X_i, \hat{\theta})][g(X_j, \theta^*) - g(X_j, \hat{\theta})]^2 + 2u_i^{*2} [g(X_j, \theta^*) - g(X_j, \hat{\theta})]^2 \\
&+ 4u_i^* u_j^{*2} [g(X_j, \theta^*) - g(X_j, \hat{\theta})] + 4u_i^* u_j^* [g(X_i, \theta^*) - g(X_i, \hat{\theta})][g(X_j, \theta^*) - g(X_j, \hat{\theta})] + u_i^{*2} u_j^{*2} \Big\} \\
&= \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 u_i^{*2} u_j^{*2} \\
&+ \frac{8}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 u_i^* u_j^{*2} [g(X_i, \theta^*) - g(X_i, \hat{\theta})] + R_n, \\
&\equiv \hat{\Sigma}_1^* + \hat{\Sigma}_2^* + R_n,
\end{aligned}
$$

where R_n represents smaller terms that is $o_p(1)$, which can be shown by using Lemma S.14 under Assumptions 1, 3, 5, 11, and 12.

Now, we show that $\hat{\Sigma}_2^* = o_p(1)$. From Assumption 3, there is $\tilde{\theta}$ between $\hat{\theta}$ and θ^* such that

$$
\hat{\Sigma}_2^* = \frac{8(\hat{\theta} - \theta^*)'}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{h^{l_z}} K\left(\frac{Z_j - Z_i}{h}\right)^2 u_i^* u_j^{*2} \frac{\partial g(X_i, \tilde{\theta})}{\partial \theta},
$$

where $\sqrt{n}(\hat{\theta} - \theta^*) = O_p(n^{-1/2})$ by Assumption 11. Since $E(u_i^{*2}|Z_i)$ is bounded from

Assumptions 1 and 7 and $E\|\frac{\partial}{\partial\theta}g(X_i,\tilde{\theta})\|^2 \leq E[\sup_{\theta\in\Theta}\|\frac{\partial}{\partial\theta}g(X_i,\tilde{\theta})\|^2]$ is bounded by Assumption 5, there is a constant C such that

$$E\left\|K\left(\frac{Z_j-Z_i}{h}\right)^2 u_i^* u_j^{*2}\frac{\partial g(X_i,\tilde{\theta})}{\partial\theta}\right\| \leq \left[E\left|K\left(\frac{Z_j-Z_i}{h}\right)^4 u_i^{*2}E(u_j^{*2}|Z_j)^2\right|\right]^{1/2}\left[E\left\|\frac{\partial g(X_i,\tilde{\theta})}{\partial\theta}\right\|^2\right]^{1/2}$$
$$\leq C\left[E\left|K\left(\frac{Z_j-Z_i}{h}\right)^4 u_i^{*2}\right|\right]^{1/2},$$

where

$$\begin{aligned}\int K\left(\frac{z_2-z_1}{h}\right)^4\sigma_{\theta^*}^2(z_1)f(z_1)f(z_2)dz_1dz_2 &= h^{l_z}\int K(u)^4\sigma_{\theta^*}^2(z_1)f(z_1)f(z_1+uh)dz_1du\\ &= h^{l_z}\int K(u)^4duE[\sigma_{\theta^*}^2(z_1)f(z_1)]+o(1)\\ &= O(h^{l_z}),\end{aligned}$$

where the last equality holds from Assumptions 1, 2, 7, and 12. Thus, $\hat{\Sigma}_2^* = O_p(n^{-1/2})$. Now, similar to the proof of Lemma 4, we can show that $\hat{\Sigma}_1^*$ is a second order U-statistic with $E[H_1^*(W_i,W_j)^2] = o(n)$, where $H_1^*(W_i,W_j) \equiv \frac{1}{h^{l_z}}K\left(\frac{Z_j-Z_i}{h}\right)^2 u_i^{*2}u_j^{*2}$. Thus, we apply Lemma 3.1 of Zheng (1996), implying that $\hat{\Sigma}_1 = E[H_1^*(W_i,W_j)] + o_p(1)$, where

$$\begin{aligned}E[H_1^*(W_i,W_j)] &= \frac{1}{h^{l_z}}\int K\left(\frac{z_2-z_1}{h}\right)^2\sigma_{\theta^*}^2(z_1)\sigma_{\theta^*}^2(z_2)f(z_1)f(z_2)dz_1dz_2\\ &= \int K(u)^2\sigma_{\theta^*}^2(z_1)\sigma_{\theta^*}^2(z_1+uh)f(z_1)f(z_1+uh)dz_1du\\ &= \int K(u)^2du\int[\sigma_{\theta^*}^2(z_1)]^2f(z_1)^2dz_1+o(1).\end{aligned}$$

□

S-7 Proof of Lemma 7

Proof. First, consider the case with $l_z = 1$. Taylor expansion of $k((Z_j - Z_i)/\hat{h})$ at $\hat{h} = h_0$ yields

$$\begin{aligned}
k\left(\frac{Z_j - Z_i}{\hat{h}}\right) &= k\left(\frac{Z_j - Z_i}{h_0}\right) - \frac{\hat{h} - h_0}{h_0}\frac{Z_j - Z_i}{h_0}k^{(1)}\left(\frac{Z_j - Z_i}{h_0}\right) \\
&\quad - \frac{1}{2!}\left(\frac{\hat{h} - h_0}{h_0}\right)^2\left[2\frac{Z_j - Z_i}{h_0}k^{(1)}\left(\frac{Z_j - Z_i}{h_0}\right) + \left(\frac{Z_j - Z_i}{h_0}\right)^2 k^{(2)}\left(\frac{Z_j - Z_i}{h_0}\right)\right] \\
&\quad - \frac{1}{3!}\left(\frac{\hat{h} - h_0}{h_0}\right)^3\left[6\frac{Z_j - Z_i}{h_0}k^{(1)}\left(\frac{Z_j - Z_i}{h_0}\right) + 2\left(\frac{Z_j - Z_i}{h_0}\right)^2 k^{(2)}\left(\frac{Z_j - Z_i}{h_0}\right)\right. \\
&\quad \left. +4\left(\frac{Z_j - Z_i}{h_0}\right)^2 k^{(2)}\left(\frac{Z_j - Z_i}{h_0}\right) + 2\left(\frac{Z_j - Z_i}{h_0}\right)^3 k^{(3)}\left(\frac{Z_j - Z_i}{h_0}\right)\right] \\
&\quad \vdots
\end{aligned} \tag{S.12}$$

which can be described by

$$\begin{aligned}
k\left(\frac{Z_j - Z_i}{\hat{h}}\right) &= k\left(\frac{Z_j - Z_i}{h_0}\right) + \sum_{s=1}^{m-1}\frac{1}{s!}\left(\frac{\hat{h} - h_0}{h_0}\right)^s \tilde{k}^{(s)}\left(\frac{Z_j - Z_i}{h_0}\right) \\
&\quad + \frac{1}{m!}\left(\frac{\hat{h} - h_0}{\tilde{h}}\right)^m \tilde{k}^{(m)}\left(\frac{Z_j - Z_i}{\tilde{h}}\right),
\end{aligned} \tag{S.13}$$

where $\tilde{h}$ is between $\hat{h}$ and h_0 and

$$\tilde{k}^{(s)}\left(\frac{Z_j - Z_i}{h}\right) \equiv h^s\frac{\partial^s}{\partial^s h}k\left(\frac{Z_j - Z_i}{h}\right) = \sum_{l=1}^{s} c_l\left(\frac{Z_j - Z_i}{h}\right)^l k^{(l)}\left(\frac{Z_j - Z_i}{h}\right), \tag{S.14}$$

for some constant c_l.

A similar results hold for $l_z > 1$, which is

$$K\left(\frac{Z_j - Z_i}{\hat{h}}\right) = K\left(\frac{Z_j - Z_i}{h_0}\right) + \sum_{s=1}^{m-1}\frac{1}{s!}\left(\frac{\hat{h} - h_0}{h_0}\right)^s \tilde{K}^{(s)}\left(\frac{Z_j - Z_i}{h_0}\right)$$

$$+\frac{1}{m!}\left(\frac{\hat{h}-h_0}{\tilde{h}}\right)^m \tilde{K}^{(m)}\left(\frac{Z_j-Z_i}{\tilde{h}}\right), \tag{S.15}$$

where

$$\begin{aligned}
\tilde{K}^{(s)}\left(\frac{Z_j-Z_i}{h}\right) &\equiv h^s \frac{\partial^s}{\partial h^s} K\left(\frac{Z_j-Z_i}{h}\right) \\
&= \sum_{1\le l_1+\cdots+l_{l_z}\le s} \prod_{k=1}^{l_z} c_{l_k}\left(\frac{Z_{kj}-Z_{ki}}{h}\right)^{l_k} k^{(l_k)}\left(\frac{Z_{kj}-Z_{ki}}{h}\right) \qquad \text{(S.16)}
\end{aligned}$$

for some constant c_{l_k}. Since $\tilde{K}^{(s)}(v)$ is an even function, $\tilde{K}^{(s)}(v)$ can be viewed as a second-oder kernel function. Thus, the above expansion yields

$$\begin{aligned}
T_n(\hat{h}) &= \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{\hat{h}^{l_z}}K\left(\frac{Z_j-Z_i}{\hat{h}}\right)\hat{u}_i\hat{u}_j \\
&= \frac{h_0^{l_z}}{\hat{h}^{l_z}}T_n(h_0) + \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{\hat{h}^{l_z}}\hat{u}_i\hat{u}_j\sum_{s=1}^{m-1}\frac{1}{s!}\left(\frac{\hat{h}-h_0}{h_0}\right)^s\tilde{K}^s\left(\frac{Z_j-Z_i}{h_0}\right) \\
&+ \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{\hat{h}^{l_z}}\hat{u}_i\hat{u}_j\frac{1}{m!}\left(\frac{\hat{h}-h_0}{\tilde{h}}\right)^m\tilde{K}^m\left(\frac{Z_j-Z_i}{\tilde{h}}\right) \\
&= \frac{h_0^{l_z}}{\hat{h}^{l_z}}T_n(h_0) + \left(\frac{\hat{h}-h_0}{h_0}\right)O_p(T_n(h_0)) \\
&+ \frac{1}{m!}\left(\frac{\hat{h}-h_0}{\tilde{h}}\right)^m\frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j\neq i}\frac{1}{\hat{h}^{l_z}}\hat{u}_i\hat{u}_j\tilde{K}^m\left(\frac{Z_j-Z_i}{\tilde{h}}\right) \\
&= \frac{h_0^{l_z}}{\hat{h}^{l_z}}T_n(h_0) + \left(\frac{\hat{h}-h_0}{h_0}\right)O_p(T_n(h_0)) + \frac{1}{\hat{h}^{l_z}}O_p\left(\left(\frac{\hat{h}-h_0}{h_0}\right)^m\right) \qquad \text{(S.17)}
\end{aligned}$$

because $\hat{h}=h_0+o_p(h_0)$, $\hat{h}/h_0-1=o_p(1)$, and $E[|\hat{u}_i\hat{u}_j||\tilde{K}^m((Z_j-Z_i)/\tilde{h})|]\le E[E(|\hat{u}_i||Z_i)]^2\le E[E(|u_i^*||Z_i)]^2<\infty$ under Assumptions 1, 5, 7, 11, and the assumption that $\sup_u|\tilde{K}^m(u)|<\infty$, where u_i^* is defined in Lemma 6. Since $\hat{h}^{-l_z}\left(\hat{h}/h_0-1\right)^m=o_p(1)$ by Assumption and $T_n(h_0) = O_p(1)$, we obtain

$$T_n(\hat{h}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{\hat{h}^{l_z}} K\left(\frac{Z_j - Z_i}{\hat{h}}\right) \hat{u}_i \hat{u}_j = \frac{h_0^{l_z}}{\hat{h}^{l_z}} T_n(h_0) + o_p(T_n(h_0)). \quad \text{(S.18)}$$

□

S-8 Proof of Lemma 8

Proof. By the expansion of the kernel given in equation (S.13), we obtain

$$\begin{aligned}
\hat{\Sigma}(\hat{h}) =& \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{\hat{h}^{l_z}} \hat{u}_i^2 \hat{u}_j^2 \left[K\left(\frac{Z_j - Z_i}{h_0}\right) \right. \\
&+ \left. \sum_{s=1}^{m-1} \frac{1}{s!} \left(\frac{\hat{h} - h_0}{h_0}\right)^s \tilde{K}^s \left(\frac{Z_j - Z_i}{h_0}\right) + \frac{1}{m!} \left(\frac{\hat{h} - h_0}{\tilde{h}}\right)^m \tilde{K}^m \left(\frac{Z_j - Z_i}{\tilde{h}}\right) \right]^2 \\
=& \frac{h_0^{l_z}}{\hat{h}^{l_z}} \hat{\Sigma}(h_0) + \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{\hat{h}^{l_z}} \hat{u}_i^2 \hat{u}_j^2 \left[\sum_{s=1}^{m-1} \frac{1}{s!} \left(\frac{\hat{h} - h_0}{h_0}\right)^s \tilde{K}^s \left(\frac{Z_j - Z_i}{h_0}\right) \right]^2 \\
&+ \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{\hat{h}^{l_z}} \hat{u}_i^2 \hat{u}_j^2 \left[\frac{1}{m!} \left(\frac{\hat{h} - h_0}{\tilde{h}}\right)^m \tilde{K}^m \left(\frac{Z_j - Z_i}{\tilde{h}}\right) \right]^2 \\
&+ \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{\hat{h}^{l_z}} \hat{u}_i^2 \hat{u}_j^2 \left[K\left(\frac{Z_j - Z_i}{h_0}\right) \sum_{s=1}^{m-1} \frac{1}{s!} \left(\frac{\hat{h} - h_0}{h_0}\right)^s \tilde{K}^s \left(\frac{Z_j - Z_i}{h_0}\right) \right] \\
&+ \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{\hat{h}^{l_z}} \hat{u}_i^2 \hat{u}_j^2 \left[K\left(\frac{Z_j - Z_i}{h_0}\right) \frac{1}{m!} \left(\frac{\hat{h} - h_0}{\tilde{h}}\right)^m \tilde{K}^m \left(\frac{Z_j - Z_i}{\tilde{h}}\right) \right] \\
&+ \frac{4}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{1}{\hat{h}^{l_z}} \hat{u}_i^2 \hat{u}_j^2 \left[\sum_{s=1}^{m-1} \frac{1}{s!m!} \left(\frac{\hat{h} - h_0}{h_0}\right)^{s+m} \tilde{K}^s \left(\frac{Z_j - Z_i}{h_0}\right) \tilde{K}^m \left(\frac{Z_j - Z_i}{\tilde{h}}\right) \right].
\end{aligned} \tag{S.19}$$

Under Assumptions 1, 5, 7, 11, and the assumption that $\sup_u |\tilde{K}^m(u)| < \infty$, we have $E[\hat{u}_i^2 \hat{u}_j^2 \tilde{K}^m((Z_j - Z_i)\tilde{h})^2] \leq E[\hat{u}_i^2]^2 \leq E[E(u_i^{*2}|Z_i)]^2 < \infty$. Since $\tilde{K}^s(v)$ is a second-oder kernel function, the above expansion yields

$$\begin{aligned}
\hat{\Sigma}(\hat{h}) =& \frac{h_0^{l_z}}{\hat{h}^{l_z}} \hat{\Sigma}(h_0) + \left(\frac{\hat{h} - h_0}{h_0}\right)^2 O_p\left(\hat{\Sigma}(h_0)\right) + \frac{1}{\hat{h}^{l_z}} O_p\left(\left(\frac{\hat{h} - h_0}{\tilde{h}}\right)^{2m}\right) \\
&+ \left(\frac{\hat{h} - h_0}{h_0}\right) O_p\left(\hat{\Sigma}(h_0)\right) + \left(\frac{\hat{h} - h_0}{h_0}\right)^m O_p\left(\hat{\Sigma}(h_0)\right) + \frac{1}{\hat{h}^{l_z}} O_p\left(\left(\frac{\hat{h} - h_0}{\tilde{h}}\right)^{m+1}\right) \\
=& \frac{h_0^{l_z}}{\hat{h}^{l_z}} \hat{\Sigma}(h_0) + o_p\left(\hat{\Sigma}(h_0)\right)
\end{aligned} \tag{S.20}$$

because $\frac{1}{\hat{h}^{l_z}}(\hat{h}/h_0 - 1)^{m+1} = o_p(1)$ and $\hat{\Sigma}(h_0) = \bar{\Sigma} + o_p(1)$ by Lemma 6. Thus, we obtain $\hat{\Sigma}(\hat{h}) = \bar{\Sigma} + o_p\,(1)$.

$\square$

Supplemental Material: Lemmas

S-1 Lemma S.15

Lemma S.15. *Under Assumptions 2, 8, and 12, we obtain* $\int K\left(\frac{x-y}{h}\right)\delta_{\theta^*}(y)f(y)dy = h^{l_z}\delta_{\theta^*}(x)f(x) + O(h^{l_z+\min\{q_k,k+q_z\}})$.

Proof. Note that $f(\cdot)$ and $\delta_{\theta^*}(\cdot)$ are q_z- and k-times differentiable by Assumptions 2 and 8, respectively. For any functions $g(\cdot):\mathbb{R}^l \to \mathbb{R}$, let $\Delta^{(q)}g(\cdot)$ indicates a vector (matrix, or cube) of q-times partial derivatives.[10] We define l_z dimensional vector $u = \{u_1, u_2, \ldots, u_{l_z}\}'$. Then,

$$\begin{aligned}
&\int K\left(\frac{x-y}{h}\right)\delta_{\theta^*}(y)f(y)dy = h^{l_z}\int K(u)\delta_{\theta^*}(x-uh)f(x-uh)du \\
&= h^{l_z}\int K(u)\left[\delta_{\theta^*}(x) - hu'\Delta^{(1)}\delta_{\theta^*}(x) + \frac{h^2}{2}u'\Delta^{(2)}\delta_{\theta^*}(x)u + \ldots\right] \\
&\qquad \left[f(x) - hu'\Delta^{(1)}f(x) + \frac{h^2}{2}u'\Delta^{(2)}f(x)u + \ldots\right]du \\
&= h^{l_z}\delta_{\theta^*}(x)f(x)\int K(u)du + O(h^{l_z+\min\{q_k,k+q_z\}}),
\end{aligned}$$

where $\int K(u)du = 1$ by Assumption 12. The last equation holds by the feature of q_kth oder kernel in Assumption 12. □

S-2 Lemma S.16

Lemma S.16. *Under Assumptions 1, 2, 3, 4, 5, 6, 8, and 11, we have*

$$\begin{aligned}
A &= \frac{2}{\sqrt{2}K_n^{l_z/2}}\sum_{k\in\mathcal{K}}\frac{\mathbb{1}\{N_k>1\}}{N_k}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}[Y_i - g(X_i,\theta^*)][g(X_j,\theta^*) - g(X_j,\hat{\theta})] \\
&= O(\sqrt{nh^{l_z}})\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2} + O(h^{l_z/2}).
\end{aligned}$$

[10] For example, $\Delta^{(1)}f(z) = \{\partial f(z)/\partial z_1, \partial f(z)/\partial z_2, \ldots, \partial f(z)/\partial z_{l_z}\}'$, and $\Delta^{(2)}f(z)$ is a l_z by l_z matrix whose (l, k) element is $\partial f(z)/\partial z_l \partial z_k$.

Proof. Let $\overline{\eta}_i \equiv E[g(X_i,\theta^*)|Z_i] - g(X_i,\theta^*) + \omega_i$, where $E(\overline{\eta}_i) = 0$ and $var(\overline{\eta}_i) = E[g(X_i,\theta^*)^2] + E(\omega_i^2) - E\{E[g(X_i,\theta^*)|Z_i]^2\} - 2E[\omega_i g(X_i,\theta^*)] < \infty$ by Assumptions 1, 2, 4, and 8, which implies $\overline{\eta}_i = O_p(1)$. Then,

$$
\begin{aligned}
A &= \frac{2}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} [m(Z_i) - g(X_i,\theta^*) + \omega_i][g(X_j,\theta^*) - g(X_j,\hat{\theta})] \\
&= \frac{2}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} [\delta_{\theta^*}(Z_i) + \overline{\eta}_i][g(X_j,\theta^*) - g(X_j,\hat{\theta})] \\
&= \frac{2}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \delta_{\theta^*}(Z_i)[g(X_j,\theta^*) - g(X_j,\hat{\theta})] \\
&+ \frac{2}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} \frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} \overline{\eta}_i[g(X_j,\theta^*) - g(X_j,\hat{\theta})] \\
&\equiv A_1 + A_2.
\end{aligned}
$$

By Assumptions 3 and 11, there is $\tilde{\theta}$ between θ^* and $\hat{\theta}$ such that $|g(X_j,\theta^*) - g(X_j,\hat{\theta})| = \left|\frac{\partial}{\partial\theta} g(X_j,\tilde{\theta})\right| O_p(n^{-1/2})$. Thus, for some positive constant C and C', we have

$$
\begin{aligned}
&E|A_1| \\
&\leq \frac{2}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} E\left[\frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} |\delta_{\theta^*}(Z_i)||g(X_j,\theta^*) - g(X_j,\hat{\theta})|\right] \\
&\leq \frac{2C}{\sqrt{2}K_n^{l_z/2}n^{1/2}} \sum_{k\in\mathcal{K}} E\left[\frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} |\delta_{\theta^*}(Z_i)| \left|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta}\right|\right] + o(1) \\
&= \frac{2C}{\sqrt{2}K_n^{l_z/2}n^{1/2}} \sum_{k\in\mathcal{K}} E\Bigg\{\frac{\mathbb{1}\{N_k>1\}}{N_k} \sum_{\{Z_i,Z_j\}\in I_k, i\neq j} E\left[|\delta_{\theta^*}(Z_i)|\,|\, Z_i\in I_k\right] \\
&\qquad E\left[\left|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta}\right|\,\middle|\, Z_j\in I_k\right]\Bigg\} + o(1) \\
&= \frac{2C}{\sqrt{2}K_n^{l_z/2}n^{1/2}} \sum_{k\in\mathcal{K}} E\left[(N_k-1)\mathbb{1}\{N_k>1\}\right] E\left[|\delta_{\theta^*}(Z_i)|\,|\, Z_i\in I_k\right] E\left[\left|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta}\right|\,\middle|\, Z_j\in I_k\right] + o(1) \\
&\leq \frac{2C'n^{1/2}}{\sqrt{2}K_n^{l_z/2}} \sum_{k\in\mathcal{K}} P(Z_j\in I_k) E\left[|\delta_{\theta^*}(Z_i)|\,|\, Z_i\in I_k\right] + o(1)
\end{aligned}
$$

$$\leq \frac{2C'n^{1/2}}{\sqrt{2}K_n^{l_z/2}}\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2} + o(1) = O(\sqrt{nh^{l_z}})\{E[\delta^2_{\theta^*}(Z_i)]\}^{1/2},$$

where $E[|\frac{\partial}{\partial\theta}g(X_j,\tilde{\theta})||Z_j \in I_k]$ is bounded by Assumption 5. Next,

$$\begin{aligned}
A_2 &= \frac{2(\theta^*-\hat{\theta})'}{\sqrt{2}K_n^{l_z/2}}\sum_{k\in\mathcal{K}}\frac{\mathbb{1}\{N_k>1\}}{N_k}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}\overline{\eta}_i\frac{\partial g(X_j,\theta)}{\partial\theta} \\
&\quad + \frac{(\theta^*-\hat{\theta})'}{\sqrt{2}K_n^{l_z/2}}\sum_{k\in\mathcal{K}}\frac{\mathbb{1}\{N_k>1\}}{N_k}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}\overline{\eta}_i\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}(\theta^*-\hat{\theta}) \\
&\equiv (\theta^*-\hat{\theta})'A_2' + (\theta^*-\hat{\theta})'A_2''(\theta^*-\hat{\theta}) \qquad \text{(S.1)}
\end{aligned}$$

Since $E(\overline{\eta}_i|Z_i)=0$, we have $E(A_2')=0$. For some constant C,

$$\begin{aligned}
E(A_2'^2) &= \frac{2}{K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left\{\frac{\mathbb{1}\{N_k>1\}}{N_k^2}\left[\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}\overline{\eta}_i\frac{\partial g(X_j,\theta)}{\partial\theta}\right]^2\right\} \\
&= \frac{2}{K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left\{\frac{\mathbb{1}\{N_k>1\}}{N_k^2}\left[\sum_{\{Z_i,Z_j,Z_l\}\in I_k, i\neq j, i\neq l}\overline{\eta}_i^2\frac{\partial g(X_j,\theta)}{\partial\theta}\frac{\partial g(X_l,\theta)}{\partial\theta'}\right]\right\} \\
&\leq \frac{2C}{K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left[\frac{(N_k-2)^2\mathbb{1}\{N_k>1\}}{N_k}\right]E(\overline{\eta}_i^2|Z_i\in I_k) \\
&\leq \frac{2C}{K_n^{l_z}}\sum_{k\in\mathcal{K}}E\left[N_k\mathbb{1}\{N_k>1\}\right]E(\overline{\eta}_i^2|Z_i\in I_k) \\
&\leq \frac{2nC}{K_n^{l_z}}\sum_{k\in\mathcal{K}}P(Z_i\in I_k)E(\overline{\eta}_i^2|Z_i\in I_k) = 2nCh^{l_z}E(\overline{\eta}_i^2) = O(nh^{l_z}), \qquad \text{(S.2)}
\end{aligned}$$

by the boundedness in Assumption 5 and boundedness of $E(\overline{\eta}_i^2|Z_i)$ as shown before. For some constant $C>0$,

$$\begin{aligned}
E|A_2''| &\leq \frac{1}{\sqrt{2}K_n^{l_z/2}}\sum_{k\in\mathcal{K}}E\left[\frac{\mathbb{1}\{N_k>1\}}{N_k}\sum_{\{Z_i,Z_j\}\in I_k, i\neq j}|\overline{\eta}_i|\left|\frac{\partial g(X_j,\tilde{\theta})}{\partial\theta\partial\theta'}\right|\right] \\
&\leq \frac{C}{\sqrt{2}K_n^{l_z/2}}\sum_{k\in\mathcal{K}}E\left[\mathbb{1}\{N_k>1\}(N_k-1)\right]E(|\overline{\eta}_i||Z_i\in I_k)
\end{aligned}$$

$$\leq \frac{nC}{\sqrt{2}K_n^{l_z/2}} \sum_{k \in \mathcal{K}} P(Z_i \in I_k) E(|\overline{\eta}_i||Z_i \in I_k) = \frac{nh^{l_z/2}C}{\sqrt{2}} E(|\overline{\eta}_i|) = O(nh^{l_z/2}), \quad \text{(S.3)}$$

where $E[|\frac{\partial}{\partial\theta\partial\theta'} g(X_j, \tilde{\theta})|Z_j \in I_k]$ is bounded by Assumption 6 and $E(|\overline{\eta}_i||Z_i) < E(|\overline{\eta}_i|^2|Z_i)^{1/2}$ is bounded as shown before. From equations (S.1), (S.2), and (S.3), we obtain $A_2 = O(h^{l_z/2})$.

□

Supplemental Material: Simulation

Table S.1 shows the simulation results with the sample size of $n = 1000$. All simulation settings but sample size are the same with those in Section 5 of the paper.

Table S.1: Size and power of T_n with $n = 1000$.

					Bootstrap		Normal	
H_0	H_1		ρ	η	h_{cv}	h_{opt}	h_{cv}	h_{opt}
H_0 is true								
(6)		DGP 1	0.8	0.1	0.054	0.047	0.020	0.033
		DGP 2	0.8	0.5	0.053	0.046	0.030	0.028
		DGP 3	0.7	0.1	0.052	0.045	0.021	0.031
(7)		DGP 1	0.8	0.1	0.060	0.051	0.022	0.025
		DGP 2	0.8	0.5	0.063	0.051	0.017	0.026
		DGP 3	0.7	0.1	0.061	0.048	0.021	0.027
H_0 is false								
(6)	(7)	DGP 1	0.8	0.1	1.000	1.000	1.000	1.000
		DGP 2	0.8	0.5	1.000	1.000	1.000	1.000
		DGP 3	0.7	0.1	1.000	0.960	1.000	0.936
(6)	(8)	DGP 1	0.8	0.1	1.000	1.000	1.000	1.000
		DGP 2	0.8	0.5	1.000	0.999	1.000	0.996
		DGP 3	0.7	0.1	1.000	0.779	1.000	0.717
(7)	(8)	DGP 1	0.8	0.1	0.999	0.957	0.994	0.938
		DGP 2	0.8	0.5	0.999	0.899	0.987	0.868
		DGP 3	0.7	0.1	0.680	0.418	0.507	0.358

Note: Critical values are obtained from bootstrapping (columns labeled by bootstrap) and the normal distribution (columns labeled by Normal).

Refrerenecs

Abramovich, F., Feis, D., Italia, S., and Theofanis. (2009). Optimal testing for additivity in multiple nonparametric regression. *Annals of the Institute of Statistical Mathematics*, 61(3), 691–714.

Guerre, E., and Lavergne, P. (2002). Optimal minimax rates for nonparametric specification testing in regression models. *Econometric Theory*, 18(5), 1139–1171.

Hall, P., and Heyde, C. C. (1980). *Martingale limit theory and its application.* Academic press.

Ingster, Y. I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I, II, III. *Mathematical Methods of Statistics*, 2(2), 85–114.

Ingster, Y. I., and Sapatinas, T. (2009). Minimax goodness-of-fit testing in multivariate nonparametric regression. *Mathematical Methods of Statistics*, 18(3), 241–269.

Lehmann, E. L., and Romano, J. P. (2005). *Testing statistical hypotheses.* New York, USA: Springer.

Lepski, O. V., and Spokoiny, V. G. (1999). Minimax nonparametric hypothesis testing: the case of an inhomogeneous alternative. *Bernoulli*, 5(2), 333–358.

Lepski, O. V., and Tsybakov, A. (2000). Asymptotically exact nonparametric hypothesis testing in sup-norm and at a fixed point. *Probability Theory and Related Fields*, 117(1), 17–48.

Li, Q., and Racine, J. S. (2007). *Nonparametric econometrics: Theory and practice.* Princeton University Press.

Spokoiny, V. G. (1996). Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, 24(6), 2477–2498.

Zheng, J. X. (1996). A consistent test of functional form via nonparametric estimation techniques. *Journal of Econometrics*, 75(2), 263–289.