Adaptive, Rate-Optimal Hypothesis Testing in Nonparametric IV Models*

CHRISTOPH BREUNIG[†] XIAOHONG CHEN[‡]

First version: August 2018; Latest revision: February 5, 2024

We propose a new adaptive hypothesis test for inequality (e.g., monotonicity, convexity) and equality (e.g., parametric, semiparametric) restrictions on a structural function in a nonparametric instrumental variables (NPIV) model. Our test statistic is based on a modified leave-one-out sample analog of a quadratic distance between the restricted and unrestricted sieve NPIV estimators. We provide computationally simple, data-driven choices of sieve tuning parameters and Bonferroni adjusted chi-squared critical values. Our test adapts to the unknown smoothness of alternative functions in the presence of unknown degree of endogeneity and unknown strength of the instruments. It attains the adaptive minimax rate of testing in L^2. That is, the sum of its type I error uniformly over the composite null and its type II error uniformly over nonparametric alternative models cannot be improved by any other hypothesis test for NPIV models of unknown regularities. Confidence sets in L^2 are obtained by inverting the adaptive test. Simulations confirm that our adaptive test controls size and its finite-sample power greatly exceeds existing non-adaptive tests for monotonicity and parametric restrictions in NPIV models. Empirical applications to test for shape restrictions of differentiated products demand and of Engel curves are presented.

Keywords: Nonparametric instrumental variables; Shape restrictions; Composite hypothesis; Nonparametric alternatives; Minimax rate of testing; Adaptive hypothesis testing; Random exponential scan; Sieve regularization; Sieve U statistics.

*We thank the Co-Editor, three anonymous referees and Denis Chetverikov for very constructive comments. Earlier versions have been presented at numerous workshops and conferences since February 2019. We thank Don Andrews, Tim Armstrong, Tim Christensen, Giovanni Compiani, Enno Mammen, Peter Mathe and other participants at various meetings for helpful comments. The empirical result reported in Section 6.1 is our own analyses using data provided through the Nielsen Datasets (from The Nielsen Company (US), LLC) at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. Nielsen is not responsible for, had no role in, and was not involved in analyzing and preparing the empirical findings reported herein.

†Department of Economics, University of Bonn, Adenauerallee 24-26, 53113 Bonn, Germany. Email: cbreunig@uni-bonn.de

‡Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520, USA. Email: xiaohong.chen@yale.edu

1. Introduction

In this paper, we propose computationally simple, optimal hypothesis testing in a nonparametric instrumental variables (NPIV) model. The maintained assumption is that there is a nonparametric structural function h satisfying the NPIV model

$$\mathrm{E}[Y - h(X)|W] = 0, \tag{1.1}$$

where X is a d_x-dimensional vector of possibly endogenous regressors, W is a d_w-dimensional vector of conditional (instrumental) variables (with $d_w \geq d_x$), and the joint distribution of (Y, X, W) is unspecified beyond (1.1). With the danger of abusing terminology, we call a function h satisfying model (1.1) a NPIV function. We are interested in testing a null hypothesis that a NPIV function h satisfies some simplifying economic restrictions, such as parametric or semiparametric equality restrictions or inequality restrictions (e.g., nonnegativity, monotonicity, convexity or supermodularity). Our new test builds on a simple data-driven choice of tuning parameters that ensures asymptotic size control and non-trivial power uniformly against a large class of nonparametric alternatives.

Before presenting the theoretical properties of our new test, we derive the *minimax rate of testing* in L^2, which is the fastest rate of separation in root-mean squared distance between the null hypothesis and the nonparametric alternatives that enables consistent testing uniformly over the latter. We establish the minimax result in two steps: First, we derive, for all possible tests, a lower bound for the type I error uniformly over distributions satisfying the null hypothesis and the type II error uniformly over the nonparametric alternative NPIV functions separated from the null hypothesis by a rate r_n that shrinks to zero as the sample size n goes to infinity. Thus, there exists no other test that provides a better performance with respect to the sum of those errors. Second, we propose a test whose sum of the type I and the type II errors are bounded from above (by the nominal level) at the same separation rate r_n. This test is based on a modified leave-one-out sample analog of a quadratic distance between the restricted and unrestricted sieve NPIV estimators of h. The test is shown to attain the minimax rate of testing r_n when the sieve dimension is chosen optimally according to the smoothness of the nonparametric alternative functions and the degree of the ill-posedness of the NPIV model (that depends on the smoothness of the conditional density of X given W). We call this test minimax rate-optimal (with known model regularities).

In practice, the smoothness of the nonparametric alternative functions and the degree of the ill-posedness of the NPIV model are both unknown. Our new test is a data-driven version of the minimax rate-optimal test that adapts to the unknown smoothness of the nonparametric alternative NPIV functions in the presence of the unknown degree of the ill-posedness. Our test rejects the null hypothesis as soon as there is a sieve dimension (say the smallest sieve dimension) in an estimated index set such that the corresponding normalized quadratic distance estimator exceeds one; and fails to reject the null otherwise. The normalization builds on Bonferroni corrected chi-squared critical values. The simple Bonferroni correction is computed using the cardinality of the estimated index set, which is in turn determined by a random exponential scan (RES) procedure that automatically takes into account the unknown degree of ill-posedness.

We show that our new test attains the minimax rate of testing in L^2 for severely ill-posed NPIV models, and is up to a $\sqrt{\log\log(n)}$ multiplicative factor of the minimax rate of testing for mildly ill-posed NPIV models. This extra $\sqrt{\log\log(n)}$ term is the necessary price to pay for adaptivity to unknown smoothness of nonparametric alternative functions.[1] A key technical part to establish this rate optimality in L^2 testing is to derive a tight upper bound on the convergence rate of a leave-one-out sieve estimator of a quadratic functional of a NPIV function h. We show that our adaptive test has asymptotic size control under composite null by deriving a tight, slowly divergent lower bound for Bonferroni corrected chi-squared critical values. By inverting our adaptive tests we obtain L^2-confidence sets on restricted structural functions. These confidence sets are free of additional choices of tuning parameters. The adaptive minimax rate of testing determines the L^2 radius of the confidence sets.

In Monte Carlo simulations, we analyze the finite sample properties of our adaptive test for the null of monotonicity or a parametric hypothesis using various simulation designs. Our simulations reveal the following patterns: First, our adaptive test delivers adequate size control under different composite null hypotheses and for varying strengths of the instruments. Second, our adaptive test is powerful in comparison to existing tests when alternative functions are relatively simple. Moreover, the finite-sample power of our adaptive test greatly exceeds that of existing tests when alternative functions become more nonlinear. The great power gains of our adaptive test are present even for relatively weak instrument strength and small sample sizes. This highlights the importance of our data-driven choice of the sieve dimension to ensure powerful performance uniformly against a large class of alternative NPIV functions. Moreover, unlike bootstrap tests using bootstrapped critical values, our adaptive test uses simple Bonferroni corrected chi-squared critical values and hence is fast to compute.

We present two empirical applications. The first is adaptive testing for connected substitutes shape restrictions in demand for differential products using market level data. The second application is adaptive testing for monotonicity, convexity and parametric forms in Engel curves using household level data.

There are many papers on testing NPIV type models by extending Bierens [1990]'s

[1]This is needed even for adaptive minimax hypothesis testing in nonparametric regressions (without endogeneity); see Spokoiny [1996], Horowitz and Spokoiny [2001] and Guerre and Lavergne [2005].

test for conditional moment restrictions to models that allow for functions depending on endogenous regressors; see, e.g., Horowitz [2006], Santos [2012], Breunig [2015], Chen and Pouzo [2015], Chernozhukov et al. [2015], Zhu [2020] and the references therein. All of the published papers on testing NPIV models assume that some non-random sequences of key tuning (regularization) parameters satisfy some theoretical rate conditions, and are not adaptive to unknown smoothness of alternative functions. Our paper makes an important contribution by providing the first data-driven choice of key tuning parameters that are minimax rate-adaptive and powerful in testing equality and inequality restrictions in NPIV models.

Shape restrictions play a central role in economics and econometrics. See, for example, Chetverikov et al. [2018] for a recent review; Horowitz and Lee [2012], Blundell et al. [2017], Chetverikov and Wilhelm [2017] and Freyberger and Reeves [2019] for recent nonparametric estimation under shape constraints; and Chetverikov [2019], Chernozhukov et al. [2015] and Fang and Seo [2021] for testing shape restrictions. Our paper is the first to provide an adaptive and rate-optimal test (in L^2) for shape restrictions in NPIV models. We establish the minimax rate-adaptivity of our new test (in L^2) using an exponential inequality for U-statistics with increasing dimensions. Our paper also complements a recent work by Chen et al. [2021], which constructs honest and near-adaptive uniform confidence bands for a NPIV function and its partial derivatives using a bootstrapped Lepski's procedure (in sup-norm). Simulation studies and real data applications indicate that our adaptive test has asymptotic size control and is powerful in finite samples, eliminating the need for computationally intensive bootstrap critical values.

The remainder of the paper is organized as follows. Section 2 describes our new hypothesis test. Section 3 establishes the oracle minimax optimal rate of testing. Section 4 shows that this minimax optimal rate is attained (within a $\sqrt{\log\log(n)}$ term) by our testing procedure. Section 5 presents two simulation studies and Section 6 provides two empirical illustrations. Appendices A and B contain proofs for the results in Sections 3 and 4. The online supplementary appendices contain additional materials: Appendix C presents additional simulation results. Appendix D provides additional proofs for the results in Section 4. Appendix E contains additional technical lemmas and their proofs.

Basic notation. For a random variable X, we let $L^2(X)$ denote the equivalence class of all real-valued measurable functions ϕ of X with finite second moment $(\mathrm{E}[\phi^2(X)] < \infty)$, which is a Hilbert space under the norm $\|\phi\|_{L^2(X)} := \sqrt{\mathrm{E}[\phi^2(X)]}$ with associated inner product $\langle\cdot,\cdot\rangle_X$. We let $L^\infty = \{\phi : \|\phi\|_\infty < \infty\}$ where $\|\cdot\|_\infty$ denotes the supremum norm. For a matrix M, M' denotes its transpose, and M^- denotes its generalized inverse. For a $J \times J$ matrix $M = (M_{jl})_{1\leq j,l\leq J}$ we define its Frobenius norm as $\|M\|_F = \sqrt{\sum_{j,l=1}^J M_{jl}^2}$. Let $\|\cdot\|$ denote the Euclidean norm when applied to a vector and the operator norm induced by the Euclidean norm when applied to a matrix. For sequences of positive real numbers $\{a_n\}$ and $\{b_n\}$, we use the notation $a_n \lesssim b_n$ if $\limsup_{n\to\infty} a_n/b_n < \infty$, and $a_n \sim b_n$ if $a_n \lesssim b_n$ and $b_n \lesssim a_n$.

2. Preview of the Adaptive Hypothesis Testing

We first introduce the null and the alternative hypotheses as well as the concept of minimax rate of testing in Subsection 2.1. We then describe our new rate-adaptive test for NPIV type models in Subsection 2.2. The theoretical justifications are postponed to Sections 3 and 4.

2.1. Null Hypotheses and Nonparametric Alternatives

Let $\mathcal{H}$ denote a closed subset of $L^2(X)$ that captures some unknown degree of smoothness. Let $\{(Y_i, X_i, W_i)\}_{i=1}^n$ denote a random sample from the distribution P_h of (Y, X, W) satisfying the NPIV model (2.1):

$$Y = h(X) + U, \quad \text{where} \quad \mathrm{E}_h[U|W] = 0 \quad \text{and} \quad h \in \mathcal{H}. \tag{2.1}$$

Here, E_h denotes the (conditional) expectation under P_h. In this paper, we assume that the joint distribution of (X, W) does not depend on $h \in \mathcal{H}$ and that the conditional density of X given W is continuous on its support. The conditional expectation operator $T : L^2(X) \mapsto L^2(W)$ given by $Th(w) := \mathrm{E}[h(X)|W = w]$ is uniquely defined by the conditional density of X given W and hence does not depend on h. Then NPIV model (2.1) can be equivalently expressed as $\mathrm{E}_h[Y|W] = (Th)(W)$ for $h \in \mathcal{H}$. For easy presentation, we mainly consider a nonparametric class of functions as the maintained hypothesis $\mathcal{H}$. Nevertheless, our theoretical results allow for semiparametric structures $\mathcal{H}$ as well (see Subsection 4.2).

Let $\mathcal{H}_0$ denote the null class of functions in $\mathcal{H}$ that satisfies a conjectured restriction in (2.1). In this paper, we assume that $\mathcal{H}_0$ is a nonempty, closed and convex, strict subset of $\mathcal{H}$. For any $h \in \mathcal{H}$ there exists a unique element $\Pi_{\mathcal{H}_0} h \in \mathcal{H}_0$ such that $\inf_{\phi\in\mathcal{H}_0} \|h - \phi\|_{L^2(X)} = \|h - \Pi_{\mathcal{H}_0} h\|_{L^2(X)}$ (by the Hilbert projection theorem). In addition to a simple null $\mathcal{H}_0 = \{h_0\}$ (with a known function $h_0 \in \mathcal{H}$), we allow for general parametric, semi/nonparametric equality and inequality composite null restrictions. We present two examples of composite null restrictions below (see Subsection 4.2 for additional examples).

Example 2.1 (Nonparametric shape restrictions). *$\mathcal{H}_0$ can be a closed convex subset of $\mathcal{H}$ determined by inequality restrictions such that $\mathcal{H}_0 = \{h \in \mathcal{H} : \partial^l h \geq 0\}$, where $\partial^l h$ denotes the l-th partial derivative of h with respect to components of x. This allows for* *hypotheses on NPIV functions, including nonnegativity ($l = 0$), monotonicity ($l = 1$), or convexity ($l = 2$). We can also test for supermodularity restrictions on NPIV functions corresponding to $\mathcal{H}_0 = \{h \in \mathcal{H} : \partial^2 h/(\partial x_1 \partial x_2) \geq 0\}$. Our framework also allows for testing these restricted function classes simultaneously since intersections of these are again closed convex subsets of $\mathcal{H}$.*

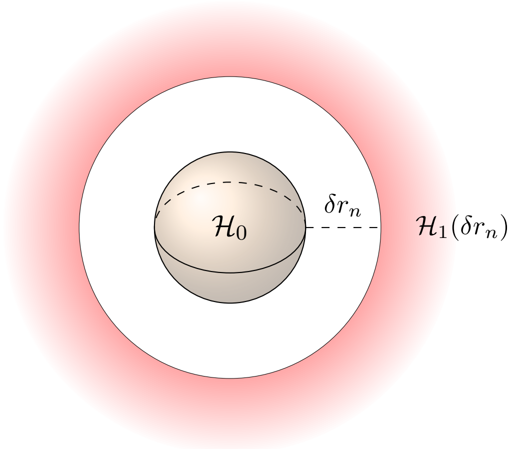
Example 2.2 (Semiparametric restrictions)**.** *Let $F(\cdot;\theta,g)$ be a known function up to unknown (θ,g), and consider the restricted class of functions $\mathcal{H}_0 = \{h \in \mathcal{H} : h(\cdot) = F(\cdot;\theta,g)$ for some $\theta \in \Theta$ and $g \in \mathcal{G}\}$, for a finite-dimensional, convex compact parameter space Θ and a nonparametric closed and convex function class $\mathcal{G}$. The known function $F(\cdot;\theta,g)$ could be nonlinear in θ but is assumed to be linear (or affine) in g and consequently, $\mathcal{H}_0$ is a closed convex subset of $\mathcal{H}$. Examples include null hypotheses of parametric form, or partially linear form, or partially parametric additive form.*

To analyze the power of any test of the null class $\mathcal{H}_0$ against nonparametric alternatives, we require some separation in $\|\cdot\|_{L^2(X)}-$distance between the null and the class of nonparametric alternatives for all $h \in \mathcal{H}$. Below, we use the notation $\|h - \mathcal{H}_0\|_{L^2(X)} := \inf_{\phi\in\mathcal{H}_0} \|h - \phi\|_{L^2(X)} = \|h - \Pi_{\mathcal{H}_0} h\|_{L^2(X)}$. We consider the following class of nonparametric alternatives

$$\mathcal{H}_1(\delta r_n) := \Big\{h \in \mathcal{H} : \|h - \mathcal{H}_0\|_{L^2(X)} \geq \delta r_n\Big\}$$

for some constant $\delta > 0$ and a *separation rate of testing* $r_n > 0$ that decreases to zero as the sample size n goes to infinity. We say that a test statistic $\mathtt{T}_n$ with values in $\{0,1\}$ is consistent uniformly over $\mathcal{H}_1(\delta r_n)$ if $\sup_{h\in\mathcal{H}_1(\delta r_n)} \mathrm{P}_h(\mathtt{T}_n = 0) = o(1)$.

In Section 3, we establish the *minimax (separation) rate of testing* r_n in the sense of Ingster [1993]: We propose a test that minimizes the sum of the supremum of the type I error over $\mathcal{H}_0$ and the supremum of the type II error over $\mathcal{H}_1(\delta r_n)$. Moreover, we show that the sum of both errors cannot be improved by any other test.



Definition 1. *A separation rate of testing r_n is called the minimax (separation) rate of testing if the following two requirements are met for every level $\alpha \in (0,1)$:*

(i) For some constant $\delta_* := \delta_*(\alpha) > 0$, it holds

$$\liminf_{n\to\infty} \inf_{\mathtt{T}_n} \Big\{ \sup_{h\in\mathcal{H}_0} \mathrm{P}_h(\mathtt{T}_n = 1) + \sup_{h\in\mathcal{H}_1(\delta_* r_n)} \mathrm{P}_h(\mathtt{T}_n = 0) \Big\} \geq \alpha, \tag{2.2}$$

where $\inf_{\mathtt{T}_n}$ is the infimum over all statistics with values in $\{0,1\}$. (ii) There exists a test statistic $\mathtt{T}_n := \mathtt{T}_n(\alpha)$ with values in $\{0,1\}$ such that

$$\limsup_{n\to\infty} \Big\{ \sup_{h\in\mathcal{H}_0} \mathrm{P}_h(\mathtt{T}_n = 1) + \sup_{h\in\mathcal{H}_1(\delta^* r_n)} \mathrm{P}_h(\mathtt{T}_n = 0) \Big\} \leq \alpha \tag{2.3}$$

for some constant $\delta^* > 0$.

We refer to Part (i) as the lower bound and Part (ii) as the upper bound, and the test statistic $\mathtt{T}_n := \mathtt{T}_n(\alpha)$ in Part (ii) attaining the matching lower and upper bound as an optimal test. We use r_n^* to denote the minimax (separation) rate of testing as the matching lower and upper bound.

In Section 3 we first establish a minimax rate of testing r_n^* assuming the knowledge of the smoothness of alternative NPIV functions $h \in \mathcal{H}$ and the inversion property of the unknown conditional expectation operator $T : L^2(X) \mapsto L^2(W)$. Both are unknown in empirical applications. The minimax rate r_n^* is attained by a sieve test statistic using an optimal choice of sieve dimension (a tuning parameter) that depends on these unknown objects, and hence is infeasible. In Section 4 we provide a data-driven modification of the optimal sieve test, i.e., a feasible testing procedure that adapts to the unknown smoothness of the unrestricted NPIV function $h \in \mathcal{H}$ in the presence of unknown smoothing properties of the inverse of the operator T. In particular, we propose a test statistic $\widehat{\mathtt{T}}_n$ with data-driven tuning parameters that nearly attains the minimax rate of testing, has asymptotic size control over the composite null, and is consistent uniformly over the class of nonparametric alternatives. Specifically, we show for some generic constant $\delta^\circ > 0$ that, for any nominal level $\alpha \in (0,1)$,

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_0} \mathrm{P}_h(\widehat{\mathtt{T}}_n = 1) \leq \alpha \qquad \text{and} \qquad \lim_{n\to\infty} \sup_{h\in\mathcal{H}_1(\delta^\circ \mathsf{r}_n)} \mathrm{P}_h(\widehat{\mathtt{T}}_n = 0) = 0, \tag{2.4}$$

where r_n coincides, up to the $\sqrt{\log\log(n)}$ multiplicative factor, with the minimax rate of testing r_n^*. We call such a feasible test $\widehat{\mathtt{T}}_n$ *adaptive* and *rate-optimal* (or sometimes simply adaptive).

2.2. Our Adaptive Test

Our test is based on a consistent estimate of the quadratic distance, $\|h - \Pi_{\mathcal{H}_0} h\|^2_{L^2(X)} = \|h - \mathcal{H}_0\|^2_{L^2(X)}$, between the NPIV function $h \in \mathcal{H}$ and its projection $\Pi_{\mathcal{H}_0} h$ onto $\mathcal{H}_0$ under the $\|\cdot\|_{L^2(X)}$. We first introduce some notation. Let $\{\psi_j\}_{j=1}^{\infty}$ and $\{b_k\}_{k=1}^{\infty}$ be complete basis functions for the Hilbert spaces $L^2(X)$ and $L^2(W)$ respectively. Let $\psi^J(\cdot)$ and $b^K(\cdot)$ be vectors of basis functions of dimensions J and $K = K(J) > J$ respectively. These can be cosine, power series, spline, or wavelet basis functions. Let $G = \mathrm{E}[\psi^J(X)\psi^J(X)']$, $G_b = \mathrm{E}[b^{K(J)}(W)b^{K(J)}(W)']$ and $S = \mathrm{E}[b^{K(J)}(W)\psi^J(X)']$. We assume that G, G_b and $S'G_b^{-1}S$ have full ranks. Then the $J \times K(J)$ matrix $A = G^{1/2}[S'G_b^{-1}S]^{-1}S'G_b^{-1}$ is well defined. Let Ψ_J denote the closed linear subspace of $L^2(X)$ spanned by $\{\psi_1, \ldots, \psi_J\}$. We define a population 2SLS projection of $h \in L^2(X)$ onto the sieve space Ψ_J as

$$Q_J h(\cdot) := \psi^J(\cdot)' G^{-1/2} A\, \mathrm{E}[b^K(W)h(X)] \ .$$

For any NPIV function $h \in \mathcal{H}$ in (2.1), we have $Q_J h(\cdot) = \psi^J(\cdot)'G^{-1/2}A\,\mathrm{E}_h[b^K(W)Y]$, and

$$\|Q_J(h - \Pi_{\mathcal{H}_0}h)\|_{L^2(X)}^2 = \left\| A\, \mathrm{E}_h\left[b^K(W)(Y - \Pi_{\mathcal{H}_0}h(X))\right]\right\|^2, \tag{2.5}$$

which approximates $\|h - \Pi_{\mathcal{H}_0}h\|_{L^2(X)}^2$ well as sieve dimension J grows large (see Lemma B.1).

For each sieve dimension J, we can construct a test based on an estimated quadratic distance $\|Q_J(h - \Pi_{\mathcal{H}_0}h)\|_{L^2(X)}^2$ between the unrestricted and restricted NPIV estimators of a function h satisfying (2.1). Let $\Psi = (\psi^J(X_1), \ldots, \psi^J(X_n))'$, $B = (b^K(W_1), \ldots, b^K(W_n))'$, $P_B = B(B'B)^- B'$, and $\widehat{A} = \sqrt{n}(\Psi'\Psi)^{1/2}[\Psi' P_B \Psi]^- \Psi' B(B'B)^-$. Let $\mathrm{Y} = (Y_1, \ldots, Y_n)'$. Our unrestricted sieve NPIV estimator solves a sample 2SLS problem (Blundell et al. [2007]):

$$\begin{aligned} \widehat{h}_J &= \arg\min_{\phi \in \Psi_J} \sum_{1 \le i, i' \le n} \big(Y_i - \phi(X_i)\big) b^{K(J)}(W_i)' \widehat{A}' \widehat{A} b^{K(J)}(W_{i'}) \big(Y_{i'} - \phi(X_{i'})\big) \\ &= \psi^J(\cdot)'[\Psi' P_B \Psi]^- \Psi' P_B \mathrm{Y} \ . \end{aligned} \tag{2.6}$$

Let $\mathcal{H}_{0,J}$ denote a nonempty, closed and convex, finite-dimensional subset of $\mathcal{H}_0$. A restricted NPIV estimator for $\Pi_{\mathcal{H}_0}h \in \mathcal{H}_0$ is given by

$$\widehat{h}_J^{\mathrm{R}} = \arg\min_{\phi \in \mathcal{H}_{0,J}} \sum_{1 \le i, i' \le n} \big(Y_i - \phi(X_i)\big) b^{K(J)}(W_i)' \widehat{A}' \widehat{A} b^{K(J)}(W_{i'}) \big(Y_{i'} - \phi(X_{i'})\big). \tag{2.7}$$

In this paper, the choice of $\mathcal{H}_{0,J}$ is allowed to depend on the structure of the null class of NPIV functions $\mathcal{H}_0$. For a general nonparametric or a semi-nonparametric composite null hypothesis, $\mathcal{H}_{0,J}$ depends on sieve dimension J and grows dense in $\mathcal{H}_0$ as the sample size increases. For instance, we let $\mathcal{H}_{0,J} = \Psi_J \cap \mathcal{H}_0$ under a nonparametric composite null whenever $\Psi_J \cap \mathcal{H}_0 \neq \emptyset$ (which holds for the nonparametric inequality restrictions in Example 2.1). We can also let $\mathcal{H}_{0,J} = \mathcal{H}_0$ under a simple null ($\mathcal{H}_0 = \{h_0\}$ for a known function h_0), or under a parametric composite null ($\mathcal{H}_0 = \{F(\cdot;\theta),\ \theta \in \Theta\}$ for some known mapping F).

For each sieve dimension J, we compute a $J-$dependent test statistic $n\widehat{D}_J/\widehat{\mathrm{v}}_J$, which is a standardized, centered (or leave-one-out) version of the sample analog of (2.5):

$$\widehat{D}_J = \frac{2}{n(n-1)} \sum_{1\leq i<i'\leq n} \big(Y_i - \widehat{h}^{\mathrm{R}}_J(X_i)\big) b^{K(J)}(W_i)'\widehat{A}'\widehat{A} b^{K(J)}(W_{i'})\big(Y_{i'} - \widehat{h}^{\mathrm{R}}_J(X_{i'})\big), \tag{2.8}$$

$$\widehat{\mathrm{v}}_J = \left\| \widehat{A}\left(\frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{h}_J(X_i))^2 b^{K(J)}(W_i) b^{K(J)}(W_i)'\right)\widehat{A}'\right\|_F , \tag{2.9}$$

where $\widehat{\mathrm{v}}_J$ estimates the population normalization factor

$$\mathrm{v}_J = \left\| A\, \mathrm{E}_h[(Y - h(X))^2 b^{K(J)}(W) b^{K(J)}(W)'] A' \right\|_F , \tag{2.10}$$

which is the variance of $\frac{2}{n(n-1)}\sum_{1\leq i<i'\leq n}\big(Y_i - h(X_i)\big)b^{K(J)}(W_i)'A'Ab^{K(J)}(W_{i'})\big(Y_{i'} - h(X_{i'})\big)$.

We compute our adaptive test for the null hypothesis $\mathcal{H}_0$ against nonparametric alternatives in three simple steps.

Step 1. Compute a *random exponential scan* (RES) index set:

$$\widehat{\mathcal{I}}_n := \left\{ J \leq \widehat{J}_{\max} : \quad J = \underline{J}2^j \text{ where } j = 0, 1, \ldots, j_{\max} \right\} \tag{2.11}$$

where $\underline{J} := \lfloor \sqrt{\log\log n} \rfloor$, $j_{\max} := \lceil \log_2(n^{1/3}/\underline{J}) \rceil$, and the empirical upper bound

$$\widehat{J}_{\max} := \min\left\{ J > \underline{J} : \quad 1.5\,[\zeta(J)]^2\sqrt{(\log J)/n} \geq \widehat{s}_J \right\}, \tag{2.12}$$

where $\widehat{s}_J$ is the minimal singular value of $(B'B)^{-1/2}B'\Psi(\Psi'\Psi)^{-1/2}$, and $\zeta(J) = \sqrt{J}$ for spline, wavelet, or trigonometric sieve basis, and $\zeta(J) = J$ for power series.

Step 2. Let $\#(\widehat{\mathcal{I}}_n)$ be the cardinality of the RES index set. For a nominal level $\alpha \in (0,1)$, we compute a Bonferroni corrected chi-squared critical value as

$$\widehat{\eta}_J(\alpha) := \big(q\big(\alpha/\#(\widehat{\mathcal{I}}_n), J\big) - J\big)/\sqrt{J},$$

where $q\big(a, J\big)$ is the $100(1-a)\%$-quantile of the standard chi-square distribution with J degrees of freedom.

Step 3. Compute $\widehat{\mathcal{W}}_J(\alpha) := \frac{n\,\widehat{D}_J}{\widehat{\eta}_J(\alpha)\,\widehat{\mathrm{v}}_J}$ for all $J \in \widehat{\mathcal{I}}_n$. Compute the test

$$\widehat{\mathrm{T}}_n := \mathbb{1}\left\{\text{there exists } J \in \widehat{\mathcal{I}}_n \text{ such that } \quad \widehat{\mathcal{W}}_J(\alpha) > 1\right\} \tag{2.13}$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function. Under the nominal level $\alpha \in (0,1)$, $\widehat{\mathtt{T}}_n = 1$ indicates rejection of the null hypothesis and $\widehat{\mathtt{T}}_n = 0$ indicates a failure to reject the null.

Remark 2.1. *The RES index set $\widehat{\mathcal{I}}_n$ in Step 1 determines a collection of candidate sieve dimensions J for our test. The data-dependent upper bound $\widehat{J}_{\max}$ ensures that the cardinality of the index set $\widehat{\mathcal{I}}_n$ is not too large relative to the sampling variability of unrestricted sieve NPIV estimation. We prove in Lemma B.8 that the empirical upper bound $\widehat{J}_{\max}$ diverges in probability at a rate much faster than that of $\underline{J}$ and thus, the search range is large enough to detect a large collection of alternative NPIV functions. In many simulations and real data applications, we find that our adaptive test results are not sensitive to the choice of the constant 1.5, and that the lower bound $\underline{J}$ is not binding in most cases.*

Remark 2.2 (Critical Values)**.** *A remarkable feature of our adaptive test is that it provides asymptotic size control for inequality restrictions without restricting the degree of freedom of the Bonferroni corrected chi-squared critical values to the number of binding constraints. This is established by the observation that our Bonferroni corrected critical values $\widehat{\eta}_J(\alpha)$ diverge slowly as $n \to \infty$ with probability approaching one; see Lemma B.5. This, along with the cardinality of $\widehat{\mathcal{I}}_n$ not becoming too large by construction, and complexity restrictions on the composite null hypotheses, enables us to establish asymptotic size control.*

Remark 2.3 (Choice of K)**.** *We let $K = K(J) = cJ$ for some finite constant $c > 1$, and our adaptive testing procedure optimizes over J given the choice of $K(J)$. We have tried $K(J) = 2J$ and $K(J) = 4J$ in simulation studies. The simulation results, in terms of size and power, are not sensitive to these choices of K. This is consistent with our theory that the choice of J is the key tuning parameter in minimax rate-optimal hypothesis testing in NPIV models using sieve methods.*

3. The Minimax Rate of Testing

This section derives the minimax rate of testing in NPIV models, when $\mathcal{H}$ coincides with the Sobolev ellipsoid of *a priori* known smoothness $p > 0$. Subsection 3.1 establishes the lower bound for the rate of testing in L^2. Subsection 3.2 shows that the lower bound can be achieved by a simple test statistic if the tuning parameter can be chosen optimally.

3.1. The Lower Bound

Before we state the lower bound for the rate of testing, we introduce additional notation and main assumptions. Based on the basis functions $\{\psi_j\}_{j=1}^{\infty}$, we denote its $L^2(X)$–normalized basis functions by $\{\widetilde{\psi}_j\}_{j=1}^{\infty}$. We assume that $\mathcal{H}$ coincides with the Sobolev ellipsoid $\mathcal{H} = \{h \in L^2(X) : \sum_{j\geq 1} j^{2p/d_x} \langle h, \widetilde{\psi}_j\rangle_X^2 \leq C_{\mathcal{H}}^2\}$ for some constant $C_{\mathcal{H}} > 0$.

Assumption 1. *(i)* $\inf_{w\in\mathcal{W}}\inf_{h\in\mathcal{H}} \mathbb{V}ar_h(Y-h(X)|W=w)\geq \underline{\sigma}^2>0$*; (ii) for any* $h\in\mathcal{H}$*,* $Th=0$ *implies that* $\|h\|^2_{L^2(X)}=0$*; (iii) the densities of* X *and* W *are uniformly bounded below from zero and from above on their rectangular support; (iv) there are a finite constant* $C>0$ *and a positive decreasing function* ν *with* $\nu_j:=\nu(j)$ *such that* $\|Th\|^2_{L^2(W)}\leq C\sum_{j\geq 1}\nu_j^2\langle h,\widetilde{\psi}_j\rangle_X^2$ *for all* $h\in\mathcal{H}$*.*

Assumptions 1(i)(ii)(iii) are basic regularity conditions imposed in the paper. Assumption 1(iv) specifies the smoothing property of the conditional expectation operator T relative to the basis $\{\widetilde{\psi}_j\}$. The smoother T is (i.e., the smoother the conditional density of X given W is), Assumption 1(iv) will be satisfied with a faster decreasing to zero sequence $\{\nu_j\}$, and the harder it is to detect properties of the NPIV function in $L^2(X)$ metric.

In this paper we call a decreasing sequence $\{\nu_j\}$ *regularly varying* if $\nu_J^{-4}J\lesssim\sum_{j=1}^J\nu_j^{-4}$. The regularly varying sequence $\{\nu_j\}$ allows for very broad decreasing patterns, and includes two leading special cases: (1) *mildly ill-posed* case where $\nu_j=j^{-a/d_x}$ for some $a>0$; and (2) *severely ill-posed* case where $\nu_j=\exp(-j^{a/d_x}/2)$ for some $a>0$.

Theorem 3.1. *Let Assumption 1 be satisfied. Consider testing a null hypothesis* $\mathcal{H}_0$ *versus* $\mathcal{H}_1(\delta r_n)=\big\{h\in\mathcal{H}:\|h-\mathcal{H}_0\|_{L^2(X)}\geq\delta r_n\big\}$ *for some constant* $\delta>0$ *and a separation rate*

$$r_n=n^{-\frac{1}{2}}\left(\sum_{j=1}^{J_*}\nu_j^{-4}\right)^{1/4},\quad \textit{with } J_*:=\max\left\{J:n^{-\frac{1}{2}}\left(\sum_{j=1}^{J}\nu_j^{-4}j^{\frac{4p}{d_x}}\right)^{1/4}\leq C_{\mathcal{H}}\right\}. \tag{3.1}$$

Then: for any $\alpha\in(0,1)$ *there exists a constant* $\delta_*:=\delta_*(\alpha)>0$ *such that*

$$\liminf_{n\to\infty}\inf_{\mathtt{T}_n}\left\{\sup_{h\in\mathcal{H}_0}\mathrm{P}_h(\mathtt{T}_n=1)+\sup_{h\in\mathcal{H}_1(\delta_* r_n)}\mathrm{P}_h(\mathtt{T}_n=0)\right\}\geq\alpha,$$

where $\sup_{h\in\mathcal{H}_\ell}\mathrm{P}_h(\cdot)$ *denotes the supremum over* $h\in\mathcal{H}_\ell$ *and distributions of* (X,W,U) *satisfying Assumption 1 for* $\ell=0,1$*.*
Further, when $\{\nu_j\}$ *is regularly varying, the separation rate* r_n *given in* (3.1) *simplifies to*

$$r_n\sim J_*^{-p/d_x},\quad \textit{with } J_*\sim\max\left\{J:n^{-1/2}J^{1/4}\nu_J^{-1}\leq J^{-p/d_x}\right\}. \tag{3.2}$$

(1) Mildly ill-posed case: $J_*\sim n^{2d_x/(4(p+a)+d_x)}$ *and* $r_n\sim n^{-2p/(4(p+a)+d_x)}$*.*

(2) Severely ill-posed case: $J_*=(c\log n)^{d_x/a}$ *for some* $c\in(0,1)$ *and* $r_n\sim(\log n)^{-p/a}$*.*

Remark 3.1. *In the literature on linear ill-posed inverse problem with a compact operator* T*, an "exact link condition" is commonly used to describe the smoothing (or compact embedding) property of* T*, which can be stated as follows:*

$$c\sum_{j\geq 1}\nu_j^2\langle h,\widetilde{\psi}_j\rangle_X^2\leq\|Th\|^2_{L^2(W)}\leq C\sum_{j\geq 1}\nu_j^2\langle h,\widetilde{\psi}_j\rangle_X^2\qquad \textit{for all } h\in\mathcal{H} \tag{3.3}$$

for some finite constants $C \geq c > 0$ and a positive decreasing function ν with $\nu_j := \nu(j)$. The RHS inequality of (3.3) *(i.e., Assumption 1(iv)) is used for the lower bound calculation, and the LHS inequality of* (3.3) *is imposed for the upper bound calculation. However, to have matching lower and upper bound, i.e., to establish the rate is minimax optimal, the exact link condition* (3.3) *or something similar is typically imposed even with a known T; see, e.g., Chen and Reiß [2011]. We note that any compact operator T has a unique singular value decomposition. If the basis $\{\widetilde{\psi}_j\}$ is an eigenfunction basis associated with the operator T, then* (3.3) *is automatically satisfied with $C = c = 1$ and $\{\nu_j\}_{j=1}^{\infty}$ being its singular values in decreasing order. More generally,* (3.3) *is also satisfied when $\{\widetilde{\psi}_j\}$ is a Riesz basis (see Blundell et al. [2007]). Since the conditional expectation operator T is compact under very mild condition (such as when the conditional density of X given W is continuous), it typically satisfies* (3.3)*, which is an alternative way to express the smoothing property of the operator T.*

In our proof of Theorem 3.1, we reduce the lower bound calculation for the NPIV model to that for a model with a known operator T. Consequently, Assumption 1(iv) is sufficient to establish the lower bound. However, for the upper bound calculation of the NPIV model, we need to estimate the unknown operator T. Therefore, in addition to the LHS inequality of (3.3), some extra sufficient conditions will be used to address the error of estimating T nonparametrically. See the next subsection for details.

3.2. An Upper Bound Under Simple Null Hypotheses

For the simple null hypothesis $\mathcal{H}_0 = \{h_0\}$, we consider the test statistic

$$\widehat{D}_J(h_0) = \frac{2}{n(n-1)} \sum_{1 \leq i < i' \leq n} \big(Y_i - h_0(X_i)\big)\big(Y_{i'} - h_0(X_{i'})\big) b^{K(J)}(W_i)' \widehat{A}' \widehat{A} b^{K(J)}(W_{i'}), \quad (3.4)$$

which is $\widehat{D}_J$ defined in (2.8) with $\widehat{h}_J^{\mathrm{R}} = h_0$. A J-dependent analog of the test $\widehat{\mathtt{T}}_n$ is given by

$$\mathtt{T}_{n,J} = \mathbb{1}\left\{\frac{n\widehat{D}_J(h_0)}{\widehat{\mathrm{V}}_J} > \eta_J(\alpha)\right\} \quad \text{with} \quad \eta_J(\alpha) = (q(\alpha, J) - J)/\sqrt{J}. \quad (3.5)$$

The test $\mathtt{T}_{n,J}$ with optimally chosen J serves as a benchmark of our adaptive testing procedure (given in (4.1)) for the simple null hypothesis.

We define the projections $\Pi_J h(\cdot) = \psi^J(\cdot)' G^{-1} \langle \psi^J, h\rangle_{L^2(X)}$ for $h \in L^2(X)$ and $\Pi_K m(\cdot) = b^K(\cdot)' G_b^{-1} \operatorname{E}[b^K(W) m(W)]$ for $m \in L^2(W)$. Further, let $s_J = \inf_{h \in \Psi_J} \|\Pi_K T h\|_{L^2(W)} / \|h\|_{L^2(X)}$, i.e., s_J coincides with the minimal singular value of $G_b^{-1/2} S G^{-1/2}$. Let $\zeta_J = \max(\zeta_{\psi,J}, \zeta_{b,K})$, $\zeta_{\psi,J} = \sup_x \|G^{-1/2}\psi^J(x)\|$ and $\zeta_{b,K} = \sup_w \|G_b^{-1/2} b^K(w)\|$. We assume throughout the paper that $\zeta_J = O(\sqrt{J})$ (which holds for polynomial spline, wavelet, and cosine bases), or

$\zeta_J = O(J)$ (which holds for orthogonal polynomial bases).

Assumption 2. *(i)* $\sup_{w\in\mathcal{W}}\sup_{h\in\mathcal{H}} \mathrm{E}_h[(Y - \widetilde{h}(X))^2|W = w] \le \overline{\sigma}^2 < \infty$ *where* $\widetilde{h} \in \{h, \Pi_{\mathcal{H}_0}h\}$ *and* $\sup_{h\in\mathcal{H}} \mathrm{E}_h[(Y-h(X))^4] < \infty$*; (ii)* $s_J^{-1}\zeta_J^2\sqrt{(\log J)/n} = O(1)$*; (iii)* $\zeta_J\sqrt{\log J} = O(J^{p/d_x})$*; (iv)* $s_J^{-1}\|\Pi_K T(\Pi_J h - h)\|_{L^2(W)} \le C_T\|\Pi_J h - h\|_{L^2(X)}$ *for a constant* $C_T > 0$*, uniformly for* $h \in \mathcal{H}$*.*

Below we denote $\Psi_{J,1} := \{h \in \Psi_J : \|h\|_{L^2(X)} = 1\}$. Then $\tau_J := \left[\inf_{h\in\Psi_{J,1}} \|Th\|_{L^2(W)}\right]^{-1}$ is the sieve measure of ill-posedness that has been used in sieve estimation of NPIV models (see, e.g., Blundell et al. [2007]).

Assumption 3. *(i)* $\sup_{h\in\Psi_{J,1}} \tau_J\|(\Pi_K T - T)h\|_{L^2(W)} = o(1)$*; (ii) the LHS inequality of* (3.3) *holds.*

Assumption 2(i) is an extra condition on the data-generating process (DGP) since it imposes upper bounds on conditional second moment and finiteness of uncondition 4th moment. This additional DGP is already allowed for in the proof of Theorem 3.1 (on the lower bound). The other conditions, Assumption 2(ii)(iii)(iv), are imposed since our test statistic involves linear sieve estimated operator T to achieve the separation rate. Precisely, Assumption 2(ii)(iii) introduces conditions on the sieve dimension J, satisfying the growth condition imposed in Theorem 3.1. Assumption 2(iv) imposes an upper bound on the smoothing properties of the conditional expectation operator T. It is the usual L^2 stability condition imposed in the sieve NPIV literature, and is satisfied by Riesz bases (see Blundell et al. [2007, Assumption 6] and Chen and Pouzo [2012, Assumption 5.2(ii)]).

Assumption 3(i) is a mild condition on the approximation properties of the basis used for the instrument space, see Chen and Christensen [2018, Assumption 4(i)]. It implies that s_J and τ_J^{-1} are equivalent:

$$\tau_J^{-1} \ge s_J = \inf_{h\in\Psi_{J,1}} \|\Pi_K Th\|_{L^2(W)} \ge \inf_{h\in\Psi_{J,1}} \|Th\|_{L^2(W)} - \sup_{h\in\Psi_{J,1}} \|(\Pi_K T - T)h\|_{L^2(W)} = \tau_J^{-1}(1-o(1)).$$

While Assumption 3(ii) implies $\tau_J^{-1} = \inf_{h\in\Psi_{J,1}} \|Th\|_{L^2(W)} \ge \sqrt{c}\nu_J$ for all J. Assumption 3 thus implies

$$s_J^{-1} \sim \tau_J \le (\sqrt{c})^{-1}\nu_J^{-1}.$$

We note that $s_J \le \tau_J^{-1} \le \|T\widetilde{\psi}_J\|_{L^2(W)} \le \sqrt{C}\nu_J$ under Assumption 1(iv) (i.e., the RHS inequality of (3.3)) and $\{\widetilde{\psi}_j\}_j$ being an orthonormal basis in $L^2(X)$. We also note that Assumption 2(iv) is satisfied under Assumption 1(iv) and Assumption 3. In summary, Assumption 2 and Assumption 3 have no effect on the lower bound calculation in Theorem 3.1.

The next theorem provides an upper bound on the separation rate of testing in L^2 under a simple null using the test statistic $\mathtt{T}_{n,J}$.

Theorem 3.2. *Let Assumptions 1(i)-(iii) and 2 be satisfied. Consider testing the simple hypothesis $\mathcal{H}_0 = \{h_0\}$ (for a known function h_0) versus $\mathcal{H}_1(\delta^\circ r_{n,J}) = \{h \in \mathcal{H} : \|h - h_0\|_{L^2(X)} \geq \delta^\circ r_{n,J}\}$ for a constant $\delta^\circ > 0$ and a separation rate*

$$r_{n,J} = \max\left\{n^{-1/2} s_J^{-1} J^{1/4}, J^{-p/d_x}\right\}. \tag{3.6}$$

Then, for any $\alpha \in (0,1)$ we have

$$\limsup_{n\to\infty} \mathrm{P}_{h_0}(\mathtt{T}_{n,J} = 1) \leq \alpha \qquad \text{and} \qquad \lim_{n\to\infty} \sup_{h\in\mathcal{H}_1(\delta^\circ r_{n,J})} \mathrm{P}_h(\mathtt{T}_{n,J} = 0) = 0. \tag{3.7}$$

In addition, let Assumption 3 hold and $J_{*0} := \max\left\{J : n^{-1/2}\nu_J^{-1} J^{1/4} \leq J^{-p/d_x}\right\}$. Then: the test statistic $\mathtt{T}_{n,J_{*0}}$ attains the optimal separation rate of

$$r_{n,J_{*0}} = (J_{*0})^{-p/d_x} \sim r_n, \tag{3.8}$$

which is the lower bound rate given in (3.2) when $\{\nu_j\}$ is regularly varying.

(1) Mildly ill-posed case: $J_{*0} \sim n^{2d_x/(4(p+a)+d_x)}$ and $r_{n,J_{*0}} \sim n^{-2p/(4(p+a)+d_x)}$.

(2) Severely ill-posed case: $J_{*0} = \big(c \log n\big)^{d_x/a}$ for some $c \in (0,1)$ and $r_{n,J_{*0}} \sim (\log n)^{-p/a}$.

Theorem 3.2 shows that, under Assumptions 1(i)-(iii) and 2, the test statistic $\mathtt{T}_{n,J}$ given in (3.5) attains the L^2-separation rate of testing $r_{n,J}$ in (3.6). Given a sieve dimension J, this rate consists of a standard deviation term $(n^{-1/2}s_J^{-1}J^{1/4})$ and a bias term (J^{-p/d_x}). A central step to achieve this rate result is to establish a rate of convergence of the quadratic distance estimator $\widehat{D}_J(h_0)$ (see Theorem B.1), which we show is sufficient for the consistency of $\mathtt{T}_{n,J}$ uniformly over $\mathcal{H}_1(\delta^\circ r_{n,J})$. In addition, under Assumption 3, Theorem 3.2 implies that the sieve test $\mathtt{T}_{n,J_{*0}}$ achieves the L^2-*minimax rate of testing* under a simple null, with known smoothness p of the nonparametric alternatives and known degree of ill-posedness.

Given a sieve dimension J, the L^2-rate of sieve estimation for any NPIV function $h \in \mathcal{H}$ is: $\max\{n^{-1/2}s_J^{-1}J^{1/2}, J^{-p/d_x}\}$ (see, e.g., Chen and Reiß [2011]). Comparing the L^2-rate of estimation and of testing via the sieve NPIV procedures, while both have the same bias term J^{-p/d_x}, the L^2 rate of testing has a smaller "standard deviation" term $n^{-1/2}s_J^{-1}J^{1/4}$. Intuitively, we may obtain a higher precision in testing as the L^2-rate of testing is determined by estimating a quadratic norm of the unrestricted NPIV function $h \in \mathcal{H}$. Interestingly, although this leads to a faster optimal L^2-rate of sieve testing $r_{n,J_{*0}} \sim n^{-2p/(4(p+a)+d_x)}$ for mildly illposed case, the optimal L^2-rate of sieve testing $r_{n,J_{*0}} \sim (\log n)^{-p/a}$ is the same as the optimal L^2-rate of sieve estimation for severely illposed case. This is because, in the severely ill-posed case, the bias term dominates the standard deviation term for the optimal sieve dimension J in both sieve testing and estimation.

Remark 3.2 (L^2-minimax rate of testing vs L^2-minimax rate of estimation)**.** *Theorems 3.1 and 3.2 together imply that the L^2-minimax rate of testing $r_n^* = (\log n)^{-p/a}$ in the severely ill-posed case coincides with the L^2-minimax rate of estimation (Chen and Reiß [2011]). For the mildly ill-posed NPIV models, the L^2-minimax rate of testing $r_n^* = n^{-2p/(4(p+a)+d_x)}$ goes to zero faster than the L^2-minimax rate of estimation, which is $n^{-p/(2(p+a)+d_x)}$ (Hall and Horowitz [2005] and Chen and Reiß [2011]).*

4. Adaptive Inference

This section establishes theoretical properties of our test statistic $\widehat{\mathtt{T}}_n$ that is constructed using data-driven choices of tuning parameters. We show that our test is able to adapt to the unknown smoothness $p > 0$ of the functions in $\mathcal{H}$. Subsection 4.1 establishes the rate optimality of our adaptive test for simple null hypotheses. Subsection 4.2 extends this result to testing for composite null problems. Subsection 4.3 proposes L^2-confidence sets by inverting the adaptive test under imposed restrictions on the NPIV function.

4.1. Adaptive Testing Under Simple Null Hypotheses

We establish an upper bound for the rate of testing using our test statistic $\widehat{\mathtt{T}}_n$ for a simple null. Under the simple null hypothesis $\mathcal{H}_0 = \{h_0\}$, for some known function h_0 satisfying (1.1), our test $\widehat{\mathtt{T}}_n$ given in (2.13) simplifies to

$$\widehat{\mathtt{T}}_n = \mathbb{1}\left\{\text{there exists } J \in \widehat{\mathcal{I}}_n \text{ such that } \frac{n\widehat{D}_J(h_0)}{\widehat{\mathrm{v}}_J} > \widehat{\eta}_J(\alpha)\right\}, \tag{4.1}$$

where $\widehat{D}_J(h_0)$ is defined in (3.4), and $\widehat{\mathcal{I}}_n$, $\widehat{\mathrm{v}}_J$, $\widehat{\eta}_J(\alpha)$ are given in Subsection 2.2.

Recall that the RES index set $\widehat{\mathcal{I}}_n$, given in (2.11), depends on an upper bound $\widehat{J}_{\max}$ given in (2.12). To establish our asymptotic results below, we introduce a non-random index set $\mathcal{I}_n$ with a deterministic upper bound $\overline{J}$ as follows:

$$\mathcal{I}_n = \left\{J \leq \overline{J} : \quad J = \underline{J}2^j \text{ where } j = 0, 1, \ldots, j_{\max}\right\} \subset [\underline{J}, \overline{J}], \tag{4.2}$$

with $\overline{J} = \sup\{J : \zeta_J^2\sqrt{(\log J)/n} \leq \overline{c}\, s_J\}$ for some sufficiently large constant $\overline{c} > 0$. We show in Lemma B.8(i) that $\widehat{J}_{\max} \leq \overline{J}$ (and thus $\widehat{\mathcal{I}}_n \subset \mathcal{I}_n$) holds with probability approaching one uniformly over all functions $h \in \mathcal{H}$. Thus $\overline{J}$ serves as a deterministic upper bound for the RES index set $\widehat{\mathcal{I}}_n$.

Assumption 4. *(i) Assumptions 2(ii)(iv) hold uniformly for all $J \in \mathcal{I}_n$; (ii) For all $J = J(n)$ and $L = L(n)$ with $L = o(J)$ and $L \to \infty$ it holds that* $\max\left(\mathrm{v}_L, s_L^{-2}\sqrt{\log\log L}\right) =$

$o(\mathrm{v}_J)$ uniformly for $h \in \mathcal{H}_0$; (iii) $p \geq 3d_x/4$ when using cosine, spline, or wavelet basis functions and $p \geq 7d_x/4$ when using power series basis functions.

Assumption 4(i)(iii) strengthen Assumption 2(ii)(iii)(iv) to hold uniformly over the deterministic index set $\mathcal{I}_n$. The uniformity over $\mathcal{I}_n$ condition is used to establish the asymptotic size control only, and is not needed for consistency uniformly over the class of nonparametric alternative functions. Assumption 4(i) restricts the growth of the deterministic upper bound $\overline{J}$ of the RES index set $\widehat{\mathcal{I}}_n$. Assumption 4(iii) imposes a lower bound on the smoothness of the function class $\mathcal{H}$. From the proof of Theorem 4.1 we see that Assumption 4(ii) is implied by $s_J^{-4} J \lesssim \sum_{j=1}^{J} s_j^{-4}$, which in turn is implied by Assumptions 1(iv) and 3 with $\{\nu_j\}$ regularly varying.

Let an integer J° be the largest sieve dimension parameter such that the squared bias dominates the variance within a $\sqrt{\log\log n}$ term, that is,

$$J^\circ := \max\left\{J : \, n^{-1}\sqrt{\log\log n}\, s_J^{-2}\sqrt{J} \leq J^{-2p/d_x}\right\}. \tag{4.3}$$

Under Assumptions 4(i)(iii), Lemma B.8(ii) establishes that the "optimal" adaptive sieve dimension $J^\circ \in \widehat{\mathcal{I}}_n$ with probability approaching one.

Theorem 4.1. *Let Assumptions 1(i)-(iii), 2(i) and 4 be satisfied. Consider testing the simple null $\mathcal{H}_0 = \{h_0\}$ (for a known function h_0) versus $\mathcal{H}_1(\delta^\circ \mathsf{r}_n) = \{h \in \mathcal{H} : \|h - h_0\|_{L^2(X)} \geq \delta^\circ \mathsf{r}_n\}$ for a constant $\delta^\circ > 0$ and an adaptive separation rate*

$$\mathsf{r}_n = (J^\circ)^{-p/d_x}. \tag{4.4}$$

Then, for any $\alpha \in (0,1)$ we have

$$\limsup_{n\to\infty} \mathrm{P}_{h_0}(\widehat{\mathtt{T}}_n = 1) \leq \alpha \qquad \text{and} \qquad \lim_{n\to\infty} \sup_{h\in\mathcal{H}_1(\delta^\circ \mathsf{r}_n)} \mathrm{P}_h(\widehat{\mathtt{T}}_n = 0) = 0. \tag{4.5}$$

In addition, let Assumptions 1(iv) and 3 hold with $\{\nu_j\}$ regularly varying. Then: $s_J \sim \nu_J$, Assumption 4(ii) holds, and $J^\circ \sim \max\left\{J : \, n^{-1}\sqrt{\log\log n}\,\nu_J^{-2}\sqrt{J} \leq J^{-2p/d_x}\right\}$.

(1) Mildly ill-posed case: $\mathsf{r}_n \sim \left(\sqrt{\log\log n}/n\right)^{2p/(4(p+a)+d_x)}$.

(2) Severely ill-posed case: $\mathsf{r}_n \sim (\log n)^{-p/a}$.

Theorem 4.1 establishes an upper bound for the testing rate of the adaptive test $\widehat{\mathtt{T}}_n$ under a simple null hypothesis. The proof of Theorem 4.1 relies on a novel exponential bound for degenerate U-statistics based on sieve estimators (see Lemma B.6). In particular, we control the type I error using tight lower bounds for adjusted chi-square critical values (see Lemma B.5) and show consistency of $\widehat{\mathtt{T}}_n$ uniformly over $\mathcal{H}_1(\delta^\circ \mathsf{r}_n)$.

From Theorem 4.1 we see that the adaptive test attains the oracle minimax rate of testing within a $\sqrt{\log\log(n)}-$term in the mildly ill-posed case. For the adaptive testing in regression models without endogeneity (i.e., when $X = W$), it is well known that the extra $\sqrt{\log\log(n)}-$term is required (see Spokoiny [1996]). In the severely ill-posed cases, our adaptive test attains the exact minimax rate of testing and hence, there is no price to pay for adaptation. This is because, in the severely ill-posed case, the bias term dominates the standard deviation term when the sieve dimension coincides with J°, irrespective of the $\sqrt{\log\log(n)}$ term.

4.2. Adaptive Testing Under Composite Null Hypotheses

We extend the results from Subsection 4.1 to adaptive testing for a general composite null hypothesis $\mathcal{H}_0$, which is a nonempty, closed and convex strict subset of $\mathcal{H} = B^p_{2,2}$. Without loss of generality we assume $0 \in \mathcal{H}_0$. This is satisfied for the inequality restrictions in Example 2.1 and the semiparametric equality restrictions considered in Example 2.2 if, for instance, $F(\cdot;\theta,g) = 0$ for some $\theta \in \Theta$ and $g \in \mathcal{G}$.

To do so, we need to impose conditions on the complexity of the class of restricted functions $\mathcal{H}_0$. Specifically, given the $(K-1)$-dimensional unit sphere $\mathcal{S}^K = \{\mathsf{e} \in \mathbb{R}^K : \mathsf{e}_1^2 + \ldots + \mathsf{e}_K^2 = 1\}$, we control the size of the function class

$$\mathcal{F}_{h,\mathsf{e}} = \left\{(\phi - \Pi_{\mathcal{H}_0} h)(X)\widetilde{b}^{K^\circ}(W)'\mathsf{e} : \phi \in \mathcal{H}_{0,J^\circ}\right\}, \quad \mathsf{e} \in \mathcal{S}^{K^\circ},$$

where $K^\circ = K(J^\circ)$ and $\widetilde{b}^K(\cdot) = G_b^{-1/2} b^K(\cdot)$. Its envelope function is denoted by $F_{h,\mathsf{e}}$. We denote $Z := (X', W')'$. Below, we impose conditions on the covering number $N_{[]}(\epsilon, \mathcal{F}, L^2(Z))$, which is the minimal number of ϵ-brackets, in $L^2(Z)$ sense, needed to cover a function class $\mathcal{F}$. We denote $\mathcal{C}_h = \max_{\mathsf{e}\in\mathcal{S}^{K^\circ}} \int_0^1 \left(1 + \log N_{[]}\left(\epsilon\|F_{h,\mathsf{e}}\|_{L^2(Z)}, \mathcal{F}_{h,\mathsf{e}}, L^2(Z)\right)\right)^{1/2} d\epsilon$.

Assumption 5. *(i) For any* $\varepsilon > 0$ *it holds* $\sup_{h\in\mathcal{H}_0} \mathrm{P}_h\big(\max_{J\in\mathcal{I}_n}(\zeta_J\|\widehat{h}^{\mathrm{R}}_J - h\|_{L^2(X)}/c_J) > \varepsilon\big) \to 0$ *with* $c_J = \max\{1, (\log\log J)^{1/4}\}$*; (ii) for some constant* $C > 0$ *it holds that* $\sup_{h\in\mathcal{H}_1(\delta^\circ \mathrm{r}_n)} \mathrm{P}_h\big(\zeta_{J^\circ}\mathcal{C}_h\|\widehat{h}^{\mathrm{R}}_{J^\circ} - \Pi_{\mathcal{H}_0} h\|_{L^2(X)} > C\big) \to 0$ *and* $\sup_{h\in\mathcal{H}_1(\delta^\circ \mathrm{r}_n)} \mathcal{C}_h \lesssim (J^\circ)^{1/4}$.

Assumption 5 restricts the complexity of the composite null hypothesis $\mathcal{H}_0$. Assumption 5(i) implies that $\widehat{\mathtt{T}}_n$ has size control uniformly over the composite null $\mathcal{H}_0$. Assumption 5(ii) ensures the consistency of $\widehat{\mathtt{T}}_n$ uniformly over $\mathcal{H}_1(\delta^\circ \mathrm{r}_n)$. Note that Assumption 5 imposes estimation rate conditions on $\widehat{h}^{\mathrm{R}}_J$ under the composite null and the nonparametric alternatives, which can be viewed as NPIV extensions of the parametric estimation rate conditions imposed in Horowitz and Spokoiny [2001, Assumption 2] for testing for a parametric regression against nonparametric regressions.

Remark 4.1 (Primitive Conditions for Assumption 5(i))**.** *Assumption 5(i) is a very mild condition on the estimation rate (in* L^2*) of the restricted sieve NPIV estimator under the*

null hypothesis.
(1) In the case of parametric restrictions, where $\|\widehat{h}_J^{\mathrm{R}} - h\|_{L^2(X)} \leq const. \times n^{-1/2}$ *with probability approaching one uniformly over* $h \in \mathcal{H}_0$*, Assumption 5(i) is automatically satisfied by Assumption 4(i).*
(2) Under nonparametric restrictions, we note that $\|\widehat{h}_J^{\mathrm{R}} - h\|_{L^2(X)} \leq \|\widehat{h}_J - h\|_{L^2(X)}$ *for all* $h \in \mathcal{H}_0$*, and that*

$$\max_{J \in \mathcal{I}_n} \frac{\zeta_J \|\widehat{h}_J - h\|_{L^2(X)}}{c_J} \leq const. \times \max_{J \in \mathcal{I}_n} \left\{ \frac{\zeta_J \sqrt{J}}{\sqrt{n} s_J c_J} + \frac{\zeta_J \|\Pi_J^{\mathcal{I}_n} h - h\|_{L^2(X)}}{c_J} \right\} \tag{4.6}$$

with probability approaching one uniformly for $h \in \mathcal{H}_0$*, where* $\Pi_J^{\mathcal{I}_n}$ *denotes the projection onto the closed linear subspace of* $L^2(X)$ *spanned by* $\{\psi_J : J \in \mathcal{I}_n\}$*. The first summand on the right hand side of (4.6) converges to zero by the definition of* $\overline{J} = \overline{J}(n)$*. For the bias part, we assume that the index set has sufficient information to approximate the NPIV function* h*. Let* p_0 *denote the smoothness and* d_0 *the dimension of the nonparametric component under* $\mathcal{H}_0$*. If* $\|\Pi_J^{\mathcal{I}_n} h - h\|_{L^2(X)} = O(J^{-\mathrm{p}_0/d_0})$ *and* $\zeta_J = O(\sqrt{J})$*, the second summand of the right hand side of (4.6) uniformly converges to zero if* $\mathrm{p}_0/d_0 \geq 1/2$*. Since the class of restricted functions* $\mathcal{H}_0$ *is a less complex subset of* $\mathcal{H}$*, it is reasonable to assume that* $\mathrm{p}_0/d_0 \geq p/d_x$ *and thus* $\mathrm{p}_0/d_0 \geq 1/2$ *is automatically satisfied given Assumption 4(iii).*

Remark 4.2 (Primitive Conditions for Assumption 5(ii))**.** *Assumption 5(ii) restricts the complexity of the null hypothesis to have no effect on the adaptive minimax rate of testing asymptotically. Note that for any* $\epsilon > 0$ *and* $\mathsf{e} \in \mathcal{S}^{K^\circ}$ *we have*

$$\mathrm{E}\Big[\sup_{\phi_1, \phi_2 \in \mathcal{H}_{0,J^\circ} : \|\phi_1 - \phi_2\|_\infty \leq \epsilon} \big| (\phi_1 - \phi_2)(X) \widetilde{b}^{K^\circ}(W)' \mathsf{e} \big|^2 \Big] \leq \epsilon^2,$$

using that $\mathrm{E}(\widetilde{b}^{K^\circ}(W)'\mathsf{e})^2 = 1$*. Thus,* $\log N_{[]}(\epsilon, \mathcal{F}_{h,\mathsf{e}}, L^2(Z)) \leq \log N_{[]}(\epsilon, \mathcal{H}_{0,J^\circ}, L^\infty)$ *where the later is bounded by* $\epsilon^{-d_x/p}$ *(within a constant) if the functions in* $\mathcal{H}_0$ *have uniformly bounded partial derivatives with highest order derivatives being Lipschitz, see van der Vaart and Wellner [2000, Theorem 2.7.1]. We obtain* $\mathcal{C}_h \lesssim 1$ *under the condition* $2p \geq d_x$*, which is satisfied given Assumption 4(iii). In this case, a sufficient condition for Assumption 5(ii) is given by* $\mathrm{P}_h\big(\zeta_{J^\circ} \|\widehat{h}_{J^\circ}^{\mathrm{R}} - \Pi_{\mathcal{H}_0} h\|_{L^2(X)} > C\big) \to 0$ *uniformly for* $h \in \mathcal{H}_1(\delta^\circ \mathrm{r}_n)$*, which is less restrictive than Assumption 5(i) since the sieve dimension is fixed at* J°*. When the basis functions in* $\widetilde{b}^{K^\circ}$ *are uniformly bounded, such as for trigonometric bases, we immediately obtain* $\mathcal{C}_h \lesssim 1$*. If* $\mathcal{H}_0$ *consists of convex functions that are Lipschitz and map a compact and convex set in* $\mathbb{R}$ *to* $[0,1]$*, then* $\mathcal{C}_h \lesssim 1$ *by van der Vaart and Wellner [2000, Corollary 2.7.10].*

The next result establishes an upper bound for the rate of testing under a composite null hypothesis using the test statistic $\widehat{\mathtt{T}}_n$ given in (2.13).

Theorem 4.2. *Let Assumptions 1(i)-(iii), 2(i), 4, and 5 hold. Consider testing the composite null* $\mathcal{H}_0$ *versus* $\mathcal{H}_1(\delta^\circ \mathsf{r}_n) = \{h \in \mathcal{H} : \|h - \mathcal{H}_0\|_{L^2(X)} \geq \delta^\circ \mathsf{r}_n\}$ *for a constant* $\delta^\circ > 0$ *and the adaptive (separation) rate* $\mathsf{r}_n = (J^\circ)^{-p/d_x}$ *given in Theorem 4.1. Then, for any* $\alpha \in (0,1)$ *we have*

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_0} \mathrm{P}_h(\widehat{\mathtt{T}}_n = 1) \leq \alpha \qquad \text{and} \qquad \lim_{n\to\infty} \sup_{h\in\mathcal{H}_1(\delta^\circ \mathsf{r}_n)} \mathrm{P}_h(\widehat{\mathtt{T}}_n = 0) = 0. \tag{4.7}$$

In addition, let Assumptions 1(iv) and 3 hold with $\{\nu_j\}$ *regularly varying. Then:* $s_J \sim \nu_J$*, Assumption 4(ii) holds, and* $J^\circ \sim \max\left\{J : n^{-1}\sqrt{\log\log n}\,\nu_J^{-2}\sqrt{J} \leq J^{-2p/d_x}\right\}$.

(1) Mildly ill-posed case: $\mathsf{r}_n \sim \left(\sqrt{\log\log n}/n\right)^{2p/(4(p+a)+d_x)}$.

(2) Severely ill-posed case: $\mathsf{r}_n \sim (\log n)^{-p/a}$.

Theorem 4.2 states that $\widehat{\mathtt{T}}_n$ attains the same adaptive rate of testing r_n for a composite null as that for a simple null. This theorem also provides the asymptotic size control for composite hypothesis testing, which is established by controlling the sieve approximation error uniformly over the index set $\widehat{\mathcal{I}}_n$ under the null, thanks to a projection property built in the construction of our test $\widehat{\mathtt{T}}_n$; see Lemma B.9.

Adaptive Testing in Semiparametric Models. Partially parametric models are often used in empirical work and can be easily incorporated in our framework either as restricted models or as the maintained models. Let $\Theta \oplus \mathcal{G} = \{h(x_1, x_2) = x_1'\theta + g(x_2) : \theta \in \Theta, g \in \mathcal{G}\}$, where Θ denotes a finite dimensional parameter space, and $\mathcal{G}$ denotes a class of nonparametric functions.

Let the NPIV model (2.1) be the maintained hypothesis. We can test inequality restrictions as in Example 2.1 and a semiparametric structure simultaneously. For example, we can test for a partial linear structure with an nondecreasing function g by setting $\mathcal{H}_0 = \{h \in \Theta \oplus \mathcal{G} : \partial_{x_2} g \geq 0\}$. The class of alternative functions can then be written as $\mathcal{H}_1(r_n) := \left\{g \in \mathcal{G} : \|g - \mathcal{G}_0\|_{L^2(X_2)} \geq r_n\right\}$ where $\mathcal{G}_0 = \{g \in \mathcal{G} : \partial_{x_2} g \geq 0\}$ and the rate of testing r_n does not depend on the dimensionality of X_1. We can also test for the nonnegativity of the coefficient θ and a partial linear restriction by setting $\mathcal{H}_0 = \{h \in \Theta \oplus \mathcal{G} : \partial_{x_1} h \geq 0\}$. As in Example 2.2, we can test semiparametric equality restriction by taking $\mathcal{H}_0 = \Theta \oplus \mathcal{G}$.

Let the partial linear IV model be the maintained hypothesis in model (2.1) with $\mathcal{H} = \Theta \oplus \mathcal{G}$. Note that the maintained partial linear structure can be easily enforced in the sieve space used to estimate the unconstrained NPIV function. Monotonicity in all arguments of h can be imposed by $\mathcal{H}_0 = \{h \in \Theta \oplus \mathcal{G} : \theta \geq 0, \partial_{x_2} g \geq 0\}$. We also allow for second or higher order derivatives in the hypotheses considered above. For instance, we impose a partial linear structure $\mathcal{H}$ in our empirical illustration on demand for differential products in Section 6.1.

4.3. Confidence Sets in L^2

One can construct L^2-confidence sets by inverting our adaptive test for a NPIV function. The resulting confidence sets impose conjectured restrictions on the function of interest h. The $(1-\alpha)-$confidence set for a NPIV function h is given by

$$\mathcal{C}_n(\alpha) = \Big\{ h \in \mathcal{H}_0 : \frac{n\widehat{D}_J(h)}{\widehat{\mathrm{v}}_J} \leq \widehat{\eta}_J(\alpha) \quad \text{for all } J \in \widehat{\mathcal{I}}_n \Big\}. \tag{4.8}$$

This confidence set does not depend on additional tuning parameters. The following corollary exploits our previous results to characterize the asymptotic size and power properties of our procedure.

Corollary 4.1. *Let Assumptions 1(i)-(iii), 2(i), 4, and 5 be satisfied. Let* $\mathsf{r}_n = (J^\circ)^{-p/d_x}$ *be the adaptive rate of testing given in Theorem 4.2. Then, for any* $\alpha \in (0,1)$ *it holds*

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_0} \mathrm{P}_h\left(h \notin \mathcal{C}_n(\alpha)\right) \leq \alpha \tag{4.9}$$

and there exists a constant $\delta^\circ > 0$ *such that*

$$\lim_{n\to\infty} \inf_{h\in\mathcal{H}_1(\delta^\circ \mathsf{r}_n)} \mathrm{P}_h\left(h \notin \mathcal{C}_n(\alpha)\right) = 1. \tag{4.10}$$

Corollary 4.1 result (4.9) shows that the L^2-confidence set $\mathcal{C}_n(\alpha)$ controls size uniformly over the class of functions $\mathcal{H}_0$. Moreover, result (4.10) establishes power uniformly over the class $\mathcal{H}_1(\delta^\circ \mathsf{r}_n)$. We immediately see from Corollary 4.1 that the diameter of the L^2-confidence ball, $\mathsf{diam}(\mathcal{C}_n(\alpha)) = \sup\big\{\|h_1 - h_2\|_{L^2(X)} : h_1, h_2 \in \mathcal{C}_n(\alpha)\big\}$, depends on the degree of ill-posedness captured by the singular value s_{J°.

Corollary 4.2. *Let Assumptions 1(i)-(iii), 2(i), 4, and 5 be satisfied. Then, for any* $\alpha \in (0,1)$ *we have* $\sup_{h\in\mathcal{H}_0} \mathrm{P}_h\left(\mathsf{diam}(\mathcal{C}_n(\alpha)) \geq C\mathsf{r}_n\right) = o(1)$, *for some constant* $C > 0$ *and the adaptive rate* $\mathsf{r}_n = (J^\circ)^{-p/d_x}$ *given in Theorem 4.2.*

Corollary 4.2 yields a confidence set whose diameter shrinks to zero at the adaptive optimal-testing rate (of the order $(J^\circ)^{-p/d_x}$) and whose implementation does not require specifying the values of any unknown regularity parameters. Our confidence set $\mathcal{C}_n(\alpha)$ thus adapts to the unknown smoothness p of the unrestricted NPIV functions.

Let $\mathcal{H}_0$ be a Sobolev class of smoothness $p_0 > p$. It is known in statistical Gaussian White noise and regression models (see Robins and Van Der Vaart [2006] and Cai and

Low [2006]) that rate adaption is only possible over submodels $\mathcal{H}_0$ such that the rate of estimation over the submodel is strictly larger than the rate of testing over the "supermodel" $\mathcal{H}$. This suggests that it is impossible to adapt to the smoothness p_0 for severely ill-posed NPIV models. In the mildly ill-posed case, this leads to the restriction $n^{-p/(2(p+a)+d_x/2)} < n^{-p_0/(2(p_0+a)+d_x)}$. This condition translates into the smoothness restriction $p_0 < p\, c_a$ where $c_a = (4a+2d_x)/(4a+d_x)$ and hence, requires $p_0 \in \big(p, c_a p\big)$. The constant c_a is close to one even under modest degrees of ill-posedness. Consequently, the gain from adaptation with respect to the smoothness of restricted classes of NPIV functions is very limited.

5. Monte Carlo Studies

This section presents Monte Carlo performance of our adaptive test for monotonicity and parametric form of a NPIV function using simulation designs based on Chernozhukov et al. [2015]. See the online Appendix C for additional simulation results using other designs. All the simulation results are based on 5000 Monte Carlo replications for every experiment. In addition, Breunig and Chen [2020] provide evidence from simulations and real data applications that, in terms of finite-sample size and power, our adaptive test and its bootstrapped version perform similarly well. We no longer present such results here due to the lack of space.

For all the Monte Carlo designs in this section, Y is generated according to the NPIV model (2.1) for scalar-valued random variables X and W. We let $X_i = \Phi(X_i^*)$ and $W_i = \Phi(W_i^*)$ where Φ denotes the standard normal distribution function, and generate the random vector (X_i^*, W_i^*, U_i) according to

$$\begin{pmatrix} X_i^* \\ W_i^* \\ U_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \xi & 0.3 \\ \xi & 1 & 0 \\ 0.3 & 0 & 1 \end{pmatrix}\right). \tag{5.1}$$

The parameter ξ captures the strength of instruments and varies in the experiments below. As ξ increases, the instrument becomes stronger (or the ill-posedness gets weaker). While Chernozhukov et al. [2015] fixed $\xi = 0.5$ in their design, we let $\xi \in \{0.3, 0.5, 0.7\}$ in our simulation studies. The functional form of h varies in different Monte Carlo designs below.

5.1. Adaptive Testing for Monotonicity

We generate Y according to (2.1) and (5.1), using h from Chernozhukov et al. [2015] design:

$$h(x) = c_0\Big[1 - 2\Phi\Big(\frac{x - 1/2}{c_0}\Big)\Big] \quad \text{for some constant} \quad 0 < c_0 \leq 1. \tag{5.2}$$

This function $h(x)$ is decreasing in x, where c_0 captures the degree of monotonicity. We note that $h(x) \approx 0$ for c_0 close to zero and $h(x) \approx \phi(0)(1-2x)$ for c_0 close to one, where ϕ denotes the standard normal probability density function. The null hypothesis is that the NPIV function h is weakly decreasing on the support of X.

We implement our adaptive test statistic $\widehat{\mathtt{T}}_n$ given in (2.13) using quadratic B-spline basis functions with varying number of knots for h. Due to piecewise linear derivatives, monotonicity constraints are easily imposed on the restricted function at the derivative at $J-1$ points. For the instrument sieve $b^{K(J)}(W)$ we also use quadratic B-spline functions with a larger number of knots with $K(J) = 2J$ or $K(J) = 4J$. Implementation of the restricted sieve NPIV estimator $\widehat{h}_J^{\mathtt{R}}$ is straightforward using the R package `coneproj`. We also compare our adaptive test to a nonadaptive bootstrap test $\mathtt{T}^{\mathtt{B}}_{n,3}$ of Fang and Seo [2021] with a deterministic choice of sieve dimension $J = 3$ and $K = 2J = 6$ or $K = 4J = 12$. Their statistic $\mathtt{T}^{\mathtt{B}}_{n,3}$ is computed using a standard Gaussian multiplier bootstrap critical values $\widehat{\eta}_J(\alpha)$ (and their other recommended tuning parameters of $c_n = (\log J)^{-1}$ and $\gamma_n = 0.01/\log n$) with $J = 3$. In our simulations we use 200 bootstrap iterations.

Size. Table 1 presents the average data-driven choice of tuning parameter J, denoted by $\widehat{J}$, at the nominal level $\alpha = 0.05$. Specifically, $\widehat{J}$ is the average choice of J that maximizes $\widehat{\mathcal{W}}_J(\alpha)$ over the RES index set $\widehat{\mathcal{I}}_n$ when the null is not rejected; and is the smallest $J \in \widehat{\mathcal{I}}_n$ such that $\widehat{\mathcal{W}}_J(\alpha) > 1$ when the null is rejected. This data-driven choice of J corresponds to *early stopping* when the null is rejected. Table 1 shows that, for the same sample size n the average data-driven choice $\widehat{J}$ increases as the instrument strength (captured by the parameter ξ) increases; while for the same instrument strength ξ, $\widehat{J}$ weakly increases as the sample size n increases. Table 1 also reveals little difference between the choices $K(J) = 2J$ and $K(J) = 4J$, especially so for larger sample sizes. This is consistent with our theory that J, the dimension of the sieve basis used to approximate the unrestricted NPIV function h, is the key tuning parameter in our minimax rate adaptive testing.

Table 1 also reports empirical rejection probabilities under the null hypothesis using our adaptive test $\widehat{\mathtt{T}}_n$ and Fang and Seo [2021]'s nonadaptive bootstrap test $\mathtt{T}^{B}_{n,3}$ (with $J = 3$). Results are presented for different nominal levels $\alpha \in \{0.05, 0.01\}$, different sample sizes $n \in \{500, 1000, 5000\}$, different instrument strength $\xi \in \{0.3, 0.5, 0.7\}$, and different degree of monotonicity $c_0 \in \{0.01, 0.1, 1\}$. Overall, we see from Table 1 that both tests provide adequate size control across different sample size n, instrument strength ξ, and degree of monotonicity c_0. In particular, both tests for the null of monotonicity are similarly undersized.

Power. We next examine the rejection probabilities of our adaptive test when the data is generated according to (2.1) and (5.1) using the NPIV function

$$h(x) = -x/5 + c_A \left(x^2 + c_B \sin(2\pi x)\right), \tag{5.3}$$

n	c_0	ξ	$\widehat{\mathtt{T}}_n$ *with* $K=2J$ 5%	1%	$\widehat{J}$ *at 5%*	$\widehat{\mathtt{T}}_n$ *with* $K=4J$ 5%	1%	$\widehat{J}$ *at 5%*	$\mathtt{T}^B_{n,3}$ with $K=2J$ 5%	1%	$\mathtt{T}^B_{n,3}$ with $K=4J$ 5%	1%
500	0.01	0.3	0.009	0.001	3.00	0.023	0.003	3.02	0.004	0.001	0.031	0.003
		0.5	0.017	0.002	3.31	0.023	0.003	3.35	0.021	0.002	0.022	0.004
		0.7	0.026	0.004	3.56	0.028	0.009	3.57	0.016	0.002	0.014	0.010
	0.1	0.3	0.006	0.001	3.00	0.015	0.002	3.03	0.004	0.000	0.025	0.002
		0.5	0.013	0.001	3.34	0.016	0.002	3.38	0.016	0.001	0.018	0.003
		0.7	0.017	0.003	3.65	0.021	0.007	3.63	0.010	0.001	0.008	0.000
	1	0.3	0.004	0.000	3.00	0.010	0.001	3.03	0.002	0.000	0.019	0.002
		0.5	0.007	0.000	3.38	0.008	0.001	3.41	0.011	0.000	0.012	0.001
		0.7	0.007	0.001	3.76	0.011	0.002	3.74	0.004	0.000	0.003	0.000
1000	0.01	0.3	0.010	0.001	3.01	0.019	0.002	3.07	0.012	0.001	0.035	0.003
		0.5	0.011	0.003	3.55	0.021	0.003	3.48	0.026	0.003	0.026	0.003
		0.7	0.028	0.006	3.99	0.031	0.005	4.08	0.014	0.001	0.014	0.001
	0.1	0.3	0.007	0.000	3.02	0.015	0.001	3.07	0.010	0.001	0.028	0.003
		0.5	0.005	0.002	3.63	0.014	0.003	3.54	0.018	0.003	0.018	0.002
		0.7	0.018	0.004	4.19	0.022	0.003	4.28	0.008	0.001	0.007	0.001
	1	0.3	0.005	0.000	3.02	0.008	0.001	3.07	0.006	0.000	0.018	0.003
		0.5	0.005	0.000	3.63	0.006	0.001	3.54	0.008	0.001	0.010	0.001
		0.7	0.006	0.002	4.19	0.008	0.002	4.28	0.002	0.000	0.002	0.000
5000	0.01	0.3	0.016	0.002	3.40	0.020	0.004	3.39	0.028	0.003	0.037	0.007
		0.5	0.030	0.006	3.67	0.031	0.008	3.75	0.029	0.005	0.028	0.004
		0.7	0.035	0.010	4.41	0.038	0.009	4.44	0.013	0.002	0.011	0.002
	0.1	0.3	0.011	0.001	3.48	0.015	0.002	3.40	0.021	0.003	0.028	0.005
		0.5	0.017	0.003	3.88	0.019	0.005	3.93	0.013	0.002	0.013	0.002
		0.7	0.022	0.008	4.77	0.023	0.007	4.77	0.003	0.000	0.003	0.000
	1	0.3	0.006	0.001	3.48	0.008	0.001	3.40	0.012	0.001	0.016	0.002
		0.5	0.004	0.001	3.88	0.007	0.002	3.93	0.003	0.000	0.002	0.000
		0.7	0.008	0.003	4.77	0.006	0.001	4.77	0.000	0.000	0.000	0.000

Table 1: Testing Monotonicity – Empirical size of our adaptive test $\widehat{\mathtt{T}}_n$ (with average value $\widehat{J}$) and of the nonadaptive bootstrap test $\mathtt{T}^B_{n,3}$ (with a fixed $J=3$). True DGP from Section 5.1 using NPIV function (5.2). Instrument strength increases in ξ.

where $c_A \in [0,2]$ and $c_B \in \{0, 0.5, 1\}$. The null hypothesis is that the NPIV function $h(\cdot)$ is weakly decreasing over the support of X. When $c_B = 0$ the null is satisfied only if $c_A \leq 0.1$. When $c_B = 0.5$ the null hypothesis is satisfied only if $c_A \leq 0.1/(1+\pi/2) \approx 0.04$. When $c_B = 1$ the null is satisfied only if $c_A \leq 0.1/(1+\pi) \approx 0.02$.

Figure 1 depicts the empirical power function of our adaptive test $\widehat{\mathtt{T}}_n$ (blue solid lines), and of the nonadaptive bootstrap test $\mathtt{T}^{\mathtt{B}}_{n,3}$ (green dashed lines, with fixed $J=3$), with $K(J) = 4J$, under the 5% nominal level for different instrument strength $\xi \in \{0.3, 0.5, 0.7\}$, and sample size $n = 500$.[2] Figure 2 shows these power curves for a larger sample size $n = 5000$. From both figures, we see that our adaptive test becomes more powerful, for $c_A > 0.1$, as the parameter of instrument strength ξ and the sample size n increase.

Figures 1 and 2 highlight the importance of adaptation for the power of nonparametric monotonicity tests. When the alternative is of a simple quadratic form (i.e., $c_B = 0$),

[2] The finite-sample power of our adaptive test with $K(J) = 2J$ is slightly smaller than that with $K = 4J$ when $n = 500$, but the power difference disappears when n becomes larger.

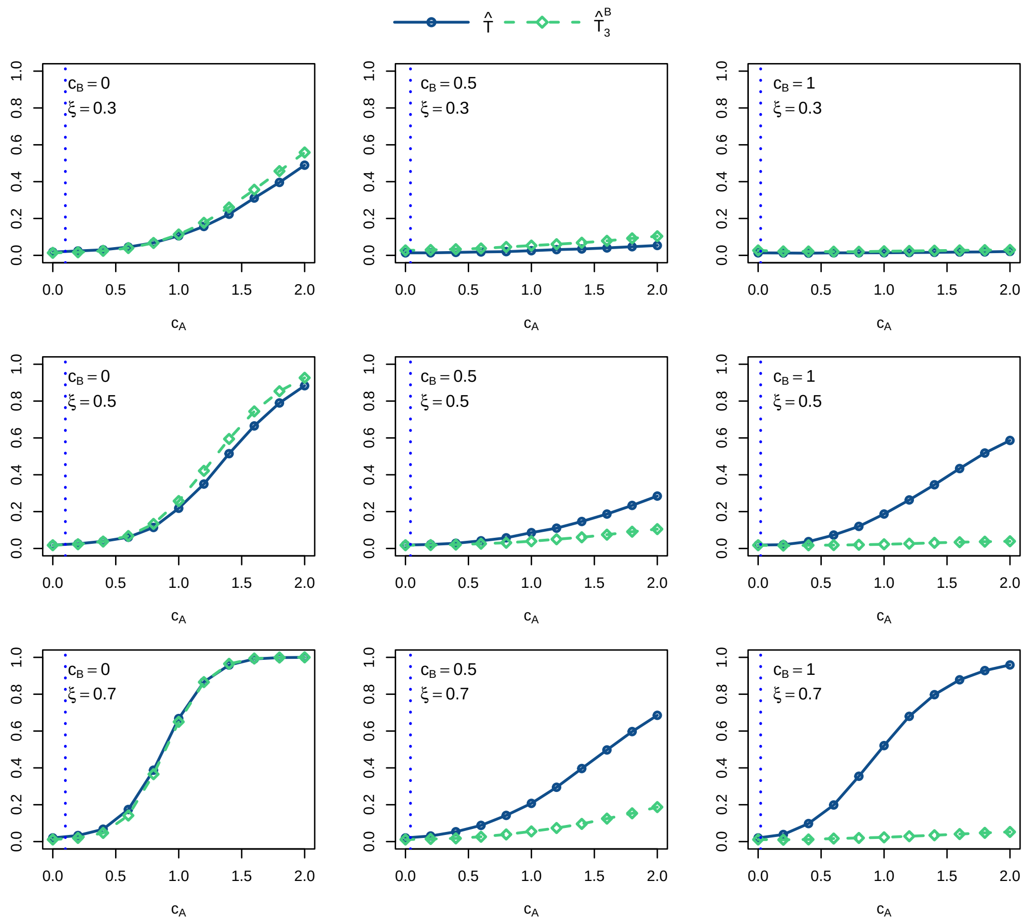


Figure 1: Testing Monotonicity – Empirical power of our adaptive test $\widehat{\mathtt{T}}_n$ (blue solid lines) and of nonadaptive bootstrap test $\mathtt{T}^{\mathtt{B}}_{n,3}$ (green dashed lines) with $K = 4J$ and $n = 500$. True DGP from Section 5.1 using NPIV function (5.3). The vertical dotted line indicates when the null hypothesis is violated. Alternatives are quadratic when $c_B = 0$ and become more complex as $c_B > 0$ increases. Instrument strength increases in ξ.

then there is little difference between our adaptive test $\widehat{\mathtt{T}}_n$ and the nonadaptive bootstrap test $\mathtt{T}^{\mathtt{B}}_{n,3}$. But, as the amount of nonlinearity increases with the constant $c_B > 0$, the nonadaptive bootstrap test becomes much less powerful than our adaptive test. For a fixed dimension parameter J, a test can have high power in a certain direction but might not be capable of detecting other nonlinearities.

Remark 5.1 (Another NPIV Monotonicity Design)**.** *In Online Appendix C we present another simulation design for testing for monotonicity. We generate Y according to the NPIV model (2.1), where $h(x) = c_0(x/5 + x^2) + c_A \sin(2\pi x)$, and (W^*, ϵ, ν) follows a multivariate standard normal distribution. Let $W = \Phi(W^*)$, $X = \Phi\big(\xi W^* + \sqrt{1-\xi^2}\epsilon\big)$, and $U = (0.3\epsilon + \sqrt{1-(0.3)^2}\nu)/2$. This design with $(c_0, c_A) = (1, 0)$ and $\xi \in \{0.3, 0.5\}$ coincides with the one in Chetverikov and Wilhelm [2017]. Simulation results for testing a null of weakly increasing NPIV function using this design reveal size and power patterns that are very similar to the ones reported in this subsection. See Appendix C for details.*

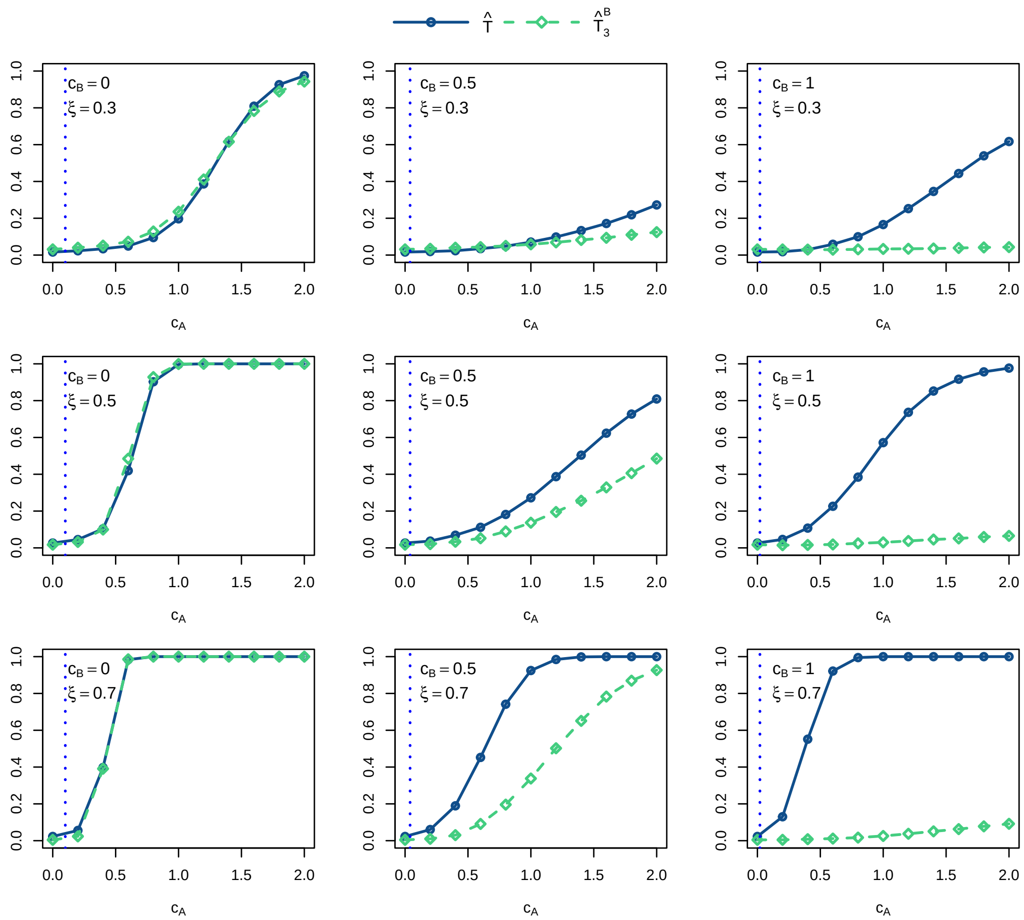


Figure 2: Testing Monotonicity – Replication of Figure 1 with $n = 5000$.

5.2. Testing for Parametric Restrictions

We now test for a parametric specification. We assume that the data is generated according to the design (2.1) and (5.1) with the NPIV function h given by (5.3) with $c_A \in [0, 4]$ and $c_B \in \{0,\ 0.5,\ 1\}$. The null hypothesis is h being linear (i.e., $c_A = c_B = 0$).

We implement our adaptive test $\widehat{\mathrm{T}}_n$ given in (2.13) using quadratic B-spline basis functions with varying number of knots and where the constrained function coincides with the parametric 2SLS estimator. The number of knots varies within the RES index set $\widehat{\mathcal{I}}_n$ as implemented in the last subsection, with $K(J) = 2J$ and $K(J) = 4J$. We compare our adaptive test to the asymptotic t-test and the test by Horowitz [2006] (denoted by JH).[3] To compute the JH test that involves kernel density estimation, we follow Horowitz [2006] to estimate the joint density f_{XW} using the kernel $K(v) = (15/16)(1 - v^2)^2 \mathbb{1}\{|v| \leq 1\}$, with the kernel bandwidth chosen via cross-validation that minimizes mean squared error of density estimation for $n = 500, 1000, 5000$. For the large sample size $n = 10000$, it

[3]Horowitz [2006] already demonstrated in his simulation studies, with a sample size $n = 500$ and 1000 Monte Carlo replications, that his test is more powerful than several existing tests including Bierens [1990]'s.

is computationally too demanding to perform cross-validated bandwidth for JH test with 5000 Monte Carlo repetitions, however. Instead, we have tried a few bandwidth choices and have found that a fixed bandwidth of 0.06 gives good empirical size control for JH test when $n = 10000$ across 5000 Monte Carlo replications, which is what we report in Table 2.

n	ξ	$\widehat{\mathtt{T}}_n$ with $K = 2J$	$\widehat{J}$	$\widehat{\mathtt{T}}_n$ with $K = 4J$	$\widehat{J}$	t-test	JH test
500	0.3	0.010	3.00	0.023	3.03	0.001	0.053
	0.5	0.023	3.34	0.030	3.50	0.024	0.057
	0.7	0.030	3.61	0.032	3.63	0.042	0.054
1000	0.3	0.013	3.01	0.023	3.07	0.005	0.055
	0.5	0.020	3.52	0.030	3.50	0.038	0.055
	0.7	0.036	3.91	0.039	4.00	0.049	0.056
5000	0.3	0.022	3.38	0.028	3.41	0.029	0.057
	0.5	0.039	3.59	0.042	3.64	0.048	0.056
	0.7	0.045	4.18	0.048	4.18	0.050	0.056
10000	0.3	0.030	3.49	0.035	3.45	0.036	0.054
	0.5	0.042	3.85	0.051	3.97	0.047	0.051
	0.7	0.055	4.18	0.055	4.17	0.048	0.056

Table 2: Testing Parametric Form – Empirical size of our adaptive test $\widehat{\mathtt{T}}_n$ (with Monte Carlo average value $\widehat{J}$), the t-test and JH test. 5% nominal level. True DGP from Section 5.2 using NPIV function (5.3) with $c_A = c_B = 0$. Instrument strength increases in ξ.

Size. Table 2 reports empirical rejection probabilities under the null hypothesis of linearity of h, of the tests at the nominal level $\alpha = 0.05$. It also reports $\widehat{J}$ (which is defined the same way as that in Table 1) at the nominal level $\alpha = 0.05$. Results are presented under different sample sizes $n \in \{500, 1000, 5000, 10000\}$ and instrument strength $\xi \in \{0.3, 0.5, 0.7\}$. We note that $\widehat{J}$ is again weakly increasing with sample size and with instrument strength. Overall, our adaptive test $\widehat{\mathtt{T}}_n$ provides adequate size control for different sample size n and different instrument strength ξ. The difference in empirical size of our adaptive tests between $K(J) = 2J$ and $K(J) = 4J$ is again small, which is consistent with our theory.

Power. Figure 3 provides empirical power curves for the 5% level tests with sample size $n = 500$. See Figure B in the online Appendix C for empirical power curves with a larger sample size $n = 5000$. From both figures, we see that our adaptive test $\widehat{\mathtt{T}}_n$ (blue solid lines) with $K(J) = 4J$ has power similar to the asymptotic t-test (red dotted lines) and the JH test (green dashed lines) for a simple quadratic alternative with $c_B = 0$. When the alternative function in (5.3) becomes more complex with $c_B \in \{0.5, 1\}$, our adaptive test becomes more powerful than the JH test. This is theoretically sensible since Horowitz [2006] test is designed to have power against $n^{-1/2}$ smooth alternative only. To sum up, our adaptive minimax test not only controls size, but also has very good finite-sample power uniformly against a large class of nonparametric alternatives.

Finally in online Appendix C we present additional simulation comparisons of our adaptive test against our adaptive version of Bierens [1990]'s type test when the dimension of conditional instrument W is larger than the dimension of the endogenous variables X.

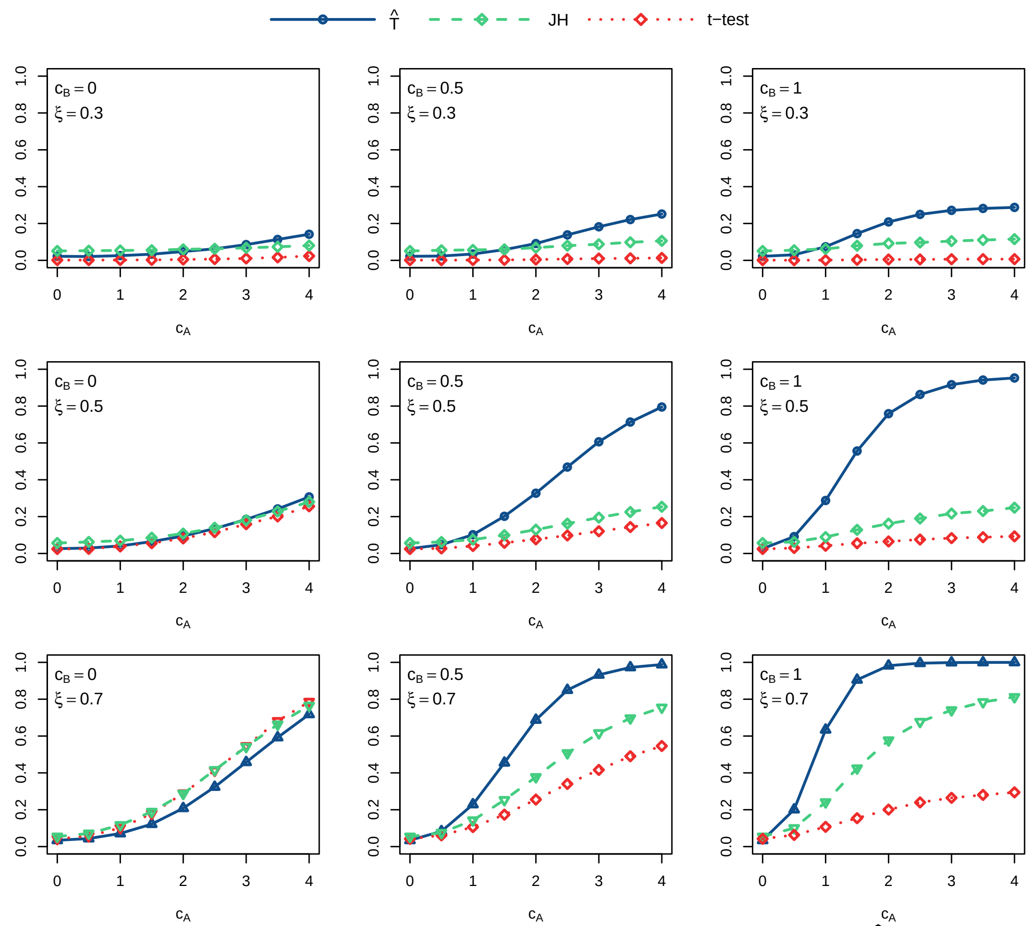


Figure 3: Testing Parametric Form – Empirical power of our adaptive test $\widehat{\mathtt{T}}_n$ (blue solid lines) with $K = 4J$, of JH test (green dashed lines) and of t-test (red dotted lines). $n = 500$. True DGP from Section 5.2 using NPIV function (5.3). Alternatives are quadratic when $c_B = 0$ and more complex as $c_B > 0$ increases. Instrument strength increases in ξ.

We observe that our adaptive test $\widehat{\mathtt{T}}_n$ again have very good size control and even better finite-sample power when $d_w > d_x$.

6. Empirical Applications

We present two empirical applications of our adaptive test for NPIV models. The first one tests for connected substitutes restrictions in differentiated products demand using market level data. The second one tests for monotonicity, convexity or parametric specification of Engel curves for non-durable good consumption using household level data.

In both empirical applications, we implement our adaptive test $\widehat{\mathtt{T}}_n$ given in (2.13) with $K(J) = 4J$. The null hypothesis is rejected at the nominal level $\alpha = 0.05$ whenever $\widehat{\mathcal{W}}_J(\alpha) > 1$ for some $J \in \widehat{\mathcal{I}}_n$ (the RES index set). Tables in this section report a set $\widehat{\mathcal{J}} \subset \widehat{\mathcal{I}}_n$, which equals to $\arg\max_{J \in \widehat{\mathcal{I}}_n} \widehat{\mathcal{W}}_J(\alpha)$ when our test fails to reject the null hypothesis and equals to $\{J \in \widehat{\mathcal{I}}_n : \widehat{\mathcal{W}}_J(\alpha) > 1\}$ when our test rejects the null. Below, we report $\widehat{\mathcal{W}}_{\widehat{\mathcal{J}}}$

with $\widehat{J}$ being the minimal integer of $\widehat{\mathcal{J}}$. We also report the corresponding p value, which should, by Bonferroni correction, be compared to the nominal level $\alpha = 0.05$ divided by the cardinality of $\widehat{\mathcal{I}}_n$. Finally, since our test is based on a leave-one-out version, the value of $\widehat{\mathcal{W}}_{\widehat{J}}$ could be negative.

6.1. Adaptive Testing for Connected Substitutes in Demand for Differential Products

Recently Berry and Haile [2014] provide conditions under which a nonparametric demand system for differentiated products can be inverted to NPIV equations using Market level data. A key restriction is what they call "connected substitutes". Compiani [2021] applies their nonparametric identification results and estimates the system of inverse demand by directly imposing the connected substitutes restrictions in his implementation of sieve NPIV estimator, and obtains informative results as an alternative to BLP demand in simulation studies and a real data application.

We revisit Compiani [2021]'s empirical application using the 2014 Nielsen scanner data set that contains market (store/week) level data of consumers in California choosing from organic strawberries, non-organic strawberries and an outside option. While Compiani [2021] directly imposes "connected substitutes" restriction in his sieve NPIV estimation of inverse demand, we want to test this restriction. Following Compiani [2021] we consider

$$X_o + U = h(\mathbf{P}, S_o, S_{no}, In), \qquad \mathrm{E}[U|\mathbf{W}_p, X_o, X_{no}, In] = 0,$$

where h denotes the inverse of the demand for organic strawberries, X_o denotes a measure of taste for organic products, X_{no} denotes the availability of other fruit, S_o and S_{no} denote the endogenous shares of the organic and non-organic strawberries, respectively. (X_o, X_{no}) are the two included instruments for the two endogenous shares (S_o, S_{no}). In denotes store-level (zip code) income and U unobserved shocks for organic produce. The vector $\mathbf{P} = (P_o, P_{no}, P_{out})$ denotes the endogenous prices of organic strawberries, non-organic strawberries, and non-strawberry fresh fruit, respectively. We follow Compiani [2021] and let $\mathbf{W}_p = (W_o, W_{no}, W_{out}, W_{s1}, W_{s2})$ be a $5-$dimensional vector of conditional instruments for the price vector $\mathbf{P}$, including 3 Hausman-type instrumental variables (W_o, W_{no}, W_{out}) and 2 shipping-point spot prices (W_{s1}, W_{s2}) (as proxies for the wholesale prices faced by retailers).

As shown by Compiani [2021, Lemma 1], the connected substitutes assumption of Berry and Haile [2014] implies the following shape restrictions on the function h: First, h is weakly increasing in the organic product price P_o. Second, h is weakly increasing in the organic product share S_o. Third, h is weakly increasing in the non-organic product share S_{no}. Fourth, $\partial h/\partial s_o \geq \partial h/\partial s_{no}$ (the so-called diagonal dominance). Below, we test for these inequality restrictions.

We consider a subset of the data set of Compiani [2021][4], where income ranges from the first and to the third quartile of its distribution and prices for organic produces are restricted to be above its 1st and below its 99th percentile. The resulting sample has size $n = 11910$. We implement our adaptive test $\widehat{\mathtt{T}}_n$ by making use of a semiparametric specification of the function h: we consider the tensor product of quadratic B-splines $\psi^{J_1}(P_o)$ and the vector $(1, In, P_{no}, \psi^3(S_o))$, where we use a cubic B-spline transformation of S_o without knots and without intercept, hence $J = 6J_1$. The variables $(P_{out}, S_{no}, S_{no}P_{no}, S_{no}S_o)$ are included additively and we set $K(J) = 4J$. We obtain the index set $\widehat{\mathcal{I}}_n = \{24, 30, 36\}$.

H_0	$\widehat{\mathcal{W}}_{\widehat{\mathcal{J}}}$	p val.	reject H_0?	$\widehat{\mathcal{J}}$
$\partial h/\partial p_o \geq 0$	0.854	0.0280	no	$\{36\}$
$\partial h/\partial p_o \leq 0$	2.635	0.000	yes	$\{30, 36\}$
$\partial h/\partial s_o \geq 0$	0.663	0.057	no	$\{36\}$
$\partial h/\partial s_o \leq 0$	2.022	0.001	yes	$\{24, 30, 36\}$
$\partial h/\partial s_{no} \geq 0$	-0.119	0.545	no	$\{24\}$
$\partial h/\partial s_{no} \leq 0$	-0.234	0.727	no	$\{24\}$
$\partial h/\partial s_o \geq \partial h/\partial s_{no}$	0.663	0.055	no	$\{36\}$
$\partial h/\partial s_o \leq \partial h/\partial s_{no}$	2.022	0.00	yes	$\{24, 30, 36\}$

Table 3: Adaptive testing for the shape of h.

According to Table 3, our adaptive test fails to reject that h is weakly increasing in the own price at the nominal level $\alpha = 0.05$, but rejects $\partial h/\partial p_o \leq 0$. Similarly, this table shows that our adaptive test also fails to reject that h is weakly increase in non-organic shares and rejects that h is weakly decreasing in S_o. When testing partial derivatives, our test fails to reject that the partial effect with respect to the non-organic share is constant. Finally, the last two rows show that our test provides empirical evidence for the diagonal dominance restriction.

6.2. Adaptive Testing for Engel Curves

The system of Engel curves plays a central role in the analysis of consumer demand for non-durable goods. It describes the $i-$th household's budget share $Y_{\ell,i}$ for non-durable goods ℓ as a function of its log-total expenditure X_i and other exogenous characteristics such as family size and age of the head of the $i-$th household. The most popular class of parametric demand systems is the almost ideal class, pioneered by Deaton and Muellbauer [1980], where budget shares are assumed to be linear in log-total expenditure. Banks et al. [1997] propose a popular extension of this system of linear Engel curves to include a squared term in log-total expenditure, and their parametric Student t test rejects linear form in favor of quadratic Engel curves.

[4]For details on the construction of the data and descriptive statistics, see Compiani [2021, Appendix F].

Blundell et al. [2007] estimated a system of nonparametric Engel curves as functions of endogenous log-total expenditure and family size, using log-gross earnings of the head of household as a conditional instrument W. We use a subset of their data from the 1995 British Family Expenditure Survey, with the head of household aged between 20 and 55 and in work, and household with one or two children. This leaves a sample of size $n = 1027$. As an illustration we consider Engel curves $h_\ell(X)$ for four non-durable goods ℓ: "food in", "fuel", "travel", and "leisure": $\mathrm{E}[Y_\ell - h_\ell(X)|W] = 0$. We use the same quadratic B-spline basis with up to 3 knots to approximate all the Engel curves and set $K(J) = 4J$. Hence the index set $\widehat{\mathcal{I}}_n = \{3, 4, 5\}$ is the same for the different Engel curves.

	H_0: h is increasing				H_0: h is decreasing			
Goods	$\widehat{\mathcal{W}}_{\widehat{\mathcal{J}}}$	p value	reject H_0?	$\widehat{\mathcal{J}}$	$\widehat{\mathcal{W}}_{\widehat{\mathcal{J}}}$	p value	reject H_0?	$\widehat{\mathcal{J}}$
"food in"	2.871	0.000	yes	$\{3\}$	-0.324	1.000	no	$\{4\}$
"fuel"	8.192	0.000	yes	$\{3,4,5\}$	0.547	0.0375	no	$\{3\}$
"travel"	2.527	0.000	yes	$\{3,4\}$	0.381	0.075	no	$\{3\}$
"leisure"	0.299	0.114	no	$\{4\}$	4.552	0.000	yes	$\{3,4\}$

Table 4: Adaptive testing for monotonicity of Engel curves.

	H_0: h is convex				H_0: h is concave			
Goods	$\widehat{\mathcal{W}}_{\widehat{\mathcal{J}}}$	p value	reject H_0?	$\widehat{\mathcal{J}}$	$\widehat{\mathcal{W}}_{\widehat{\mathcal{J}}}$	p value	reject H_0?	$\widehat{\mathcal{J}}$
"food in"	-0.287	1.000	no	$\{4\}$	-0.324	1.000	no	$\{3\}$
"fuel"	-0.325	1.000	no	$\{3\}$	1.621	0.000	yes	$\{3\}$
"travel"	1.190	0.003	yes	$\{3\}$	-0.329	1.000	no	$\{5\}$
"leisure"	-0.197	0.818	no	$\{5\}$	0.690	0.022	no	$\{4\}$

Table 5: Adaptive testing for convexity/concavity of Engel curves.

	H_0: h is linear				H_0: h is quadratic			
Goods	$\widehat{\mathcal{W}}_{\widehat{\mathcal{J}}}$	p value	reject H_0?	$\widehat{\mathcal{J}}$	$\widehat{\mathcal{W}}_{\widehat{\mathcal{J}}}$	p value	reject H_0?	$\widehat{\mathcal{J}}$
"food in"	-0.273	0.922	no	$\{3\}$	0.125	0.233	no	$\{3\}$
"fuel"	1.623	0.000	yes	$\{3\}$	-0.120	0.612	no	$\{5\}$
"travel"	1.210	0.001	yes	$\{3\}$	-0.014	0.415	no	$\{4\}$
"leisure"	0.691	0.074	no	$\{4\}$	0.513	0.041	no	$\{4\}$

Table 6: Adaptive testing for linear/quadratic specification of Engel curves.

Table 4 reports our adaptive test for weak monotonicity of Engel curves. It shows that our test rejects increasing Engel curves for "food in", "fuel", and "travel" categories, and also rejects decreasing Engel curve for "leisure" at the 0.05 nominal level. Previously, to decide whether the Engel curves are strictly monotonic, estimated derivatives of these function together with their 95% uniform confidence bands were also provided in Chen and Christensen [2018, Figure 4]. Those uniform confidence bands are constructed using a sieve score bootstrapped critical values with non-data-driven choice of sieve dimension J, and contain zero almost over the whole support of household expenditure. It is interesting to see that our adaptive test is more informative about monotonicity in certain directions that are not obvious from their 95% uniform confidence bands. Table 5 reports our adaptive test for convexity and concavity of these Engel curves. At the 5% nominal level, we reject convexity of travel goods and reject concavity of Engel curves for fuel consumption. These are in line with Chen and Christensen [2018, Figure 4], but again, significant statements about the convexity/concavity of Engel curves are only possible using our adaptive testing procedure. Finally, Table 6 presents our adaptive tests for linear or quadratic specifications (against nonparametric alternatives) of the Engel curves for the four goods. At the nominal level $\alpha = 0.05$, this table shows that our adaptive test fails to reject a quadratic form for all the goods, while it rejects a linear Engel curve for fuel and travel goods. Our results are consistent with the conclusions obtained by Banks et al. [1997] using Student t-test for linear against quadratic forms of Engel curves.

References

J. Banks, R. Blundell, and A. Lewbel. Quadratic engel curves and consumer demand. *Review of Economics and Statistics*, 79(4):527–539, 1997.

H. H. Bauschke and J. M. Borwein. On the convergence of von neumann's alternating projection algorithm for two sets. *Set-Valued Analysis*, 1:185–212, 1993.

S. T. Berry and P. A. Haile. Identification in differentiated products markets using market level data. *Econometrica*, 82(5):1749–1797, 2014.

H. J. Bierens. A consistent conditional moment test of functional form. *Econometrica*, 58 (6):1443–1458, 1990.

R. Blundell, X. Chen, and D. Kristensen. Semi-nonparametric iv estimation of shape-invariant engel curves. *Econometrica*, 75(6):1613–1669, 2007.

R. Blundell, J. Horowitz, and M. Parey. Nonparametric estimation of a nonseparable demand function under the slutsky inequality restriction. *Review of Economics and Statistics*, 99(2):291–304, 2017.

C. Breunig. Goodness-of-fit tests based on series estimators in nonparametric instrumental regression. *Journal of Econometrics*, 184(2):328–346, 2015.

C. Breunig and X. Chen. Adaptive, rate-optimal testing in instrumental variables models. *arXiv preprint arXiv:2006.09587v1*, 2020.

T. T. Cai and M. G. Low. Adaptive confidence balls. *The Annals of Statistics*, 34(1): 202–228, 2006.

X. Chen and T. M. Christensen. Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric iv regression. *Quantitative Economics*, 9(1):39–84, 2018.

X. Chen and D. Pouzo. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80(1):277–321, 01 2012.

X. Chen and D. Pouzo. Sieve quasi likelihood ratio inference on semi/nonparametric conditional moment models. *Econometrica*, 83(3):1013–1079, 2015.

X. Chen and M. Reiß. On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 27(03):497–521, 2011.

X. Chen, T. Christensen, and S. Kankanala. Adaptive estimation and uniform confidence bands for nonparametric structural functions and elasticities. *arXiv preprint arXiv:2107.11869*, 2021.

V. Chernozhukov, W. K. Newey, and A. Santos. Constrained conditional moment restriction models. *arXiv preprint arXiv:1509.06311*, 2015.

D. Chetverikov. Testing regression monotonicity in econometric models. *Econometric Theory*, 35(4):729–776, 2019.

D. Chetverikov and D. Wilhelm. Nonparametric instrumental variable estimation under monotonicity. *Econometrica*, 85:1303–1320, 2017.

D. Chetverikov, A. Santos, and A. M. Shaikh. The econometrics of shape restrictions. *Annual Review of Economics*, 10:31–63, 2018.

O. Collier, L. Comminges, and A. B. Tsybakov. Minimax estimation of linear and quadratic functionals on sparsity classes. *The Annals of Statistics*, 45(3):923–958, 2017.

G. Compiani. Market counterfactuals and the specification of multi-product demand: A nonparametric approach. *Quantitative Economics*, forthcoming, 2021.

A. Deaton and J. Muellbauer. An almost ideal demand system. *American Economic Review*, pages 312–326, 1980.

Z. Fang and J. Seo. A projection framework for testing shape restrictions that form convex cones. *Econometrica*, 89(5):2439–2458, 2021.

J. Freyberger and B. Reeves. Inference under shape restrictions. Working paper, University of Wisconsin, 2019.

E. Guerre and P. Lavergne. Data-driven rate-optimal specification testing in regression models. *The Annals of Statistics*, 33(2):840–870, 2005.

P. Hall and J. L. Horowitz. Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics*, 33:2904–2929, 2005.

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

J. L. Horowitz and S. Lee. Uniform confidence bands for functions estimated nonparametrically with instrumental variables. *Journal of Econometrics*, 168(2):175–188, 2012.

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

Y. I. Ingster. Asymptotically minimax hypothesis testing for nonparametric alternatives. i, ii, iii. *Math. Methods Statist*, 2(2):85–114, 1993.

M. Reiß. Asymptotic equivalence for nonparametric regression with multivariate and random design. *The Annals of Statistics*, 36(4):1957–1982, 2008.

J. Robins and A. Van Der Vaart. Adaptive nonparametric confidence sets. *The Annals of Statistics*, 34(1):229–253, 2006.

A. Santos. Inference in nonparametric instrumental variables with partial identification. *Econometrica*, 80(1):213–275, 2012.

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

A. B. Tsybakov. *Introduction to nonparametric estimation.* Springer, 2009.

A. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics).* Springer, 2000.

Y. Zhu. Inference in nonparametric/semiparametric moment equality models with shape restrictions. *Quantitative Economics*, 11(2):609–636, 2020.

A. Proofs of Minimax Testing Results in Section 3

Proof of Theorem 3.1. We first derive the lower bound for testing simple null hypotheses. Let P_θ denote the joint distribution of (Y, X, W) satisfying $Y = Th_\theta + V$ with known operator T and $V|W \sim \mathcal{N}(0, \sigma^2)$, the so-called reduced-form nonparametric indirection regression (NPIR) model as in Chen and Reiß [2011] with fixed variance $\sigma^2 > 0$. To establish the lower bound, a consideration of the NPIR model is sufficient, as we show in the first inequality of (A.4) below.

By Reiß [2008], the reduced-form NPIR is asymptotic equivalent to the Gaussian white noise model $dY(w) = Th_\theta(w)dw + \frac{\sigma}{\sqrt{n}}dB(w)$ where dB is a Gaussian white noise in $L^2_{\mathcal{W}} := \{\phi : \int_{\mathcal{W}}[\phi(w)]^2 dw < \infty\}$ and, in particular, to the Gaussian sequence model $y_k = \int Th_\theta(w)\widetilde{b}_k(w)dw + \frac{\sigma}{\sqrt{n}}\xi_k$ where $y_k := \int \widetilde{b}_k(w)dY(w)$ and $\xi_k \sim \mathcal{N}(0, 1)$. Without loss of generality we let $h_0 = 0$. We introduce $\theta = (\theta_j)_{j\geq 1}$ with $\theta_j \in \{-1, 1\}$ and introduce the test function

$$h_\theta(\cdot) = \frac{\delta_*}{\sqrt{n}} \sum_{j=1}^{J_*} \nu_j^{-2}\theta_j\widetilde{\psi}_j(\cdot)\left(\sum_{j=1}^{J_*}\nu_j^{-4}\right)^{-1/4}, \tag{A.1}$$

for some sufficiently small $\delta_* > 0$. Here, $\{\widetilde{\psi}_j\}_{j\geq 1}$ forms an orthonormal basis in $L^2(X)$ and the dimension parameter J_* satisfies the inequality restriction

$$\frac{1}{n}\left(\sum_{j=1}^{J_*}\nu_j^{-4}j^{4p/d_x}\right)^{1/2}\leq C_{\mathcal{H}}^2. \tag{A.2}$$

Therefore, orthonormality of the basis functions $\{\widetilde{\psi}_j\}_{j\geq1}$ in $L^2(X)$ together with the Cauchy-Schwarz inequality implies for any $\theta\in\{\pm1\}^J$:

$$\sum_{j=1}^{\infty}\langle h_\theta,\widetilde{\psi}_j\rangle_X^2 j^{2p/d_x}=\frac{\delta_*^2}{n}\sum_{j=1}^{J_*}\nu_j^{-4}j^{2p/d_x}\left(\sum_{l=1}^{J_*}\nu_l^{-4}\right)^{-1/2}\leq\frac{\delta_*^2}{n}\left(\sum_{j=1}^{J_*}\nu_j^{-4}j^{4p/d_x}\right)^{1/2}\leq C_{\mathcal{H}}^2$$

for all $\delta_*\leq1$ and thus, we conclude that $h_\theta\in\mathcal{H}$ by the definition of $\mathcal{H}$. For any $\theta\in\{\pm1\}^{J_*}$ we have

$$\|h_\theta-\mathcal{H}_0\|_{L^2(X)}=\|h_\theta\|_{L^2(X)}=\frac{\delta_*}{\sqrt{n}}\left(\sum_{j=1}^{J_*}\nu_j^{-4}\right)^{1/4}=\delta_* r_n \tag{A.3}$$

and hence, $h_\theta\in\mathcal{H}_1(\delta_* r_n)$.

Let P^* denote the probability distribution obtained of the NPIR model by assigning the uniform distribution on $\{\pm1\}^{J_*}$ and P_0 the probability distribution when $h_\theta=0$. From the proof of Collier et al. [2017, Lemma 3] we infer the following reduction to testing between two probability measures under a simple null hypothesis. Using that $h_\theta\in\mathcal{H}_1(\delta_* r_n)$ for all $\theta\in\{\pm1\}^{J_*}$, we thus evaluate

$$\begin{aligned}\inf_{\mathtt{T}_n}\Big\{\sup_{h\in\mathcal{H}_0}\mathrm{P}_h(\mathtt{T}_n=1)+\sup_{h\in\mathcal{H}_1(\delta_* r_n)}\mathrm{P}_h(\mathtt{T}_n=0)\Big\}&\geq\inf_{\mathtt{T}_n}\big\{\mathrm{P}_0(\mathtt{T}_n=1)+\sup_{\theta\in\{\pm1\}^J}\mathrm{P}_\theta(\mathtt{T}_n=0)\big\}\\&\geq\inf_{\mathtt{T}_n}\big\{\mathrm{P}_0(\mathtt{T}_n=1)+\mathrm{P}^*(\mathtt{T}_n=0)\big\}\geq1-\mathcal{V}(\mathrm{P}^*,\mathrm{P}_0)\geq1-\sqrt{\chi^2(\mathrm{P}^*,\mathrm{P}_0)},\end{aligned} \tag{A.4}$$

where $\mathcal{V}(\cdot,\cdot)$ denotes the total variation distance and $\chi^2(\cdot,\cdot)$ denotes the χ^2 divergence.

We can write that $y_k=\gamma_k\theta_k+\frac{\sigma}{\sqrt{n}}\xi_k$ where $\gamma_k:=\delta_* n^{-1/2}\langle T\phi_{J_*},\widetilde{b}_k\rangle_{\mathcal{W}}$ and $\phi_J(\cdot):=\big(\sum_{j=1}^J\nu_j^{-4}\big)^{-1/4}\sum_{j=1}^J\nu_j^{-2}\widetilde{\psi}_j(\cdot)$. Consequently, by the derivation of equation (2.106) in Tsybakov [2009], the χ^2 divergence between P^* and P_0 satisfies

$$\chi^2(\mathrm{P}^*,\mathrm{P}_0)=\int\left(\frac{d\mathrm{P}^*}{d\mathrm{P}_0}\right)^2d\mathrm{P}_0-1=\prod_{k=1}^{J_*}\frac{\exp(-n\gamma_k^2/\sigma^2)+\exp(n\gamma_k^2/\sigma^2)}{2}-1.$$

By Tsybakov [2009, Section 2.7.5] there exists a constant $c_1>0$ such that $\exp(-n\gamma_k^2/\sigma^2)+\exp(n\gamma_k^2/\sigma^2)\leq2\exp\left(c_1n^2\gamma_k^4\right)$. Assumption 1(iv) implies for some constant $c>0$ that

$\sum_{k=1}^{J_*}\langle T\phi_{J_*},\widetilde{b}_k\rangle_{\mathcal{W}}^4 \leq c\sum_{j\geq 1}\nu_j^4\langle \phi_{J_*},\widetilde{\psi}_j\rangle_X^4 = c$ and we thus obtain by the definition of γ_k:

$$\chi^2(\mathrm{P}^*,\mathrm{P}_0) \leq \exp\Big(c_1 n^2 \sum_{k=1}^{J_*}\gamma_k^4\Big) - 1 \leq \exp\Big(\delta_*^4 c_1 c c_{\mathcal{X}}^{-2}\Big) - 1 \leq 1-\alpha,$$

for $\delta_* = \delta_*(\alpha) > 0$ sufficiently small. Consequently, the result follows by making use of inequality (A.4).

In the regularly varying case ($\nu_{J_*}^{-4}J_* \lesssim \sum_{j=1}^{J_*}\nu_j^{-4}$) for $J_* \sim \max\big\{J : n^{-1/2}J^{1/4}\nu_J^{-1} \leq J^{-p/d_x}\big\}$, we note that inequality(A.2) holds within a constant and we have $r_n = n^{-1/2}\big(\sum_{j=1}^{J_*}\nu_j^{-4}\big)^{1/4} \sim n^{-1/2}J_*^{1/4}\nu_{J_*}^{-1} \sim J_*^{-p/d_x}$. Consider the mildly ill-posed case ($\nu_j = j^{-a/d_x}$). The choice of $J_* \sim n^{2d_x/(4(p+a)+d_x)}$ ensures constraint (A.2) within a constant and implies $r_n \sim n^{-2p/(4(p+a)+d_x)}$. Consider the severely ill-posed case ($\nu_j = \exp(-j^{a/d_x}/2)$). The choice of $J_* = (c\log n)^{d_x/a}$ satisfies (A.2) within a constant and implies $r_n \sim (\log n)^{-p/\zeta}$, which completes the proof for the simple null case.

We now turn to the lower bound for testing composite null hypotheses. Consider the test function given in equation (A.1). Using that $\mathcal{H}_0$ is a nonempty, closed and convex, strict subset of $\mathcal{H}$ there exists a unique element $\Pi_{\mathcal{H}_0}h \in \mathcal{H}_0$ (by the Hilbert projection theorem) such that

$$\|h_\theta - \mathcal{H}_0\|_{L^2(X)} = \|h_\theta - \Pi_{\mathcal{H}_0}h_\theta\|_{L^2(X)} \geq \|h_{\theta_*} - \Pi_{\mathcal{H}_0}h_{\theta_*}\|_{L^2(X)} \tag{A.5}$$

for some $\theta_* \in \{\pm 1\}^{J_*}$. As above, we may assume $\Pi_{\mathcal{H}_0}h_{\theta_*} = 0$ without loss of generality (otherwise, consider $\widetilde{Y} = Y - T\Pi_{\mathcal{H}_0}h_{\theta_*}$ in the reduced-form NPIR model). Given the inequality (A.5), we thus conclude $\|h_\theta - \mathcal{H}_0\|_{L^2(X)} \geq \|h_{\theta_*}\|_{L^2(X)} \geq \delta_* r_n$, by following inequality (A.3). Therefore, we may proceed with the proof of the lower bound as in the simple null case. □

Proof of Theorem 3.2. First, by Lemma A.1 we control the type I error of the test $\mathtt{T}_{n,J}$ given in (3.5): $\limsup_{n\to\infty}\mathrm{P}_{h_0}(\mathtt{T}_{n,J}=1) = \limsup_{n\to\infty}\mathrm{P}_{h_0}\Big(n\widehat{D}_J(h_0) > \eta_J(\alpha)\widehat{\mathrm{v}}_J\Big) \leq \alpha$. To control the type II error, we have uniformly for $h \in \mathcal{H}_1(\delta^\circ r_{n,J})$,

$$\begin{aligned}\mathrm{P}_h\left(\mathtt{T}_{n,J}=0\right) &\leq \mathrm{P}_h\Big(n\widehat{D}_J(h_0) \leq \eta_J(\alpha)\widehat{\mathrm{v}}_J,\ \ \widehat{\mathrm{v}}_J \leq (1+c_0)\mathrm{v}_J\Big) + \mathrm{P}_h\left(\widehat{\mathrm{v}}_J > (1+c_0)\mathrm{v}_J\right)\\ &\leq \mathrm{P}_h\Big(n\widehat{D}_J(h_0) \leq (1+c_0)\eta_J(\alpha)\mathrm{v}_J\Big) + o(1) = o(1),\end{aligned}$$

where the second equation is due to Lemma B.4(i) and the last equation is due to Lemma B.7(i) in Appendix B. Finally recall that Assumption 3 implies $\nu_j^{-2} \geq cs_J^{-2}$. Given the definition of J_{*0}, the rate results for the mildly ill-posed case ($\nu_j = j^{-a/d_x}$) and for the severely ill-posed case ($\nu_j = \exp(-j^{a/d_x}/2)$) follow from $r_{n,J_{*0}} = (J_{*0})^{-p/d_x}$ directly. □

Lemma A.1. *Let Assumptions 1(i)-(iii) and 2 be satisfied. Then, under the simple hypothesis $\mathcal{H}_0 = \{h_0\}$ for a known function h_0, we have* $\mathrm{P}_{h_0}\left(\frac{n\widehat{D}_J(h_0)}{\widehat{\mathrm{v}}_J} > \eta_J(\alpha)\right) = \alpha + o(1)$.

A proof of Lemma A.1 is given in online Appendix E.

B. Proofs of Adaptive Testing Results in Section 4

We first introduce additional notation. For a $r \times c$ matrix M with $r \leq c$ and full row rank r we let M_l^- denote its left pseudoinverse, namely $(M'M)^- M'$. The $J \times K$ matrices $\widehat{A}$ and A defined in Sections 2.2 can be written as $\widehat{A} = (\widehat{G}_b^{-1/2}\widehat{S}\widehat{G}^{-1/2})_l^- \widehat{G}_b^{-1/2}$ and $A = (G_b^{-1/2}SG^{-1/2})_l^- G_b^{-1/2}$. For any $J \geq 1$ let $s_J = s_{\min}(G_b^{-1/2}SG^{-1/2}) > 0$ denote the J-th smallest singular value of $G_b^{-1/2}SG^{-1/2}$. Then $\|AG_b^{1/2}\| = \|(G_b^{-1/2}SG^{-1/2})_l^-\| = s_J^{-1}$. Let $\widetilde{b}^K(\cdot) = G_b^{-1/2}b^K(\cdot)$ and $\widetilde{\psi}^J(\cdot) = G^{-1/2}\psi^J(\cdot)$. For any $h \in L^2(X)$, its population 2SLS projection onto the sieve space Ψ_J is:

$$Q_J h(\cdot) = \widetilde{\psi}^J(\cdot)' A\, \mathrm{E}[b^K(W)h(X)] = \widetilde{\psi}^J(\cdot)'(G_b^{-1/2}SG^{-1/2})_l^- \,\mathrm{E}[\widetilde{b}^K(W)h(X)]\ . \tag{B.1}$$

We next present nine results that are used to establish our adaptive testing upper bounds. The proofs of these results are postponed to the online Appendix E. Below, we shorten "with probability P_h approaching one uniformly for $h \in \mathcal{H}$" to "wpa1 uniformly for $h \in \mathcal{H}$".

Theorem B.1. *Let Assumptions 1(ii)-(iii) and 2 hold. Then, wpa1 uniformly for $h \in \mathcal{H}$:*

$$\widehat{D}_J(\Pi_{\mathcal{H}_0}h) - \|Q_J(h - \Pi_{\mathcal{H}_0}h)\|^2_{L^2(X)} \lesssim n^{-1}s_J^{-2}\sqrt{J} + n^{-1/2}s_J^{-1}\big(\|h - \Pi_{\mathcal{H}_0}h\|_{L^2(X)} + J^{-p/d_x}\big).$$

Theorem B.1 provides an upper bound for quadratic distance estimation, which is essential for our upper bound on the minimax rate of testing in L^2.

Lemma B.1. *Let Assumption 2(iv) hold. Then we have uniformly for $h \in \mathcal{H}$: (i) $\|Q_J(h - \Pi_{\mathcal{H}_0}h)\|_{L^2(X)} = \|h - \Pi_{\mathcal{H}_0}h\|_{L^2(X)} + O(J^{-p/d_x})$ and (ii) $\|Q_J h - h\|_{L^2(X)} = O(J^{-p/d_x})$.*

Lemma B.2. *Let Assumption 2(i) hold. Then: $\mathrm{v}_J \leq \overline{\sigma}^2 s_J^{-2}\sqrt{J}$ uniformly for $h \in \mathcal{H}$ and $J \in \mathcal{I}_n$.*

Lemma B.3. *Let Assumption 1(i) hold. Then: $J \leq \sum_{j=1}^J s_j^{-4} \leq \underline{\sigma}^{-4}\mathrm{v}_J^2$ uniformly for $h \in \mathcal{H}$ and $J \in \mathcal{I}_n$.*

Lemma B.4. *Let Assumption 1(i)-(iii) be satisfied.*
(i) If in addition Assumption 2 holds, then for any $c > 0$ we have

$$\sup_{h \in \mathcal{H}} \mathrm{P}_h\left(\left|1 - \frac{\widehat{\mathrm{v}}_J}{\mathrm{v}_J}\right| > c\right) = o(1).$$

(ii) If in addition Assumptions 2(i) and 4(i) hold, then for any $c > 0$ we have

$$\sup_{h\in\mathcal{H}} \mathrm{P}_h\left(\max_{J\in\mathcal{I}_n}\left|1-\frac{\widehat{\mathrm{v}}_J}{\mathrm{v}_J}\right| > c\right) = o(1).$$

Lemma B.5. *For all $\alpha \in (0,1)$ and $J \in \widehat{\mathcal{I}}_n$ we have for n sufficiently large and almost surely that*

$$\frac{\sqrt{\log\log(J)-\log(\alpha)}}{4} \leq \widehat{\eta}_J(\alpha) \leq 4\sqrt{\log\log(n)-\log(\alpha)}.$$

For any $h \in \mathcal{H}$ let $V_i^J := Ab^K(W_i)[Y_i - \Pi_{\mathcal{H}_0}h(X_i)]$ with V_{ij} as its jth entry, $1 \leq j \leq J$. Then $Q_J(h - \Pi_{\mathcal{H}_0}h) = \mathrm{E}_h[V^J]'\widetilde{\psi}^J$ and $\|\mathrm{E}_h[V^J]\|^2 = \|Q_J(h-\Pi_{\mathcal{H}_0}h)\|^2_{L^2(X)}$ for any NPIV function $h \in \mathcal{H}$. Let $Z_i = (Y_i, X_i, W_i)$. For any set D_i we define

$$R(Z_i, Z_{i'}, D_i) := (V_i^J \mathbb{1}_{D_i})'(V_{i'}^J \mathbb{1}_{D_{i'}}) - \mathrm{E}_h(V_i^J \mathbb{1}_{D_i})'\,\mathrm{E}_h(V_i^J \mathbb{1}_{D_i})\ ,$$

$R_1(Z_i, Z_{i'}) := R(Z_i, Z_{i'}, M_i)$ and $R_2(Z_i, Z_{i'}) := R(Z_i, Z_{i'}, M_i^c)$ where $M_i = \{|Y_i - \Pi_{\mathcal{H}_0}h(X_i)| \leq M_n\}$ and $M_n = \sqrt{n}\,\zeta_{\overline{J}}^{-1}(\log\log\overline{J})^{-3/4}$. Let

$$\Lambda_1 := \Big(\frac{n(n-1)}{2}\,\mathrm{E}[R_1^2(Z_1,Z_2)]\Big)^{1/2}, \qquad \Lambda_2 := n \sup_{\|\nu\|_{L^2(Z)}\leq 1, \|\kappa\|_{L^2(Z)}\leq 1} \mathrm{E}[R_1(Z_1,Z_2)\nu(Z_1)\kappa(Z_2)],$$

$$\Lambda_3 := \big(n \sup_z |\,\mathrm{E}[R_1^2(Z_1,z)]|\big)^{1/2}, \qquad \text{and} \qquad \Lambda_4 := \sup_{z_1,z_2}|R_1(z_1,z_2)|.$$

Lemma B.6. *(i) There exists a generic constant $C_{R_1} > 0$, such that for all $u > 0$ and $n \in \mathbb{N}$ we have:*

$$\mathrm{P}_h\left(\Big|\sum_{1\leq i<i'\leq n} R_1(Z_i, Z_{i'})\Big| \geq C_{R_1}\Big(\Lambda_1\sqrt{u} + \Lambda_2 u + \Lambda_3 u^{3/2} + \Lambda_4 u^2\Big)\right) \leq 6\exp(-u).$$

(ii) Let Assumption 2(i) hold. Then for the kernel R_1 the following holds under $\mathcal{H}_0$:

$$\Lambda_1 \leq \sqrt{n(n-1)/2}\,\mathrm{v}_J, \quad \Lambda_2 \leq \overline{\sigma}^2\, n\, s_J^{-2}\ , \ \Lambda_3 \leq \overline{\sigma}^2\sqrt{n}\, M_n\, \zeta_{b,K}\, s_J^{-2}, \quad \Lambda_4 \leq M_n^2 \zeta_{b,K}^2 s_J^{-2}\ .$$

Lemma B.7. *(i) Under the conditions of Theorem 3.2 we have for some constant $c_0 > 0$ that $\mathrm{P}_h\big(n\widehat{D}_J(h_0) \leq (1+c_0)\eta_J(\alpha)\mathrm{v}_J\big) = o(1)$ uniformly for $h \in \mathcal{H}_1(\delta^\circ r_{n,J})$.*
(ii) Under the conditions of Theorem 4.1 we have $\mathrm{P}_h\big(n\widehat{D}_{J^*}(h_0) \leq 2c_1\sqrt{\log\log n}\,\mathrm{v}_{J^*}\big) = o(1)$ uniformly for $h \in \mathcal{H}_1(\delta^\circ r_n)$, where J^* and c_1 are given in the proof of Theorem 4.1.

Lemma B.8. *Let Assumption 4(i)(iii) be satisfied. Then $\widehat{J}_{\max}$ given in (2.12) satisfies (i) $\sup_{h\in\mathcal{H}} \mathrm{P}_h\left(\widehat{J}_{\max} > \overline{J}\right) = o(1)$; and (ii) $\sup_{h\in\mathcal{H}} \mathrm{P}_h\left(2J^\circ > \widehat{J}_{\max}\right) = o(1)$.*

Proof of Theorem 4.1. We prove this result in three steps. First, we bound the type

I error of the test statistic $\widetilde{\mathtt{T}}_n = \mathbb{1}\big\{ \max_{J\in\mathcal{I}_n} \big(n\widehat{D}_J(h_0)/(\eta'_J(\alpha)\mathrm{v}_J)\big) > 1\big\}$, $\eta'_J(\alpha) := (1-c_0)\sqrt{\log\log J - \log\alpha}/4$ for some constant $0 < c_0 < 1$. Second, we bound the type II error of $\widetilde{\mathtt{T}}_n$ where $\eta'_J(\alpha)$ is replaced by $\eta''(\alpha) := 4(1+c_0)\sqrt{\log\log n - \log\alpha}$. Third, we show that the derived bounds in Steps 1 and 2 are sufficient to control the type I and type II errors of our adaptive test $\widehat{\mathtt{T}}_n$ for a simple null hypothesis $\mathcal{H}_0 = \{h_0\}$.

Step 1: To control the type I error of $\widetilde{\mathtt{T}}_n$, we use a decomposition under $\mathcal{H}_0 = \{h_0\}$ via the U-statistic $\mathcal{U}_{J,l} = \frac{2}{n(n-1)}\sum_{1\le i<i'\le n} R_l(Z_i, Z_{i'})$ for $l = 1, 2$ and $U_i = Y_i - h_0(X_i)$:

$$\begin{aligned}\mathrm{P}_{h_0}\big(\widetilde{\mathtt{T}}_n = 1\big) \le \mathrm{P}_{h_0}\Big(&\max_{J\in\mathcal{I}_n}\Big|\frac{1}{\eta'_J(\alpha)\mathrm{v}_J(n-1)}\sum_{j=1}^{J}\sum_{i\neq i'} V_{ij}V_{i'j}\Big| \\ &+ \max_{J\in\mathcal{I}_n}\Big|\frac{1}{\eta'_J(\alpha)\mathrm{v}_J(n-1)}\sum_{i\neq i'} U_iU_{i'}b^K(W_i)'\Big(A'A - \widehat{A}'\widehat{A}\Big)b^K(W_{i'})\Big| > 1\Big) \le I + II + III,\end{aligned}$$

with $I := \mathrm{P}_{h_0}\Big(\max_{J\in\mathcal{I}_n}\big|n\mathcal{U}_{J,1}/(\eta'_J(\alpha)\mathrm{v}_J)\big| > \frac{1}{4}\Big)$, $II := \mathrm{P}_{h_0}\big(\max_{J\in\mathcal{I}_n}\big|n\mathcal{U}_{J,2}/(\eta'_J(\alpha)\mathrm{v}_J)\big| > \frac{1}{4}\big)$,

$$III := \mathrm{P}_{h_0}\Big(\max_{J\in\mathcal{I}_n}\Big|\frac{1}{\eta'_J(\alpha)\mathrm{v}_J(n-1)}\sum_{i\neq i'} U_iU_{i'}b^K(W_i)'\big(A'A - \widehat{A}'\widehat{A}\big)b^K(W_{i'})\Big| > \frac{1}{2}\Big)\,.$$

First we consider term III. Using the definition of $\eta'_J(\alpha)$ and the fact that $\sqrt{\log\log J - \log\alpha} > \sqrt{\log\log J}$ for any $\alpha \in (0,1)$, we obtain $III = o(1)$ by applying Lemma E.6.
Next we consider term I. Define $\Lambda(u, J) := \Lambda_1\sqrt{u} + \Lambda_2 u + \Lambda_3 u^{3/2} + \Lambda_4 u^2$. By Lemma B.6(ii) with $M_n = \sqrt{n}\,\zeta_{\overline{J}}^{-1}(\log\log\overline{J})^{-3/4}$ we have for all $J \in \mathcal{I}_n$,

$$\Lambda(u, J) \le n\mathrm{v}_J\sqrt{u/2} + \overline{\sigma}^2 n s_J^{-2} u + \overline{\sigma}^2 n s_J^{-2}\left(\log\log\overline{J}\right)^{-3/4}u^{3/2} + n s_J^{-2}\left(\log\log\overline{J}\right)^{-3/2}u^2$$

for n sufficiently large. Replacing in the previous inequality u by $u_J = 2\log\log J^{c_\alpha}$ where $c_\alpha = \sqrt{1 + (\pi/\log 2)^2}/\sqrt{\alpha}$, we obtain for n sufficiently large:

$$\begin{aligned}\Lambda(u_J, J) &\le n\mathrm{v}_J\sqrt{\log\log J^{c_\alpha}} + \frac{2\overline{\sigma}^2 n}{s_J^2}\log\log J^{c_\alpha} + \frac{\overline{\sigma}^2 n}{s_J^2}\left(2\log\log J^{c_\alpha}\right)^{3/4} + \frac{4n}{s_J^2}\sqrt{\log\log J^{c_\alpha}} \\ &\le \frac{5}{4}n\mathrm{v}_J\sqrt{\log\log J - \log\alpha} + 3\overline{\sigma}^2 n s_J^{-2}(\log\log J - \log\alpha) \\ &\le \frac{5}{1-c_0}n\mathrm{v}_J\eta'_J(\alpha) + \frac{12\overline{\sigma}^2}{1-c_0}n s_J^{-2}\eta'_J(\alpha)\sqrt{\log\log J},\end{aligned}$$

by the definition of $\eta'_J(\alpha)$. Given $s_J^{-2} \le \underline{\sigma}^{-2}\mathrm{v}_J$ (by Lemma B.3) and Assumption 4(ii), we have for all $J \in \mathcal{I}_n$ and for n sufficiently large: $\Lambda(u_J, L(J)) \le C_{R_1}\frac{n-1}{8}\mathrm{v}_J\eta'_J(\alpha)$ with $L(J) = \exp(1/6)J\underline{J}^{-1/2}$. By Lemma B.6(i) with $u = 2\log\log J^{c_\alpha}$ and the fact that $J = \underline{J}2^j$

for all $J \in \mathcal{I}_n$, we obtain for n sufficiently large:

$$I \leq \sum_{J \in \mathcal{I}_n} \mathrm{P}_{h_0}\left(\left|n\,\mathcal{U}_{J,1}\right| > \frac{\eta'_J(\alpha)}{4}\,\mathrm{v}_J\right) = \sum_{J \in \mathcal{I}_n} \mathrm{P}_{h_0}\left(\Big|\sum_{i<i'} R_1(Z_i, Z_{i'})\Big| \geq \frac{\eta'_J(\alpha)}{4}\frac{n-1}{2}\mathrm{v}_J\right)$$
$$\leq \sum_{J \in \mathcal{I}_n} \mathrm{P}_{h_0}\left(\Big|\sum_{i<i'} R_1(Z_i, Z_{i'})\Big| \geq C_{R_1}\Lambda(u_J, L(J))\right) \leq 6 \sum_{J \in \mathcal{I}_n} \exp\big(-2\log\log(L(J)^{c_\alpha})\big).$$

Using the fact that $\sum_{j \geq 1} j^{-2} = \pi^2/6$, we obtain:

$$I \leq 6\, c_\alpha^{-2} \sum_{J \in \mathcal{I}_n} \big(\log L(J)\big)^{-2} \leq \alpha\, \frac{6}{1 + (\pi/\log 2)^2} \sum_{j \geq 0} (1/6 + j \log 2)^{-2}$$
$$\leq \alpha\, \frac{6}{1 + (\pi/\log 2)^2}\Big(1/6 + (\log 2)^{-2} \sum_{j \geq 1} j^{-2}\Big) = \alpha\ .$$

Consider term II. Since $\mathrm{E}_{h_0}|U \mathbb{1}_{\{|U|>M_n\}}| \leq M_n^{-3}\,\mathrm{E}_{h_0}[U^4 \mathbb{1}_{\{|U|>M_n\}}] \leq M_n^{-3}\,\mathrm{E}_{h_0}[U^4]$, Markov's inequality yields

$$II \leq \mathrm{E}_{h_0} \max_{J \in \mathcal{I}_n}\left|\frac{4}{\eta'_J(\alpha)\mathrm{v}_J(n-1)} \sum_{i<i'} U_i \mathbb{1}_{M_i^c} U_{i'} \mathbb{1}_{M_{i'}^c} b^K(W_i)' A' A b^K(W_{i'})\right|$$
$$\leq 4n\,\mathrm{E}_{h_0}|U \mathbb{1}_{\{|U|>M_n\}}|\ \mathrm{E}_{h_0}|U \mathbb{1}_{\{|U|>M_n\}}| \max_{J \in \mathcal{I}_n} \frac{\zeta_J^2 \big\|(G_b^{-1/2} S G^{-1/2})_l^- \big\|^2}{\eta'_J(\alpha)\mathrm{v}_J}$$
$$\leq 4n M_n^{-6}\big(\mathrm{E}_{h_0}[U^4]\big)^2 \zeta_{\overline{J}}^2 \max_{J \in \mathcal{I}_n} \frac{s_J^{-2}}{\eta'_J(\alpha)\mathrm{v}_J},$$

where the fourth moment of $U = Y - h_0(X)$ is bounded under Assumption 2(i). Lemma B.3 implies $s_J^{-2} \leq \underline{\sigma}^{-2}\mathrm{v}_J$. By definition $M_n = \sqrt{n}\,\zeta_{\overline{J}}^{-1}(\log\log\overline{J})^{-3/4}$ and Assumption 4(i), we obtain $II = o\left(n^{-2}(\log\log\overline{J})^{9/2}\zeta_{\overline{J}}^8\right) = o(1)$.

Step 2: We control the type II error of the test statistic $\widetilde{\mathtt{T}}_n$ where $\eta'_J(\alpha)$ is replaced by $\eta''(\alpha) > 0$. From the definition $\overline{J} = \sup\{J : s_J^{-1}\zeta_J^2\sqrt{(\log J)/n} \leq \overline{c}\}$ we infer that the dimension parameter J° given in (4.3) satisfies $\underline{J} \leq J^\circ \leq \overline{J}/2$ for $\overline{c}$ sufficiently large by Assumption 4(iii). Thus, by the construction of the set $\mathcal{I}_n$ there exists $J^* \in \mathcal{I}_n$ such that $J^\circ \leq J^* < 2J^\circ$. Let $K^* = K(J^*)$. We note that for all $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$

$$\mathrm{P}_h\big(\widetilde{\mathtt{T}}_n = 0\big) = \mathrm{P}_h\Big(n\,\widehat{D}_J(h_0) \leq \eta''(\alpha)\,\mathrm{v}_J \text{ for all } J \in \mathcal{I}_n\Big)$$
$$\leq \mathrm{P}_h\Big(n\,\widehat{D}_{J^*}(h_0) \leq c_1\sqrt{\log\log n - \log\alpha}\,\mathrm{v}_{J^*}\Big)$$

with $c_1 = 4(1 + c_0)$, by the definition of $\eta''(\alpha)$. Note that $\log\log n - \log\alpha = (\log\log n)[1 - (\log\alpha)/(\log\log n)] \leq 2\log\log n$ for all n sufficiently large. Consequently, we may apply Lemma B.7(ii) which implies $\mathrm{P}_h\big(\widetilde{\mathtt{T}}_n = 0\big) = o(1)$ uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$.

Step 3: Finally, we account for estimation of the normalization factor v_J and for estimation of upper bound of the RES index $\widehat{\mathcal{I}}_n$. We control the type I error of the test $\widehat{\mathtt{T}}_n$ under simple null hypotheses as follows. The lower bound in Lemma B.5 implies

$$\begin{aligned}
\mathrm{P}_{h_0}\big(\widehat{\mathtt{T}}_n = 1\big) &\le \mathrm{P}_{h_0}\left(\max_{J\in\widehat{\mathcal{I}}_n}\Big\{n\widehat{D}_J(h_0)/(\eta'_J(\alpha)\widehat{\mathrm{v}}_J)\Big\} > (1-c_0)^{-1}\right)\\
&\le \mathrm{P}_{h_0}\left(\max_{J\in\mathcal{I}_n}\Big\{n\widehat{D}_J(h_0)/(\eta'_J(\alpha)\widehat{\mathrm{v}}_J)\Big\} > (1-c_0)^{-1},\quad \widehat{\mathrm{v}}_J \ge (1-c_0)\mathrm{v}_J \text{ for all } J\in\mathcal{I}_n\right)\\
&\qquad + \mathrm{P}_{h_0}\left(\widehat{\mathrm{v}}_J < (1-c_0)\mathrm{v}_J \text{ for all } J\in\mathcal{I}_n\right) + \mathrm{P}_{h_0}\big(\widehat{J}_{\max} > \overline{J}\big)\\
&\le \mathrm{P}_{h_0}\left(\max_{J\in\mathcal{I}_n}\Big\{n\widehat{D}_J(h_0)/(\eta'_J(\alpha)\mathrm{v}_J)\Big\} > 1\right) + \mathrm{P}_{h_0}\left(\max_{J\in\mathcal{I}_n}|\widehat{\mathrm{v}}_J/\mathrm{v}_J - 1| > c_0\right) + o(1) \le \alpha + o(1)
\end{aligned}$$

where the third inequality is due to Lemmas B.8(i) and B.4(ii), and the last inequality is due to Step 1 of this proof. To bound the type II error of the test $\widehat{\mathtt{T}}_n$ recall the definition of $J^* \in \mathcal{I}_n$ given in Step 2 of this proof. Using the upper bound of Lemma B.5 together with Lemmas B.8(ii) and B.4 we evaluate uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:

$$\begin{aligned}
\mathrm{P}_h\big(\widehat{\mathtt{T}}_n = 0\big) &\le \mathrm{P}_h\left(n\widehat{D}_{J^*}(h_0) \le (1+c_0)^{-1}\eta''(\alpha)\,\widehat{\mathrm{v}}_{J^*}\right) + \mathrm{P}_h\big(J^* > \widehat{J}_{\max}\big)\\
&\le \mathrm{P}_h\left(n\widehat{D}_{J^*}(h_0) \le (1+c_0)^{-1}\eta''(\alpha)\,\widehat{\mathrm{v}}_{J^*},\quad \widehat{\mathrm{v}}_{J^*} \le (1+c_0)\mathrm{v}_{J^*}\right) + \mathrm{P}_h\left(\widehat{\mathrm{v}}_{J^*} > (1+c_0)\mathrm{v}_{J^*}\right) + o(1)\\
&\le \mathrm{P}_h\left(n\widehat{D}_{J^*}(h_0) \le \eta''(\alpha)\mathrm{v}_{J^*}\right) + o(1) = o(1),
\end{aligned}$$

where the last equation is due to Step 2 of this proof.

We show that Assumption 4(ii) is satisfied if $\{s_j\}$ is regularly varying. From Lemmas B.2 and B.3 we infer $\mathrm{v}_L \le \overline{\sigma}^2 s_L^{-2}\sqrt{L}$ and $\mathrm{v}_J \ge \underline{\sigma}^2\big(\sum_{j=1}^J s_j^{-4}\big)^{1/2}$ uniformly for $h \in \mathcal{H}$ and $J \in \mathcal{I}_n$. Then

$$\frac{\mathrm{v}_L}{\mathrm{v}_J} \lesssim \left(\frac{L s_L^{-4}}{\sum_{j=1}^J s_j^{-4}}\right)^{1/2} \lesssim \left(\frac{L s_L^{-4}}{J s_J^{-4}}\right)^{1/2} = o(1)$$

for all $L = o(J)$. Recall that Assumption 1(iv) and Assumption 3 imply that $\nu_j^{-2} \ge c s_j^{-2}$ and $\nu_j^{-2} \le C s_j^{-2}$ for $0 < c < C$. Thus, we obtain regularly varying $\{\nu_j\}$ implies regularly varying $\{s_j\}$, and hence $\mathrm{v}_L/\mathrm{v}_J = o(1)$ for all $L = o(J)$. Since both the mildly ill-posed and severely ill-posed are special cases of regularly varying, the rest of the results follows. In the mildly ill-posed case, we obtain $J^\circ \sim \big(n/\sqrt{\log\log n}\big)^{2d_x/(4(p+a)+d_x)}$ which implies $\mathsf{r}_n \sim \big(\sqrt{\log\log n}/n\big)^{2p/(4(p+a)+d_x)}$. In the severely ill-posed case, note that if $J^\circ \sim \big(c\log n\big)^{d_x/a}$ for some constant $c \in (0,1)$ then we obtain $n^{-1}\sqrt{\log\log n}\, s_{J^\circ}^{-2}\sqrt{J^\circ} \lesssim (J^\circ)^{-2p/d_x} \sim \big(\log n\big)^{-2p/d_x}$.
□

Proof of Theorem 4.2. We prove this result in three steps. First, we bound the type I error of the test statistic $\widetilde{\mathtt{T}}_n = \mathbb{1}\big\{\max_{J\in\mathcal{I}_n}\big\{n\widehat{D}_J/(\eta'_J(\alpha)\mathrm{v}_J)\big\} > 1\big\}$, where $\eta'_J(\alpha)$ is given in the proof of Theorem 4.1. Second, we bound the type II error of $\widetilde{\mathtt{T}}_n$ where $\eta'_J(\alpha)$ is replaced by $\eta''(\alpha)$ given in the proof of Theorem 4.1. Third, we show that Steps 1 and 2 are sufficient to control the type I and type II errors of our adaptive test $\widehat{\mathtt{T}}_n$ for the composite null.

Step 1: We control the type I error of the test statistic $\widetilde{\mathtt{T}}_n$ using the decomposition

$$n(n-1)\,\widehat{D}_J = \sum_{i\neq i'} \big(Y_i - \widehat{h}^{\mathrm{R}}_J(X_i)\big)\big(Y_{i'} - \widehat{h}^{\mathrm{R}}_J(X_{i'})\big) b^K(W_i)'\widehat{A}'\widehat{A} b^K(W_{i'})$$

$$= \Big\| \sum_i \big(Y_i - \widehat{h}^{\mathrm{R}}_J(X_i)\big)\widehat{A} b^K(W_i)\Big\|^2 - \sum_i \Big\|\big(Y_i - \widehat{h}^{\mathrm{R}}_J(X_i)\big)\widehat{A} b^K(W_i)\Big\|^2.$$

For any $h \in \mathcal{H}_0$ we define $h^*_J := \arg\min_{\phi\in\mathcal{H}_{0,J}} \|\sum_i (\phi - h)(X_i)\widehat{A} b^K(W_i)\|$. The definition of the restricted NPIV estimator $\widehat{h}^{\mathrm{R}}_J \in \mathcal{H}_{0,J}$ in (2.7) yields for all $h \in \mathcal{H}_0$:

$$\Big\| \sum_i \big(Y_i - \widehat{h}^{\mathrm{R}}_J(X_i)\big)\widehat{A} b^K(W_i)\Big\| \le \Big\| \sum_i \big(Y_i - h^*_J(X_i)\big)\widehat{A} b^K(W_i)\Big\|$$

$$\le \Big\| \sum_i \big(Y_i - h(X_i)\big)\widehat{A} b^K(W_i)\Big\| + \Big\| \sum_i (h - h^*_J)(X_i)\widehat{A} b^K(W_i)\Big\|.$$

By Lemma B.9, uniformly for $J \in \mathcal{I}_n$, we have

$$\frac{n\widehat{D}_J}{\eta'_J(\alpha)\mathrm{v}_J} - \frac{n\widehat{D}_J(h)}{\eta'_J(\alpha)\mathrm{v}_J} \lesssim \big(\mathrm{v}_J\sqrt{(\log\log J)/J}\big)^{-1/2} n^{-1}\sum_i \big(Y_i - h(X_i)\big) b^K(W_i)'\widehat{A}'\widehat{A} b^K(W_i)\big(\widehat{h}^{\mathrm{R}}_J - h\big)(X_i)$$

$$+ \big(\mathrm{v}_J\sqrt{(\log\log J)/J}\big)^{-1/2}\Big\| \frac{1}{\sqrt{n}}\sum_i \big(Y_i - h(X_i)\big)\widehat{A} b^K(W_i)\Big\|$$

$$=: \big(\mathrm{v}_J\sqrt{(\log\log J)/J}\big)^{-1/2}\big(T_{1,J} + 2T_{2,J}\big)$$

wpa1 uniformly for $h \in \mathcal{H}_0$, where $\widehat{D}_J(h)$ is given in (3.4) (with h_0 replaced by $h = \Pi_{\mathcal{H}_0} h$ under $\mathcal{H}_0$). Now we may follow step 1 of the proof of Theorem 4.1 and obtain

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_0} \mathrm{P}_h\left(\max_{J\in\mathcal{I}_n}\Big\{n\widehat{D}_J(h)/(\eta'_J(\alpha)\mathrm{v}_J)\Big\} > 1/4\right) \le \alpha.$$

It remains to control $T_{1,J}$ and $T_{2,J}$. Consider $T_{1,J}$. For all $J \in \mathcal{I}_n$ we evaluate

$$T_{1,J} = \frac{1}{n}\sum_i \big(Y_i - h(X_i)\big) b^K(W_i)' A' A b^K(W_i)\big(\widehat{h}^{\mathrm{R}}_J - h\big)(X_i)$$

$$+ \frac{1}{n}\sum_i \big(Y_i - h(X_i)\big) b^K(W_i)'\big(\widehat{A}'\widehat{A} - A'A\big) b^K(W_i)\big(\widehat{h}^{\mathrm{R}}_J - h\big)(X_i) := T_{11,J} + T_{12,J}.$$

Consider $T_{11,J}$. We first observe by the CauchySchwarz inequality that

$$T_{11,J} \le \Big(\frac{1}{n}\sum_i \big(Y_i - h(X_i)\big)^2 \|A b^K(W_i)\|^2\Big)^{1/2}\Big(\frac{1}{n}\sum_i \big\|A b^K(W_i)\big(\widehat{h}^{\mathrm{R}}_J - h\big)(X_i)\big\|^2\Big)^{1/2}.$$

Further, another application of the Cauchy-Schwarz inequality implies

$$\mathrm{E}_h \max_{J\in\mathcal{I}_n} \big\|\big(Y-h(X)\big)Ab^K(W)\|^2 \leq \max_{J\in\mathcal{I}_n} \sqrt{J}\big\|A\,\mathrm{E}_h\left[\big(Y-h(X)\big)^2\widetilde{b}^K(W)\widetilde{b}^K(W)'\right]A'\big\|_F = \max_{J\in\mathcal{I}_n}\big\{\sqrt{J}\mathrm{v}_J\big\},$$

using the definition of the normalization term v_J. Consequently, we evaluate

$$\max_{J\in\mathcal{I}_n} \frac{T_{11,J}}{\mathrm{v}_J\sqrt{\log\log J}} \lesssim \max_{J\in\mathcal{I}_n} \frac{\zeta_J\|\widehat{h}_J^{\mathrm{R}}-h\|_{L^2(X)}}{\sqrt{\log\log J}} \times \max_{J\in\mathcal{I}_n} \frac{\sqrt{\mathrm{E}_h\left[\|(Y-h(X))Ab^K(W)\|^2\right]}}{\zeta_J s_J \mathrm{v}_J}$$

wpa1 uniformly for $h\in\mathcal{H}_0$, where the right hand side tends to zero by the rate condition imposed in Assumption 5(i), i.e., $\mathrm{P}_h(\max_{J\in\mathcal{I}_n}\|\widehat{h}_J^{\mathrm{R}}-h\|_{L^2(X)}\zeta_J/(\log\log J)^{1/4} > \varepsilon) \to 0$ uniformly for $h\in\mathcal{H}_0$ for any $\varepsilon>0$. Similarly, $\max_{J\in\mathcal{I}_n} T_{12,J}/(\mathrm{v}_J\sqrt{\log\log J})$ vanishes wpa1 uniformly for $h\in\mathcal{H}_0$, using that

$$\mathrm{P}\left(\max_{J\in\mathcal{I}_n}\Big\{s_J^2\zeta_J^{-1}\sqrt{n/(\log J)}\big\|(\widehat{A}-A)G_b^{1/2}\big\|\Big\} > C\right)$$
$$= \mathrm{P}\left(\max_{J\in\mathcal{I}_n}\Big\{s_J^2\zeta_J^{-1}\sqrt{\frac{n}{\log J}}\Big\|(\widehat{G}_b^{-1/2}\widehat{S}\widehat{G}^{-1/2})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}SG^{-1/2})_l^-\Big\|\Big\} > C\right) = o(1),$$

by Lemma E.5(i). Consider $T_{2,J}$. We have

$$T_{2,J} \leq \Big\|\frac{1}{\sqrt{n}}\sum_i \big(Y_i-h(X_i)\big)Ab^K(W_i)\Big\| + \Big\|\frac{1}{\sqrt{n}}\sum_i \big(Y_i-h(X_i)\big)(\widehat{A}-A)b^K(W_i)\Big\| := T_{21,J}+T_{22,J}.$$

We have $\mathrm{E}_h\max_{J\in\mathcal{I}_n} T_{21,J} \leq \sqrt{\mathrm{E}_h\max_{J\in\mathcal{I}_n}\|(Y-h(X))Ab^{K(J)}(W)\|^2} \leq \max_{J\in\mathcal{I}_n}\{J^{1/4}\sqrt{\mathrm{v}_J}\}$ as derived above and conclude

$$\mathrm{E}_h \max_{J\in\mathcal{I}_n}\frac{T_{21,J}}{\big(\mathrm{v}_J\sqrt{J(\log\log J)}\big)^{1/2}} \lesssim \max_{J\in\mathcal{I}_n}\frac{J^{1/4}\sqrt{\mathrm{v}_J}}{\big(\mathrm{v}_J\sqrt{J(\log\log J)}\big)^{1/2}} = o(1)$$

uniformly for $h\in\mathcal{H}_0$. Concerning the second summand $T_{22,J}$, by another application of Lemma E.5, $\max_{J\in\mathcal{I}_n} T_{22,J}/(\mathrm{v}_J\sqrt{J(\log\log J)})$ vanishes wpa1 uniformly for $h\in\mathcal{H}_0$.

Step 2: We control the type II error of the test statistic $\widetilde{\mathtt{T}}_n$. Let J^* be as in the proof of Theorem 4.1. We evaluate for all $h\in\mathcal{H}_1(\delta^\circ\mathrm{r}_n)$ that

$$\mathrm{P}_h\left(\widetilde{\mathtt{T}}_n=0\right) = \mathrm{P}_h\big(n\,\widehat{D}_J \leq \eta''(\alpha)\,\mathrm{v}_J \text{ for all } J\in\mathcal{I}_n\big) \leq \mathrm{P}_h\big(n\,\widehat{D}_{J^*} \leq c_1\sqrt{\log\log n-\log\alpha}\,\mathrm{v}_{J^*}\big),$$

with $c_1 = 4(1+c_0)$, by definition $\eta''(\alpha)$. Let $\widehat{V}_i^J := (Y_i-\widehat{h}_J^{\mathrm{R}}(X_i))Ab^K(W_i)$ then

$$\|\,\mathrm{E}_h[\widehat{V}^{J^*}]\|^2 = \mathrm{E}_h[(Y-\widehat{h}_{J^*}^{\mathrm{R}}(X))b^{K^*}(W)']A'A\,\mathrm{E}_h[(Y-\widehat{h}_{J^*}^{\mathrm{R}}(X))b^{K^*}(W)] = \|Q_{J^*}(h-\widehat{h}_{J^*}^{\mathrm{R}})\|_{L^2(X)}^2.$$

The triangular inequality implies $\big|\|Q_{J^*}(h-\widehat{h}_{J^*}^{\mathrm{R}})\|_{L^2(X)}-\|h-\widehat{h}_{J^*}^{\mathrm{R}}\|_{L^2(X)}\big| \leq \sup_{\phi\in\mathcal{H}}\|Q_{J^*}\phi-\phi\|_{L^2(X)}$ uniformly for $h\in\mathcal{H}_1(\delta^\circ\mathrm{r}_n)$. Consequently, Lemma B.1(ii) together with the definition of J^* implies $\sup_{h\in\mathcal{H}_1(\delta^\circ \mathsf{r}_n)}(\|\operatorname{E}_h[\widehat{V}^{J^*}]\| - \|h-\widehat{h}^{\mathrm{R}}_{J^*}\|_{L^2(X)})^2 \leq C_B \mathsf{r}_n^2$ for some constant $C_B>0$. Using this bound, we derive

$$\begin{aligned}
\mathrm{P}_h\left(n\,\widehat{D}_{J^*} \leq 2c_1\sqrt{\log\log n}\,\mathrm{v}_{J^*}\right) &= \mathrm{P}_h\left(\|\operatorname{E}_h[\widehat{V}^{J^*}]\|^2 - \widehat{D}_{J^*} > \|\operatorname{E}_h[\widehat{V}^{J^*}]\|^2 - \frac{2c_1\sqrt{\log\log n}\,\mathrm{v}_{J^*}}{n}\right)\\
&\leq T_1+T_2,
\end{aligned}$$

$$T_1 := \mathrm{P}_h\Big(\Big|\frac{4}{n(n-1)}\sum_{j=1}^{J^*}\sum_{i<i'}\big(\widehat{V}_{ij}\widehat{V}_{i'j} - \operatorname{E}_h[\widehat{V}_{1j}]^2\big)\Big| > \rho_h\Big)$$

$$T_2 := \mathrm{P}_h\Big(\Big|\frac{4}{n(n-1)}\sum_{i<i'}(Y_i - \widehat{h}^{\mathrm{R}}_{J^*}(X_i))(Y_{i'} - \widehat{h}^{\mathrm{R}}_{J^*}(X_{i'}))b^{K^*}(W_i)'\big(A'A - \widehat{A}'\widehat{A}\big)b^{K^*}(W_{i'})\Big| > \rho_h\Big),$$

where $\rho_h = \|h-\mathcal{H}_0\|^2_{L^2(X)}/2 - 2c_1 n^{-1}\sqrt{\log\log n}\,\mathrm{v}_{J^*} - C_B\mathsf{r}_n^2$. To establish an upper bound of T_1, we make use of Lemma E.3 which yields

$$T_1 \lesssim n^{-1}s_{J^*}^{-2}\rho_h^{-2}\mathcal{C}_h^2\big(\|h-\mathcal{H}_0\|^2_{L^2(X)} + (J^*)^{-2p/d_x}\big) + n^{-2}s_{J^*}^{-4}J^*\rho_h^{-2}. \tag{B.2}$$

First, consider the case where $n^{-2}s_{J^*}^{-4}J^*\rho_h^{-2}$ dominates the right hand side. For any $h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)$ we have $\|h-\mathcal{H}_0\|_{L^2(X)} \geq \delta^\circ\mathsf{r}_n$ for some sufficiently large $\delta^\circ>0$ and hence, we obtain the lower bound $\rho_h \geq ((\delta^\circ)^2/2 - C - C_B)\mathsf{r}_n^2$ for some constant $C>0$. Consequently, we have $T_1 \lesssim n^{-2}s_{J^*}^{-4}J^*(J^*)^{4p/d_x} = o(1)$. Second, consider the case where $n^{-1}s_{J^*}^{-2}\rho_h^{-2}\mathcal{C}_h^2\big(\|h-\mathcal{H}_0\|^2_{L^2(X)} + (J^*)^{-2p/d_x}\big)$ dominates. For any $h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)$ we have $\|h-\mathcal{H}_0\|^2_{L^2(X)} \geq (\delta^\circ)^2\mathsf{r}_n^2 \geq 5c_1 n^{-1}\mathrm{v}_{J^*}\sqrt{\log\log n}$ and we obtain the lower bound $\rho_h \geq (1/5 - C_B/(\delta^\circ)^2)\,\|h-\mathcal{H}_0\|^2_{L^2(X)}$. Hence, (B.2) yields uniformly for $h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)$ that

$$T_1 \lesssim n^{-1}s_{J^*}^{-2}\mathcal{C}_h^2\left(\|h-\mathcal{H}_0\|^{-2}_{L^2(X)} + \|h-\mathcal{H}_0\|^{-4}_{L^2(X)}(J^*)^{-2p/d}\right) \lesssim n^{-1}s_{J^*}^{-2}\sqrt{J^*}\mathsf{r}_n^{-2} = o(1)$$

using that $\sup_{h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)}\mathcal{C}_h^2 \lesssim \sqrt{J^*}$ by Assumption 5(ii). Finally, $T_2 = o(1)$ uniformly for $h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)$ by making use of Lemma E.4.

Step 3: Finally, we account for estimation of the normalization factor v_J and for estimation of the upper bound of the RES index set $\widehat{\mathcal{I}}_n$. Lemma B.8(i) implies $\sup_{h\in\mathcal{H}_0}\mathrm{P}_h\big(\widehat{J}_{\max} > \overline{J}\big) = o(1)$. We thus control the type I error of the test $\widehat{\mathtt{T}}_n$ for testing composite hypotheses, as follows. By the lower bound of Lemma B.5 we have

$$\mathrm{P}_h\left(\widehat{\mathtt{T}}_n = 1\right) \leq \mathrm{P}_h\left(\max_{J\in\mathcal{I}_n}\frac{n\widehat{D}_J}{\eta'_J(\alpha)\mathrm{v}_J} > 1\right) + \mathrm{P}_h\left(\max_{J\in\mathcal{I}_n}|\widehat{\mathrm{v}}_J/\mathrm{v}_J - 1| > c_0\right) + o(1) \leq \alpha + o(1)$$

uniformly for $h\in\mathcal{H}_0$, where the last inequality is due to Step 1 of this proof and Lemma B.4(ii). To bound the type II error of the test $\widehat{\mathtt{T}}_n$ recall the definition of $J^*\in\mathcal{I}_n$ introduced in Step 2 and note that $\sup_{h\in\mathcal{H}}\mathrm{P}_h\big(J^* > \widehat{J}_{\max}\big) = o(1)$ by Lemma B.8(ii). Consequently, the upper bound of Lemma B.5 and another application of Lemma B.8(ii) give uniformly for $h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)$: $\mathrm{P}_h\left(\widehat{\mathtt{T}}_n = 0\right) \leq \mathrm{P}_h\big(n\widehat{D}_{J^*} \leq \eta''(\alpha)\mathrm{v}_{J^*}\big) + \mathrm{P}_h\left(|\widehat{\mathrm{v}}_{J^*}/\mathrm{v}_{J^*} - 1| > c_0\right) + o(1) = o(1)$, where the last equation is due to Step 2 and Lemma B.4(i). □

Lemma B.9. *Let Assumptions 1(i)-(iii), 2(i), 4, and 5(i) be satisfied. Recall the notation* $h_J^* = \arg\min_{\phi\in\mathcal{H}_{0,J}} \|\sum_i(\phi - h)(X_i)\widehat{A}b^K(W_i)\|$. *Then, for all* $\varepsilon > 0$ *we have*

$$\sup_{h\in\mathcal{H}_0} \mathrm{P}_h\Big(\max_{J\in\mathcal{I}_n} \big\|\big(n v_J\sqrt{(\log\log J)/J}\big)^{-1/2}\sum_i (h - h_J^*)(X_i)\widehat{A}b^K(W_i)\big\| > \varepsilon\Big) = o(1).$$

Proof. The result is immediate under parametric null hypotheses. We now consider the nonparametric case, where the semiparametric situation follows analogously. Define $\widetilde{\Pi}_{\mathcal{B}}h := \arg\min_{\phi\in\mathcal{B}} \big\|\sum_i(\phi - h)(X_i)\widehat{A}b^K(W_i)\big\|$ for any closed, convex set $\mathcal{B}\subset\mathcal{H}$ and $\Psi_{J,h} := \{\phi : \phi = \kappa_1 Q_1 h + \ldots + \kappa_J Q_J h \text{ where } \sum_{j=1}^J |\kappa_j| \le 1\} \subset \Psi_J$ for any $h\in\mathcal{H}$. We have $0\in\Psi_{J,h}$, in particular, the zero function belongs to the interior of $\Psi_{J,h}$. Thus, $0\in\mathcal{H}_0$ implies that the zero function belongs to the interior of $\Psi_{J,h} - \mathcal{H}_0$. Now using that $\mathcal{H}_0$ and $\Psi_{J,h}$ are closed and convex subsets of $\mathcal{H}$ we may apply Bauschke and Borwein [1993, Corollary 4.5(i)]: there exists $h_J \in \Psi_{J,h}\cap\mathcal{H}_0 \neq \emptyset$ and $0 < c < 1$ such that

$$\sup_{h\in\mathcal{H}_0} \mathrm{P}_h\Big(\max_{J\in\mathcal{I}_n}\Big\{\Big\|n^{-1}\sum_i \big(h_J - (\widetilde{\Pi}_{\Psi_{J,h}}\widetilde{\Pi}_{\mathcal{H}_0})^m h\big)(X_i)\widehat{A}b^K(W_i)\Big\| \lesssim c^m\Big\}\Big) = 1 - o(1) \quad \text{(B.3)}$$

for all $m\ge 1$. Here, we used also that $\Psi_{J,h}\subset\Psi_{J',h}$ whenever $J < J'$. The definition of h_J^* implies

$$\begin{aligned}\Big\|\sum_i (h - h_J^*)(X_i)\widehat{A}b^K(W_i)\Big\| &\le \Big\|\sum_i (h - h_J)(X_i)\widehat{A}b^K(W_i)\Big\| \\ &\le \Big\|\sum_i \big(h - (\widetilde{\Pi}_{\Psi_{J,h}}\widetilde{\Pi}_{\mathcal{H}_0})^m h\big)(X_i)\widehat{A}b^K(W_i)\Big\| + \Big\|\sum_i \big((\widetilde{\Pi}_{\Psi_{J,h}}\widetilde{\Pi}_{\mathcal{H}_0})^m h - h_J\big)(X_i)\widehat{A}b^K(W_i)\Big\|.\end{aligned}$$

We make use of the decomposition $h - (\widetilde{\Pi}_{\Psi_{J,h}}\widetilde{\Pi}_{\mathcal{H}_0})^m h = \big(\mathrm{id} + \widetilde{\Pi}_{\Psi_{J,h}}\widetilde{\Pi}_{\mathcal{H}_0} + \ldots + (\widetilde{\Pi}_{\Psi_{J,h}}\widetilde{\Pi}_{\mathcal{H}_0})^m\big)(h - \widetilde{\Pi}_{\Psi_{J,h}}\widetilde{\Pi}_{\mathcal{H}_0} h)$. We may assume that $h\in\mathcal{H}_0$ does not belong to $\Psi_{J,h}$ and thus, $\widetilde{\Pi}_{\Psi_{J,h}}\widetilde{\Pi}_{\mathcal{H}_0}$ forms a contraction satisfying

$$\Big\|\sum_i \big(h - (\widetilde{\Pi}_{\Psi_{J,h}}\widetilde{\Pi}_{\mathcal{H}_0})^m h\big)(X_i)\widehat{A}b^K(W_i)\Big\| \le \Big\|\sum_i \big(h - \widetilde{\Pi}_{\Psi_{J,h}} h\big)(X_i)\widehat{A}b^K(W_i)\Big\|.$$

Choosing $m = \lfloor \log_c(J^{-1/2}\sqrt{v_J/n})\rfloor$ we have $m\ge 1$ for n sufficiently large by the upper bound on v_J established in Lemma B.2, Assumption 4(ii), and using that $0 < c < 1$. Plugging this choice of m in equation (B.3) thus implies

$$\begin{aligned}&\Big\|\big(n v_J\sqrt{(\log\log J)/J}\big)^{-1/2}\sum_i (h - h_J^*)(X_i)\widehat{A}b^K(W_i)\Big\| \\ &\quad\lesssim \Big\|\big(n v_J\sqrt{(\log\log J)/J}\big)^{-1/2}\sum_i (h - Q_J h)(X_i)\widehat{A}b^K(W_i)\Big\| + J^{-1/2}\end{aligned}$$

with probability approaching one, uniformly for $h \in \mathcal{H}_0$, using that $Q_J h \in \Psi_{J,h}$. It is sufficient to consider the first summand on the right hand side since $\max_{J\in\mathcal{I}_n} J^{-1/2} = \underline{J}^{-1/2} = o(1)$. First, we consider the off-diagonal summands:

$$\frac{\sqrt{J}}{n}\sum_{i\neq i'}(h-Q_Jh)(X_i)(h-Q_Jh)(X_{i'})b^K(W_i)'A'Ab^K(W_{i'})$$
$$+\frac{\sqrt{J}}{n}\sum_{i\neq i'}(h-Q_Jh)(X_i)(h-Q_Jh)(X_{i'})b^K(W_i)'\big(\widehat{A}'\widehat{A}-A'A\big)b^K(W_{i'}) =: T_{31,J}+T_{32,J}.$$

Consider $T_{31,J}$. By the definition of $Q_Jh(\cdot)=\widetilde{\psi}^J(\cdot)'A\,\mathrm{E}[b^K(W)h(X)]$ we observe

$$\mathrm{E}\left[(h-Q_Jh)(X)Ab^K(W)\right] = \mathrm{E}\left[Q_J(h-Q_Jh)(X)\widetilde{\psi}^J(X)\right]=0.$$

Further, we infer for all $J\in\mathcal{I}_n$ that $\sqrt{\mathrm{E}[(Q_Jh-h)^2(X)|W]}\lesssim \|Q_Jh-h\|_{L^2(X)}\lesssim J^{-p/d_x}$ wpa1 uniformly for $h\in\mathcal{H}_0$ by Lemma B.1(ii) and thus, $\mathrm{E}\,|\sqrt{J}\,\mathrm{E}[(Q_Jh-h)^2(X)|W]| = o(1)$ by Assumption 4(iii). Further, we obtain for all $J\in\mathcal{I}_n$ and uniformly for $h\in\mathcal{H}_0$:

$$\mathrm{E}[(Q_J(h-\Pi_Jh))^4(X)]\lesssim \zeta_J^2\big\|(G_b^{-1/2}SG^{-1/2})_\ell^-\,\mathrm{E}[(h-\Pi_Jh)(X)\widetilde{b}^K(W)]\big\|^4\lesssim \zeta_J^2J^{-4p/d_x}$$

and $J\,\mathrm{E}[(Q_J(h-\Pi_Jh))^4(X)]=o(1)$ by Assumption 4(iii). Using these moment bounds, we may follow step 1 of the proof of Theorem 4.1 by replacing $Y_i-h(X_i)$ with $J^{1/4}(Q_Jh-h)(X_i)$ for $h\in\mathcal{H}_0$ and for any $\varepsilon>0$ obtain $\mathrm{P}_h(\max_{J\in\mathcal{I}_n}T_{31,J}/(\mathrm{v}_J\sqrt{\log\log J})>\varepsilon)=o(1)$ uniformly for $h\in\mathcal{H}_0$. Consider $T_{32,J}$. For any $\varepsilon>0$, we have $\mathrm{P}_h(\max_{J\in\mathcal{I}_n}T_{32,J}/(\mathrm{v}_J\sqrt{\log\log J})>\varepsilon)=o(1)$ uniformly for $h\in\mathcal{H}_0$, following Lemma E.6 again by replacing $Y_i-h(X_i)$ with $J^{1/4}(Q_Jh-h)(X_i)$ for $h\in\mathcal{H}_0$.

Finally, we control the diagonal elements of $J^{1/4}\|n^{-1/2}\sum_i(h-Q_Jh)(X_i)\widehat{A}b^K(W_i)\|$. To do so, we make use of the decomposition

$$\frac{\sqrt{J}}{n}\sum_i\Big\|(h-Q_Jh)(X_i)Ab^K(W_i)\Big\|^2+\frac{\sqrt{J}}{n}\sum_i\Big\|(h-Q_Jh)(X_i)\big(\widehat{A}-A\big)b^K(W_i)\Big\|^2=:T_{41,J}+T_{42,J}.$$

Using Lemma E.5(i), for any $\varepsilon>0$ we obtain $\mathrm{P}_h\big(\max_{J\in\mathcal{I}_n}T_{42,J}/(\mathrm{v}_J\sqrt{\log\log J})>\varepsilon\big)=o(1)$ uniformly for $h\in\mathcal{H}_0$ and thus it is sufficient to consider $T_{41,J}$. We have

$$\max_{J\in\mathcal{I}_n}\frac{T_{41,J}}{\mathrm{v}_J\sqrt{\log\log J}}\lesssim\max_{J\in\mathcal{I}_n}\frac{\sqrt{J}\big(\|h-Q_Jh\|_{L^2(X)}\zeta_Js_J^{-1}\big)^2}{\mathrm{v}_J\sqrt{\log\log J}}$$

wpa1 uniformly for $h\in\mathcal{H}_0$. The right hand side tends to zero using that $\|h-Q_Jh\|_{L^2(X)}=O(J^{-p/d_x})$ and Assumption 4(iii) together with $s_J^{-2}\le\underline{\sigma}^{-2}\mathrm{v}_J$ (by Lemma B.3). $\square$

Supplement to "Adaptive, Rate-Optimal Hypothesis Testing in Nonparametric IV Models"

Christoph Breunig Xiaohong Chen

First version: August 2018; Latest revision: February 5, 2024

This supplementary appendix contains materials to support our main paper. Appendix C presents additional simulation results. Appendix D provides proofs of our results on confidence sets in Subsection 4.3. Appendix E presents additional technical lemmas and all the proofs.

C. Additional Simulations

This section provides additional simulation results. All the simulation results are based on 5000 Monte Carlo replications for every experiment. Due to the lack of space we report simulation results for the nominal level $\alpha = 0.05$ unless stated otherwise.

C.1. Adaptive Testing for Monotonicity: Simulation Design II

We generate the dependent variable Y according to the NPIV model (2.1), where

$$h(x) = c_0(x/5 + x^2) + c_A \sin(2\pi x) \ , \tag{C.1}$$

$c_0 \in \{0, 1\}$, $c_A \in [0, 1]$, and $W = \Phi(W^*)$, $X = \Phi\big(\xi W^* + \sqrt{1-\xi^2}\epsilon\big)$, $U = (0.3\epsilon + \sqrt{1-(0.3)^2}\nu)/2$, where (W^*, ϵ, ν) follows a multivariate standard normal distribution.[1] The null hypothesis is that the NPIV function $h(\cdot)$ is weakly increasing on the support of X. The null is satisfied when $c_A \in [0, 0.184)$, and is violated when $c_A \geq 0.184$. Note that the degree of nonlinearity/complexity of h given in (C.1) becomes larger as c_A increases.

Table A reports the empirical size of our adaptive test $\widehat{\mathtt{T}}_n$ given in (2.13) in the main paper, with the 5% nominal level, using quadratic B-spline basis functions with varying number of knots for h. We report simulation size results for our adaptive test $\widehat{\mathtt{T}}_n$ with $K(J) = 2J$ and $K(J) = 4J$ since its computation is fast. As comparison, we also report the empirical size of the nonadaptive bootstrap test $\mathtt{T}^{\mathtt{B}}_{n,3}$ for monotonicity with $J = 3$ and $K = 4J = 12$. It is time consuming to compute $\mathtt{T}^{\mathtt{B}}_{n,3}$ in simulations especially for a larger sample size $n = 5000$. Since the empirical size results for $\mathtt{T}^{\mathtt{B}}_{n,3}$ using $K = 2J = 6$ are similar

[1] This design with $(c_0, c_A) = (1, 0)$ and $\xi \in \{0.3, 0.5\}$ becomes the one considered by Chetverikov and Wilhelm [2017].

n	c_0	c_A	ξ	$\widehat{\mathtt{T}}_n$ *with* $K=2J$	$\widehat{J}$	$\widehat{\mathtt{T}}_n$ *with* $K=4J$	$\widehat{J}$	$\mathtt{T}^{\mathtt{B}}_{n,3}$ *with* $K=4J=12$
500	0	0.0	0.3	0.006	3.00	0.013	3.03	0.009
			0.5	0.018	3.33	0.019	3.39	0.018
			0.7	0.028	3.57	0.027	3.59	0.012
	1	0.0	0.3	0.002	3.00	0.004	3.03	0.005
			0.5	0.007	3.38	0.005	3.36	0.009
			0.7	0.005	3.71	0.004	3.64	0.004
	1	0.1	0.3	0.002	3.00	0.006	3.03	0.006
			0.5	0.009	3.38	0.009	3.35	0.010
			0.7	0.008	3.65	0.009	3.59	0.007
1000	0	0.0	0.3	0.010	3.02	0.013	3.07	0.015
			0.5	0.023	3.50	0.023	3.46	0.020
			0.7	0.032	3.87	0.035	3.96	0.012
	1	0.0	0.3	0.003	3.02	0.005	3.06	0.007
			0.5	0.006	3.63	0.006	3.46	0.007
			0.7	0.004	4.21	0.004	4.21	0.006
	1	0.1	0.3	0.004	3.02	0.006	3.06	0.009
			0.5	0.010	3.59	0.010	3.45	0.012
			0.7	0.011	4.09	0.009	4.09	0.006
5000	0	0.0	0.3	0.019	3.37	0.024	3.40	0.033
			0.5	0.036	3.57	0.043	3.63	0.028
			0.7	0.041	4.14	0.044	4.13	0.014
	1	0.0	0.3	0.006	3.42	0.006	3.35	0.005
			0.5	0.003	3.80	0.004	3.80	0.009
			0.7	0.002	4.73	0.002	4.70	0.006
	1	0.1	0.3	0.008	3.41	0.011	3.35	0.006
			0.5	0.012	3.70	0.014	3.70	0.009
			0.7	0.007	4.51	0.006	4.46	0.006

Table A: Testing Monotonicity - Empirical Size of our adaptive test $\widehat{\mathtt{T}}_n$ and of the nonadaptive bootstrap test $\mathtt{T}^{B}_{n,3}$. Nominal level $\alpha = 0.05$. Design from Appendix C.1 with NPIV function (C.1). Instrument strength increases in ξ.

to the ones using $K = 4J = 12$, we skip reporting the results for $\mathtt{T}^{\mathtt{B}}_{n,3}$ with $K = 2J = 6$. Again we observe that our adaptive test $\widehat{\mathtt{T}}_n$ and the nonadaptive bootstrap test $\mathtt{T}^{\mathtt{B}}_{n,3}$ provide adequate size control across different design specifications. Both are similarly undersized.

Figure A provides empirical rejection probabilities of our adaptive test $\widehat{\mathtt{T}}_n$ (blue solid lines) and of the nonadaptive bootstrap test $\mathtt{T}^{\mathtt{B}}_{n,3}$ (green dashed lines with a fixed sieve dimension $J = 3$), both use $K(J) = 4J$. The powers of both tests improve as the instrument strength ξ increases. For instrument strength $\xi = 0.3, 0.5$, the nonadaptive bootstrap test $\mathtt{T}^{\mathtt{B}}_{n,3}$ has almost trivial power for all $c_A \geq 0.2$, but, our adaptive test $\widehat{\mathtt{T}}_n$ has non-trivial power for all $c_A > 0.2$. Moreover, the finite sample power of our adaptive test $\widehat{\mathtt{T}}_n$ increases much faster than that of the nonadaptive bootstrap test as $c_A > 0.2$ becomes larger. Figure A shows the substantial finite sample power gains through adaptation even in small sample size $n = 500$. The same patterns are also present when we compare the two tests using size-adjusted empirical power curves (see our arXiv:2006.09587v3 version, Appendix C.1).

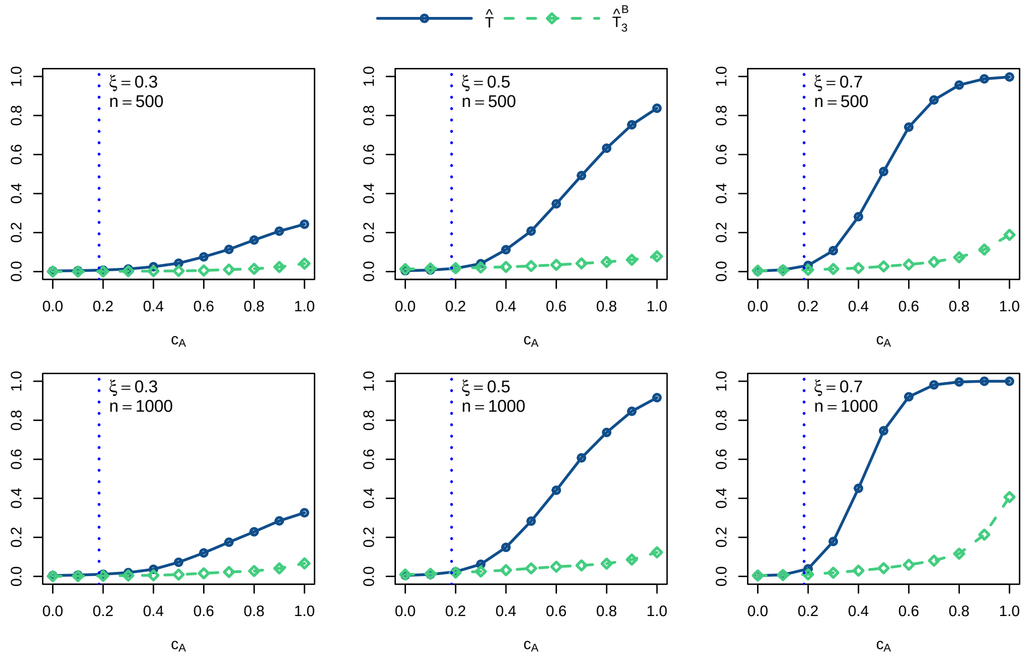


Figure A: Testing Monotonicity – Empirical power of our adaptive $\widehat{\mathtt{T}}_n$ (blue solid lines) and of the nonadaptive bootstrap test $\mathtt{T}^{\mathtt{B}}_{n,3}$ (green dashed lines), both with $K = 4J$. Design from Appendix C.1 model (C.1) with $c_0 = 1$. The vertical dotted line indicates when the null hypothesis is violated (when $c_A \geq 0.184$). Instrument strength increases in ξ.

Remark C.1. *When testing for inequality restrictions (IR)* $\mathcal{H}_0 = \{h \in \mathcal{H} : \partial^l h \geq 0\}$, *such as monotonicity and convexity, we could also compute our adaptive test* $\widehat{\mathtt{T}}_n$ *using modified critical values in Step 2 as follows: The estimator in* (2.7) *can be written as* $\widehat{h}^{\mathrm{R}}_J(\cdot) = \psi^J(\cdot)'\widehat{\beta}^{\mathrm{R}}$. *By construction of the estimator we have* $\partial^l \widehat{h}^{\mathrm{R}}_J(X_i) \geq 0$, *for all* $1 \leq i \leq n$, *or equivalently* $\partial^l \Psi \widehat{\beta}^{\mathrm{R}} \geq 0$, *where the application of the derivative operator is understood elementwise and rank*$(\partial^l \Psi) \leq J$. *Let* Ψ_{act} *be a submatrix of* Ψ *such that* $\partial^l \Psi_{act}\widehat{\beta}^{\mathrm{R}} = 0$. *Set* $\widehat{\gamma}_J = \max\big(1, rank(\partial^l \Psi_{act})\big)$ *and compute for a given nominal level* $\alpha \in (0,1)$:

$$\widehat{\eta}_J(\alpha) = \frac{q\big(\alpha/\#(\widehat{\mathcal{I}}_n), \widehat{\gamma}_J\big) - \widehat{\gamma}_J}{\sqrt{\widehat{\gamma}_J}}, \tag{C.2}$$

where $q(a, \gamma)$ *denotes the* $100(1-a)\%$*-quantile of the chi-square distribution with* γ *degrees of freedom. Assuming that* $J^c \leq \widehat{\gamma}_J$, $J \in \mathcal{I}_n$, *for some constant* $0 < c \leq 1$ *with probability approaching one uniformly for* $h \in \mathcal{H}$, *Breunig and Chen [2021] establishes size control of the test statistic using the modified critical values given in* (C.2). *See Breunig and Chen [2021] also for simulations and real data application of testing for monotonicity and convexity using this modified critical values. The simulations and empirical findings reported in Breunig and Chen [2021] are virtually the same, in terms of empirical size and power, as the ones reported in this revised version for testing inequalities.*

C.2. Empirical Power for Subsection 5.2 with Larger Sample Size

Figure B below provides additional power comparison for the simulation design stated in Subsection 5.2. It replicates Figure 3 using a larger sample size $n = 5000$.

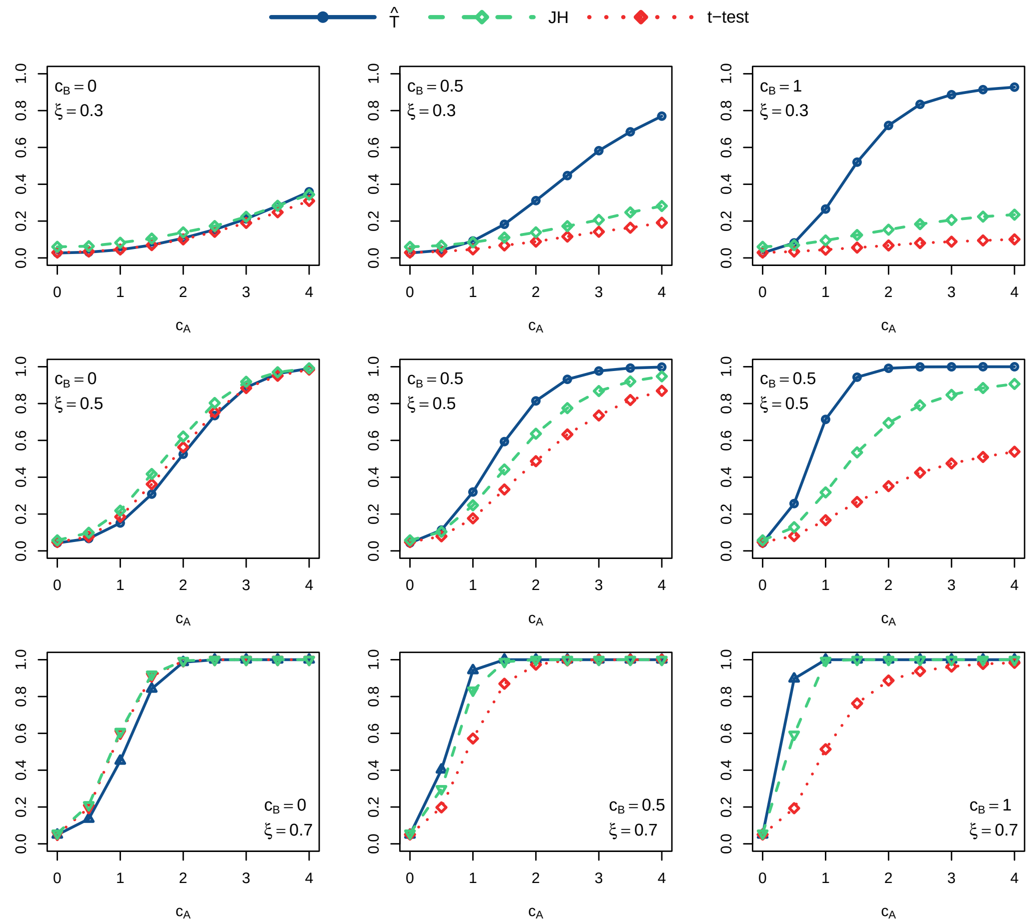


Figure B: Testing Parametric Form - Empirical power of our adaptive test $\widehat{\mathtt{T}}_n$ (blue solid lines) with $K = 4J$, of JH test (green dashed lines) and of t-test (red dotted lines). True DGP from Section 5.2 using NPIV function (5.3). Alternatives are quadratic when $c_B = 0$ and become more complex as $c_B > 0$ increases. Instrument strength increases in ξ. Replication of Figure 3 with $n = 5000$.

C.3. Simulations for Multivariate Instruments

This section presents additional simulations for testing parametric hypotheses in the presence of multivariate conditioning variable $W = (W_1, W_2)$. We set $X_i = \Phi(X_i^*)$, $W_{1i} = \Phi(W_{1i}^*)$, and $W_{2i} = \Phi(W_{2i}^*)$, where

$$\begin{pmatrix} X_i^* \\ W_{1i}^* \\ W_{2i}^* \\ U_i \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \xi & 0.4 & 0.3 \\ \xi & 1 & 0 & 0 \\ 0.4 & 0 & 1 & 0 \\ 0.3 & 0 & 0 & 1 \end{pmatrix} \right). \tag{C.3}$$

We generate the dependent variable Y according to the NPIV model (2.1) where $h(x) = -x/5 + c_A x^2$. We test the null hypothesis of linearity, i.e., whether $c_A = 0$.

Horowitz [2006] assumes $d_x = d_w$ and hence we cannot compare our adaptive test with his for Design (C.3). Instead we will compare our adaptive test $\widehat{\mathtt{T}}_n$ against an adaptive image-space test (IT), which is our proposed adaptive version of Bierens [1990]'s type test for semi-nonparametric conditional moment restrictions.[2] Specifically, our image-space test (IT) is based on a leave-one-out sieve estimator of the quadratic functional $\mathrm{E}[\mathrm{E}[Y - h^{\mathrm{R}}(X)|W]^2]$, given by

$$\widehat{D}_K = \frac{2}{n(n-1)} \sum_{1 \le i < i' \le n} \big(Y_i - \widehat{h}^{\mathrm{R}}(X_i)\big)\big(Y_{i'} - \widehat{h}^{\mathrm{R}}(X_{i'})\big) b^K(W_i)'(B'B/n)^{-} b^K(W_{i'}),$$

where $\widehat{h}^{\mathrm{R}}$ is a null restricted parametric estimator for the null parametric function h^{R}. The data-driven IT statistic is:

$$\widehat{\mathtt{IT}}_n = \mathbb{1}\left\{\text{there exists } K \in \widehat{\mathcal{I}}_n \text{ such that } n\widehat{D}_K/\widehat{\mathrm{v}}_K > \big(q\big(\alpha/\#(\widehat{\mathcal{I}}_n), K\big) - K\big)/\sqrt{K}\right\}$$

with the estimator $\widehat{\mathrm{v}}_K = \big\|(B'B)^{-1/2} \sum_{i=1}^n (Y_i - \widehat{h}^{\mathrm{R}}(X_i))^2 b^K(W_i) b^K(W_i)'(B'B)^{-1/2}\big\|_F$, and the adjusted index set $\widehat{\mathcal{I}}_n = \{K \le \widehat{K}_{\max} : K = \underline{K}2^k \text{ where } k = 0, 1, \ldots, k_{\max}\}$, where $\underline{K} := \lfloor \sqrt{\log\log n} \rfloor$, $k_{\max} := \lceil \log_2(n^{1/3}/\underline{K}) \rceil$, and the empirical upper bound $\widehat{K}_{\max} = \min\big\{K > \underline{K} : 1.5\,\zeta^2(K)\sqrt{(\log K)/n} \ge s_{\min}\big((B'B/n)^{-1/2}\big)\big\}$. Finally $q\big(a, K\big)$ is the $100(1-a)\%$-quantile of the chi-square distribution with K degrees of freedom. Table B compares the empirical size of the adaptive image space test $\widehat{\mathtt{IT}}_n$ with our adaptive structural space test $\widehat{\mathtt{T}}_n$, at the 5% nominal level. We see that both tests provide accurate size control. We also report the average choices of sieve dimension parameters, as described in Section 5. The multivariate design (C.3) leads to larger sieve dimension choices $\widehat{K}$ in adaptive image-space tests $\widehat{\mathtt{IT}}_n$ while the sieve dimension choices $\widehat{J}$ of our adaptive structural-space test $\widehat{\mathtt{T}}_n$ is not sensitive to the dimensionality (d_w) of the conditional instruments.

Figure C compares the empirical power of $\widehat{\mathtt{IT}}_n$ and of $\widehat{\mathtt{T}}_n$, at the 5% nominal level, using the sample sizes $n = 500$ (1st and 2nd rows) and $n = 1000$ (3rd and 4th rows). The finite sample empirical powers of both tests increase with ξ and sample size n. For the scalar conditional instrument case, while our adaptive structural space test $\widehat{\mathtt{T}}_n$ is more powerful when $\xi = 0.3, 0.5$ (weaker strength of instruments), the finite sample powers of both tests are similar when $\xi = 0.7$. For the multivariate conditional instruments case, while the power of our adaptive structural space test $\widehat{\mathtt{T}}_n$ increases with larger dimension d_w, the adaptive image space test $\widehat{\mathtt{IT}}_n$ suffers from larger d_w and has lower power. The same patterns are also present when we compare the two tests using size-adjusted empirical

[2] We refer readers to Breunig and Chen [2020] for the theoretical properties of the adaptive image-space test.

n	Design	ξ	$\widehat{\mathtt{T}}_n$ *with* $K=4J$	$\widehat{J}$	$\widehat{\mathtt{IT}}_n$	$\widehat{K}$
500	(5.1)	0.3	0.023	3.12	0.051	4.44
	$d_x = d_w$	0.5	0.030	3.46	0.050	4.44
		0.7	0.032	3.87	0.051	4.42
	(C.3)	0.3	0.035	3.46	0.038	8.99
	$d_x < d_w$	0.5	0.039	3.49	0.042	8.97
		0.7	0.039	3.88	0.037	8.89
1000	(5.1)	0.3	0.023	3.17	0.045	4.40
		0.5	0.030	3.51	0.051	4.39
		0.7	0.039	4.09	0.052	4.40
	(C.3)	0.3	0.037	3.49	0.035	9.03
		0.5	0.042	3.57	0.042	8.91
		0.7	0.041	4.07	0.043	8.96
5000	(5.1)	0.3	0.028	3.41	0.053	5.10
		0.5	0.042	3.64	0.055	5.10
		0.7	0.048	4.18	0.053	5.10
	(C.3)	0.3	0.050	3.84	0.045	10.17
		0.5	0.054	4.00	0.049	10.14
		0.7	0.055	4.15	0.054	10.14

Table B: Testing Parametric Form - Empirical size of our adaptive tests $\widehat{\mathtt{T}}_n$ and of $\widehat{\mathtt{IT}}_n$. Nominal level $\alpha = 0.05$. Design from Appendix C.3. Instrument strength increases in ξ.

power curves (see our arXiv:2006.09587v3 version, Appendix C.3).

D. Proofs of Inference Results in Subsection 4.3

Proof of Corollary 4.1. Proof of (4.9). We observe

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_0} \mathrm{P}_h\left(h \notin \mathcal{C}_n(\alpha)\right) = \limsup_{n\to\infty} \sup_{h\in\mathcal{H}_0} \mathrm{P}_h\left(\max_{J\in\widehat{\mathcal{I}}_n} \frac{n\widehat{D}_J(h)}{\widehat{\eta}_J(\alpha)\,\widehat{\mathrm{v}}_J} > 1\right) \le \alpha,$$

where the last inequality is due to step 1 of the proof of Theorem 4.1 and step 3 of the proof of Theorem 4.2.

Proof of (4.10). Let J^* be as be as in step 2 of the proof of Theorem 4.1. We observe uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathrm{r}_n)$ that

$$\mathrm{P}_h\left(h \notin \mathcal{C}_n(\alpha)\right) = \mathrm{P}_h\left(\max_{J\in\widehat{\mathcal{I}}_n} \frac{n\widehat{D}_J(h)}{\widehat{\eta}_J(\alpha)\widehat{\mathrm{v}}_J} > 1\right) = 1 - \mathrm{P}_h\left(\max_{J\in\widehat{\mathcal{I}}_n} \frac{n\widehat{D}_J(h)}{\widehat{\eta}_J(\alpha)\,\widehat{\mathrm{v}}_J} \le 1\right) = 1 - o(1),$$

where the last equation is due to step 2 of the proof of Theorem 4.1 and step 3 of the proof of Theorem 4.2. □

Proof of Corollary 4.2. For any $h \in \mathcal{H}_0$, we analyze the diameter of the confidence set $\mathcal{C}_n(\alpha)$ under P_h. Lemma B.8 implies $\sup_{h\in\mathcal{H}_0} \mathrm{P}_h\big(\widehat{J}_{\max} > \overline{J}\big) = o(1)$ and hence, it is sufficient to consider the deterministic index set $\mathcal{I}_n$ given in (4.2). For all $h_1 \in \mathcal{C}_n(\alpha) \subset \mathcal{H}_0$ it holds

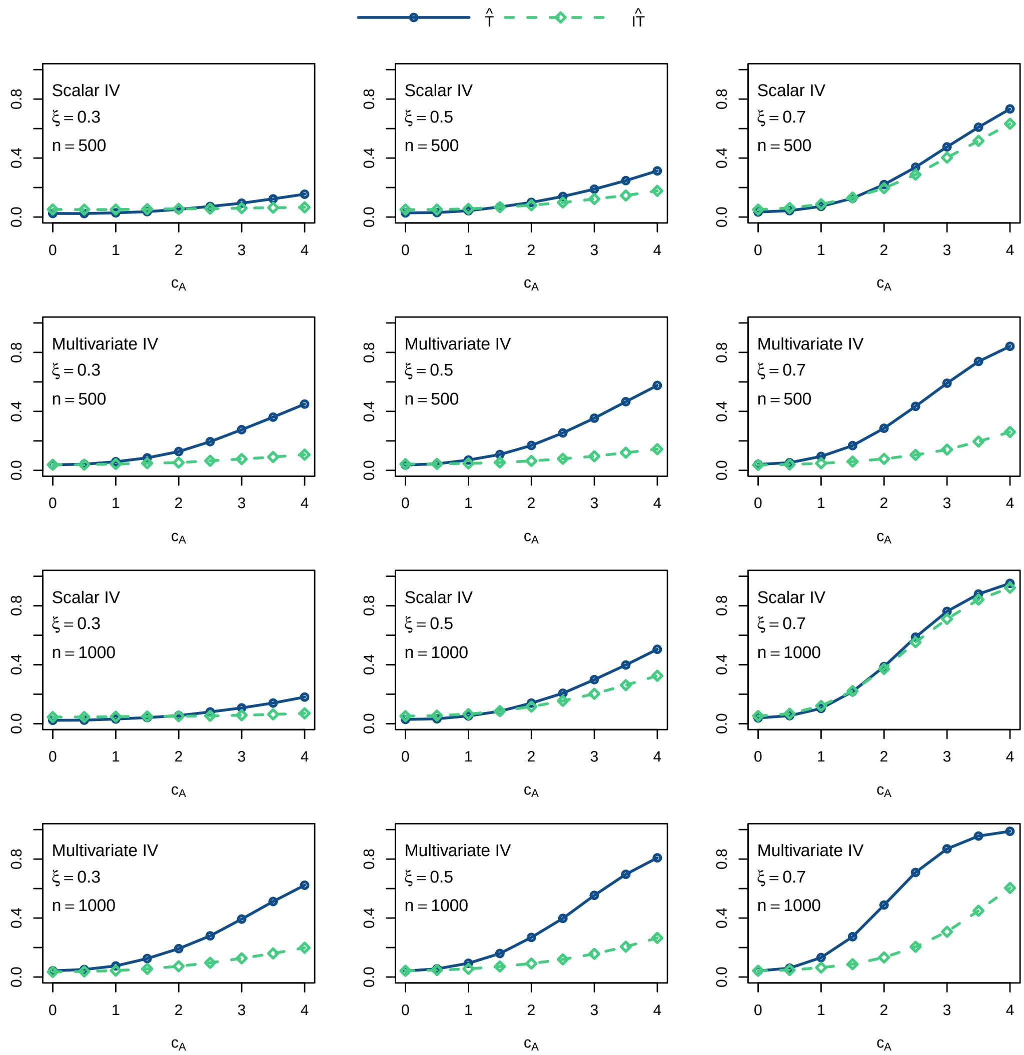


Figure C: Testing Parametric Form - Empirical power of our adaptive tests $\widehat{\mathtt{T}}_n$ (blue solid lines) and of $\widehat{\mathtt{IT}}_n$ (green dashed lines). 1st and 3rd rows: power comparisons in scalar IV case ($d_w = 1$); 2nd and 4th rows: power comparisons in multivariate IV case ($d_w > 1$). Design from Appendix C.3. Instrument strength increases in ξ.

for all $J \in \mathcal{I}_n$ by using the definition of the projection Q_J given in (B.1):

$$\begin{aligned}\|h - h_1\|_{L^2(X)} &\leq \|Q_J \Pi_J (h - h_1)\|_{L^2(X)} + \|\Pi_J h - h\|_{L^2(X)} + \|\Pi_J h_1 - h_1\|_{L^2(X)} \\ &\leq \|Q_J(h - h_1)\|_{L^2(X)} + O(J^{-p/d_x}),\end{aligned} \tag{D.1}$$

where the second inequality due to the triangular inequality and the sieve approximation bound from the smoothness restrictions imposed on $\mathcal{H}$. By Theorem B.1 we have

$$\left| \|Q_J(h - h_1)\|^2_{L^2(X)} - \widehat{D}_J(h_1) \right| \lesssim n^{-1/2} s_J^{-1} \big(\|h - h_1\|_{L^2(X)} + J^{-p/d_x} \big) + n^{-1} s_J^{-2} \sqrt{J}$$

wpa1 uniformly for $h \in \mathcal{H}_0$. Consequently, the definition of the confidence set $\mathcal{C}_n(\alpha)$ with

$h_1 \in \mathcal{C}_n(\alpha)$ gives for all $J \in \mathcal{I}_n$:

$$\begin{aligned}\|Q_J(h-h_1)\|^2_{L^2(X)} &\lesssim n^{-1}\widehat{\eta}_J(\alpha)\,\widehat{\mathrm{v}}_J + n^{-1/2}s_J^{-1}\big(\|h-h_1\|_{L^2(X)} + J^{-p/d_x}\big) + n^{-1}s_J^{-2}\sqrt{J}\\ &\lesssim n^{-1}\sqrt{\log\log n}\, s_J^{-2}\sqrt{J} + n^{-1/2}s_J^{-1}\big(\|h-h_1\|_{L^2(X)} + J^{-p/d_x}\big)\end{aligned}$$

wpa1 uniformly for $h \in \mathcal{H}_0$ by using Lemmas B.2, B.5 and B.4(ii). Consequently, inequality (D.1) yields

$$\|h-h_1\|^2_{L^2(X)} \lesssim \frac{n^{-1}\sqrt{\log\log n}\, s_J^{-2}\sqrt{J} + J^{-2p/d_x}}{1 - C_B n^{-1/2}s_J^{-1}}.$$

wpa1 uniformly for $h \in \mathcal{H}_0$. Now using that $n^{-1/2}s_J^{-1} = o(1)$ for all $J \in \mathcal{I}_n$ by Assumption 4(i) we obtain $\|h-h_1\|_{L^2(X)} \lesssim n^{-1/2}(\log\log n)^{1/4}\, s_J^{-1}J^{1/4} + J^{-p/d_x}$ with probability approaching one uniformly for $h \in \mathcal{H}_0$. We may choose $J = J^\circ \in \mathcal{I}_n$ for n sufficiently large and hence, the result follows. □

E. Technical Results

Below, $\lambda_{\max}(\cdot)$ denotes the maximal eigenvalue of a matrix.

Lemma E.1. *Let Assumptions 1(ii)-(iii) and 2 hold. Then, wpa1 uniformly for $h \in \mathcal{H}$:*

$$\begin{aligned}&\frac{1}{n(n-1)}\sum_{i\neq i'}\big(Y_i - \Pi_{\mathcal{H}_0}h(X_i)\big)\big(Y_{i'} - \Pi_{\mathcal{H}_0}h(X_{i'})\big)b^K(W_i)'\Big(A'A - \widehat{A}'\widehat{A}\Big)b^K(W_{i'})\\ &\qquad\qquad \lesssim n^{-1}\mathrm{v}_J + n^{-1/2}s_J^{-1}\big(\|h - \Pi_{\mathcal{H}_0}h\|_{L^2(X)} + J^{-p/d_x}\big).\end{aligned}$$

Proof. Let $\Pi^\perp_{\mathcal{H}_0} := \mathrm{id} - \Pi_{\mathcal{H}_0}$. We establish an upper bound of

$$\begin{aligned}&\frac{1}{n^2}\sum_{i,i'}\big(Y_i - \Pi_{\mathcal{H}_0}h(X_i)\big)\big(Y_{i'} - \Pi_{\mathcal{H}_0}h(X_{i'})\big)b^K(W_i)'\Big(A'A - \widehat{A}'\widehat{A}\Big)b^K(W_{i'})\\ &= \mathrm{E}[\Pi^\perp_{\mathcal{H}_0}h(X)b^K(W)]'\Big(A'A - \widehat{A}'\widehat{A}\Big)\,\mathrm{E}[\Pi^\perp_{\mathcal{H}_0}h(X)b^K(W)]\\ &+ 2\Big(\frac{1}{n}\sum_i\big(Y_i - \Pi_{\mathcal{H}_0}h(X_i)\big)b^K(W_i) - \mathrm{E}[\Pi^\perp_{\mathcal{H}_0}h(X)b^K(W)]\Big)'\Big(A'A - \widehat{A}'\widehat{A}\Big)\,\mathrm{E}[\Pi^\perp_{\mathcal{H}_0}h(X)b^K(W)]\\ &+ \Big(\frac{1}{n}\sum_i\big(Y_i - \Pi_{\mathcal{H}_0}h(X_i)\big)b^K(W_i)' - \mathrm{E}[\Pi^\perp_{\mathcal{H}_0}h(X)b^K(W)]'\Big)\Big(A'A - \widehat{A}'\widehat{A}\Big)\\ &\qquad\qquad \times\Big(\frac{1}{n}\sum_i\big(Y_i - \Pi_{\mathcal{H}_0}h(X_i)\big)b^K(W_i)' - \mathrm{E}[\Pi^\perp_{\mathcal{H}_0}h(X)b^K(W)]'\Big)\end{aligned}$$

uniformly for $h \in \mathcal{H}$. It is sufficient to bound the first summand on the right hand side.

We make use of the decomposition

$$
\begin{aligned}
&\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)]' \Big(A'A - \widehat{A}'\widehat{A}\Big)\, \mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] \\
&\qquad = 2\, \mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)]' A'(A - \widehat{A})\, \mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] \\
&\qquad\quad - \mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)]'(A - \widehat{A})'(A - \widehat{A})\, \mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] =: 2T_1 - T_2.
\end{aligned}
$$

We first consider the term T_1 as follows:

$$
\begin{aligned}
T_1 &= \mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)]' A'(\widehat{A} - A)\, \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] \\
&\quad + \mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)]' A'(\widehat{A} - A)\, \mathrm{E}[(\Pi_{\mathcal{H}_0}^{\perp} h - \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h)(X) b^K(W)] := A_1 + A_2. \qquad \text{(E.1)}
\end{aligned}
$$

We now consider the term A_1. Recall that $Q_J \Pi_J h = \Pi_J h$ and $\widehat{S} G^{-1} \langle h, \psi^J \rangle_{L^2(X)} = n^{-1} \sum_i \Pi_J h(X_i) b^K(W_i)$ we have:

$$
\begin{aligned}
&\big((G_b^{-1/2} S)_l^- \,\mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W)]\big)' G \big((G_b^{-1/2} S)_l^- - (\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2}\big)\, \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W)] \\
&= \big\langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h - (\psi^J)'(\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2}\, \mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] \big\rangle_{L^2(X)} \\
&= \langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle_{L^2(X)}' (\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} \Big(\frac{1}{n} \sum_i \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X_i) b^K(W_i) - \mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] \Big) \\
&= \langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle_{L^2(X)}' (G_b^{-1/2} S)_l^- \Big(\frac{1}{n} \sum_i \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X_i) \widetilde{b}^K(W_i) - \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W)] \Big) \\
&\quad + \langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle_{L^2(X)}' (G_b^{-1/2} S)_l^- G_b^{-1/2} S' \Big((\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^- \Big) \\
&\qquad \times \Big(\frac{1}{n} \sum_i \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X_i) \widetilde{b}^K(W_i) - \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W)] \Big) =: A_{11} + A_{12},
\end{aligned}
$$

where we used the notation $\widetilde{b}^K(\cdot) = G_b^{-1/2} b^K(\cdot)$. Consider A_{11} we have:

$$
\begin{aligned}
\mathrm{E}\, |A_{11}|^2 &\le n^{-1}\, \mathrm{E} \Big| \langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle_{L^2(X)}' (G_b^{-1/2} S)_l^- \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W) \Big|^2 \\
&\le 2n^{-1} \big\| \langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle_{L^2(X)}' (G_b^{-1/2} S)_l^- \big\|^2 \|\Pi_K T \Pi_{\mathcal{H}_0}^{\perp} h\|_{L^2(W)}^2 \\
&\quad + 2n^{-1} \big\| \langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle_{L^2(X)}' (G_b^{-1/2} S)_l^- \big\|^2 \|\Pi_K T (\Pi_{\mathcal{H}_0}^{\perp} h - \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h)\|_{L^2(W)}^2 \\
&\lesssim n^{-1} \big\| \langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle_{L^2(X)}' (G_b^{-1/2} S)_l^- \big\|^2,
\end{aligned}
$$

where the second bound is due to the Cauchy-Schwarz inequality and the third bound is due to Assumption 2(iv). Consider A_{12} we infer from Chen and Christensen [2018, Lemma

F.10(c)] and Assumption 2(ii) that

$$\begin{aligned}
|A_{12}|^2 \leq & \big\|\langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_l^- \big\|^2 \Big\| G_b^{-1/2} S' \Big((\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^- \Big) \Big\|^2 \\
& \times \Big\| \frac{1}{n} \sum_i \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X_i) b^K(W_i) - \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] \Big\|^2 \\
\lesssim & \big\|\langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_l^- \big\|^2 \times n^{-1} s_J^{-2} \zeta_J^2 (\log J) \times n^{-1} \zeta_J^2 \\
\lesssim & n^{-1} \big\|\langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_l^- \big\|^2
\end{aligned}$$

wpa1 uniformly for $h \in \mathcal{H}$. Next we consider the term A_2 of (E.1). Following the upper bound of A_{12} we obtain wpa1 uniformly for $h \in \mathcal{H}$:

$$\begin{aligned}
& \Big| \mathrm{E}[\Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)]' A' G (\widehat{A} - A) \, \mathrm{E}[(h - \Pi_{\mathcal{H}_0} h - \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h)(X) b^K(W)] \Big|^2 \\
& \leq \big\|\langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_l^- \big\|^2 \Big\| G_b^{-1/2} S \Big((\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} G_b^{1/2} - (G_b^{-1/2} S)_l^- \Big) \Big\|^2 \\
& \quad \times \big\| \langle T(\Pi_{\mathcal{H}_0}^{\perp} h - \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h), \widetilde{b}^K \rangle_{L^2(W)} \big\|^2 \\
& \lesssim \big\|\langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_l^- \big\|^2 \| \Pi_K T (\Pi_{\mathcal{H}_0}^{\perp} h - \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h) \|_{L^2(W)}^2 \times n^{-1} s_J^{-2} \zeta_J^2 (\log J) \\
& \lesssim n^{-1} \big\|\langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_l^- \big\|^2
\end{aligned}$$

using that $s_J^{-2} \| \Pi_K T (\Pi_{\mathcal{H}_0}^{\perp} h - \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h) \|_{L^2(W)}^2 \lesssim \| \Pi_{\mathcal{H}_0}^{\perp} h - \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h \|_{L^2(X)}^2$ by Assumption 2(iv) and $\zeta_J^2 (\log J) \| h - \Pi_J h \|_{L^2(X)}^2 = O(1)$ by Assumption 2(iii). Finally, we obtain $|T_1| \leq |A_1| + |A_2| \lesssim n^{-1/2} \| \langle Q_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle'_{L^2(X)} (G_b^{-1/2} S)_l^- \|$ wpa1 uniformly for $h \in \mathcal{H}$.

We next consider the term T_2 using the decomposition:

$$\begin{aligned}
T_2 \leq & \, 2 \, \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)]' (\widehat{A} - A)' G (\widehat{A} - A) \, \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] \\
& + 2 \, \mathrm{E}[\Pi_J^{\perp} \Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)]' (\widehat{A} - A)' G (\widehat{A} - A) \, \mathrm{E}[\Pi_J^{\perp} \Pi_{\mathcal{H}_0}^{\perp} h(X) b^K(W)] =: 2T_{21} + 2T_{22}
\end{aligned}$$

where $\Pi_J^{\perp} = \mathrm{id} - \Pi_J$ is the projection. We first bound T_{21} using Assumption 2(ii):

$$\begin{aligned}
T_{21} \leq & \Big| \langle \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle'_{L^2(X)} \Big((\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} S - I_J \Big)' (\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} \\
& \quad \times \Big(\frac{1}{n} \sum_i \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X_i) \widetilde{b}^K(W_i) - \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W)] \Big) \Big| \\
\leq & \big\| \langle \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h, \psi^J \rangle_{L^2(X)} \big\| \| S - \widehat{S} \| \Big\| (\widehat{G}_b^{-1/2} \widehat{S})_l^- \widehat{G}_b^{-1/2} \Big\|^2 \\
& \quad \times \Big\| \frac{1}{n} \sum_i \Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X_i) \widetilde{b}^K(W_i) - \mathrm{E}[\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W)] \Big\| \\
\lesssim & \| \Pi_J (h - \Pi_{\mathcal{H}_0} h) \|_{L^2(X)} n^{-1/2} s_J^{-2} \zeta_J \sqrt{\log J} \times n^{-1/2} s_J^{-1} \zeta_J \lesssim n^{-1/2} s_J^{-1} \| \Pi_J (h - \Pi_{\mathcal{H}_0} h) \|_{L^2(X)}
\end{aligned}$$

wpa1 uniformly for $h \in \mathcal{H}$. For T_{22}, we note that uniformly in $h \in \mathcal{H}$, $\| \mathrm{E}[\Pi_J^{\perp} \Pi_{\mathcal{H}_0}^{\perp} h(X) \widetilde{b}^K(W)] \| = \| \Pi_K T (\Pi_J \Pi_{\mathcal{H}_0}^{\perp} h - \Pi_{\mathcal{H}_0}^{\perp} h) \|_{L^2(W)} \lesssim s_J J^{-p/d_x}$ by Assumption 2(iv). Thus, following the upper bound derivations of T_{21}, we obtain $T_{22} \lesssim n^{-1/2} s_J^{-1} J^{-p/d_x}$ wpa1 uniformly for $h \in \mathcal{H}$. $\square$

Lemma E.2. *Under Assumptions 2(i) it holds for $\widetilde{h} \in \{h, \Pi_{\mathcal{H}_0} h\}$ that*

$$\sup_{J \in \mathcal{I}_n} \sup_{h \in \mathcal{H}} \lambda_{\max}\Big(\mathrm{E}_h \left[(Y - \widetilde{h}(X))^2 \widetilde{b}^{K(J)}(W) \widetilde{b}^{K(J)}(W)'\right]\Big) \leq \overline{\sigma}^2 < \infty.$$

Proof. We have for any $\gamma \in \mathbb{R}^K$ where $K = K(J)$ that

$$\begin{aligned}\gamma' \mathrm{E}_h \left[(Y - \widetilde{h}(X))^2 \widetilde{b}^K(W) \widetilde{b}^K(W)'\right]\gamma &\leq \mathrm{E}\left[\mathrm{E}_h[(Y - \widetilde{h}(X))^2|W]\big(\gamma'\widetilde{b}^K(W)\big)^2\right] \\ &\leq \overline{\sigma}^2 \,\mathrm{E}\left[\big(\gamma'\widetilde{b}^K(W)\big)^2\right] = \overline{\sigma}^2 \gamma' G_b^{-1/2}\, \mathrm{E}\left[b^K(W) b^K(W)'\right] G_b^{-1/2}\gamma = \overline{\sigma}^2 \|\gamma\|^2\end{aligned}$$

uniformly for $h \in \mathcal{H}$ and $J \in \mathcal{I}_n$, where the second inequality is due to Assumption 2(i). $\square$

Proof of Theorem B.1. From the definition of Q_J given in (B.1) we infer

$$\|Q_J(h - \Pi_{\mathcal{H}_0} h)\|^2_{L^2(X)} = \big\|A\, \mathrm{E}_h[(Y - \Pi_{\mathcal{H}_0} h(X)) b^K(W)]\big\|^2 = \|\,\mathrm{E}_h[V^J]\|^2$$

using the notation $V_i^J = (Y_i - \Pi_{\mathcal{H}_0} h(X_i)) A b^K(W_i)$. The definition of $\widehat{D}_J$ implies

$$\widehat{D}_J(\Pi_{\mathcal{H}_0} h) - \|Q_J(h - \Pi_{\mathcal{H}_0} h)\|^2_{L^2(X)} = \frac{1}{n(n-1)} \sum_{j=1}^J \sum_{i \neq i'} \big(V_{ij} V_{i'j} - \mathrm{E}_h[V_{1j}]^2\big) \tag{E.2}$$

$$+ \frac{1}{n(n-1)} \sum_{i \neq i'} \big(Y_i - \Pi_{\mathcal{H}_0} h(X_i)\big)\big(Y_{i'} - \Pi_{\mathcal{H}_0} h(X_{i'})\big) b^K(W_i)' \Big(A'A - \widehat{A}'\widehat{A}\Big) b^K(W_{i'}). \tag{E.3}$$

Consider the summand in (E.2), we observe

$$\Big|\sum_{j=1}^J \sum_{i \neq i'} \big(V_{ij} V_{i'j} - \mathrm{E}_h[V_{1j}]^2\big)\Big|^2 = \sum_{j,j'=1}^J \sum_{i \neq i'} \sum_{i'' \neq i'''} \big(V_{ij} V_{i'j} - \mathrm{E}_h[V_{1j}]^2\big)\big(V_{i''j'} V_{i'''j'} - \mathrm{E}_h[V_{1j'}]^2\big)$$

We distinguish three different cases. First: i, i', i'', i''' are all different, second: either $i = i''$ or $i' = i'''$, or third: $i = i'$ and $i' = i'''$. We thus calculate for each $j, j' \geq 1$ that

$$\begin{aligned}&\sum_{i \neq i'} \sum_{i'' \neq i'''} \big(V_{ij} V_{i'j} - \mathrm{E}_h[V_{1j}]^2\big)\big(V_{i''j'} V_{i'''j'} - \mathrm{E}_h[V_{1j'}]^2\big) \\ &\qquad = \sum_{i,i',i'',i''' \text{all different}} \big(V_{ij} V_{i'j} - \mathrm{E}_h[V_{1j}]^2\big)\big(V_{i''j'} V_{i'''j'} - \mathrm{E}_h[V_{1j'}]^2\big) \\ &\qquad\quad + 2 \sum_{i \neq i' \neq i''} \big(V_{ij} V_{i'j} - \mathrm{E}_h[V_{1j}]^2\big)\big(V_{i''j'} V_{i'j'} - \mathrm{E}_h[V_{1j'}]^2\big) \\ &\qquad\quad + \sum_{i \neq i'} \big(V_{ij} V_{i'j} - \mathrm{E}_h[V_{1j}]^2\big)\big(V_{ij'} V_{i'j'} - \mathrm{E}_h[V_{1j'}]^2\big).\end{aligned}$$

The expectation of the first term on the right hand side vanishes due to independent observations and thus, we have

$$
\begin{aligned}
&\mathrm{E}_h\Big|\sum_{j=1}^J\sum_{i\neq i'}\big(V_{ij}V_{i'j}-\mathrm{E}_h[V_{1j}]^2\big)\Big|^2\\
&=2n(n-1)(n-2)\underbrace{\sum_{j,j'=1}^J\mathrm{E}_h\Big[\big(V_{1j}V_{2j}-\mathrm{E}_h[V_{1j}]^2\big)\big(V_{3j'}V_{2j'}-\mathrm{E}_h[V_{1j'}]^2\big)\Big]}_{I}\\
&+n(n-1)\underbrace{\sum_{j,j'=1}^J\mathrm{E}_h\Big[\big(V_{1j}V_{2j}-\mathrm{E}_h[V_{1j}]^2\big)\big(V_{1j'}V_{2j'}-\mathrm{E}_h[V_{1j'}]^2\big)\Big]}_{II}.
\end{aligned}
$$

Now using $\|(G_b^{-1/2}SG^{-1/2})_l^-\|=s_J^{-1}$ together with the notation $\widetilde{\psi}^J=G^{-1/2}\psi^J$ we obtain

$$
\begin{aligned}
\big\|\langle Q_J(h-\Pi_{\mathcal{H}_0}h),\psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_l^-\big\|&=\big\|\langle Q_J(h-\Pi_{\mathcal{H}_0}h),\widetilde{\psi}^J\rangle'_{L^2(X)}(G_b^{-1/2}SG^{-1/2})_l^-\big\|\\
&\leq s_J^{-1}\big\|\langle Q_J(h-\Pi_{\mathcal{H}_0}h),\widetilde{\psi}^J\rangle_{L^2(X)}\big\|\lesssim s_J^{-1}\big(\|h-\Pi_{\mathcal{H}_0}h\|_{L^2(X)}+J^{-p/d_x}\big),
\end{aligned}\tag{E.4}
$$

where the last equation is due to Lemma B.1(i). To bound the term I we observe that

$$
\begin{aligned}
I&=\sum_{j,j'=1}^J\mathrm{E}_h[V_{1j}]\,\mathrm{E}_h[V_{1j'}]\,\mathbb{C}ov_h(V_{1j},V_{1j'})=\mathrm{E}_h[V_1^J]'\,\mathbb{C}ov_h(V_1^J,V_1^J)\,\mathrm{E}_h[V_1^J]\\
&\leq\lambda_{\max}\big(\mathbb{V}ar_h((Y-\Pi_{\mathcal{H}_0}h(X))\widetilde{b}^K(W))\big)\big\|(G_b^{-1/2}SG^{-1/2})_l^-\,\mathrm{E}_h[V_1^J]\big\|^2\\
&\leq\overline{\sigma}^2\Big\|\Big((G_b^{-1/2}S)_l^-\,\mathrm{E}_h[(Y-\Pi_{\mathcal{H}_0}h(X))\widetilde{b}^K(W)]\Big)'G(G_b^{-1/2}S)_l^-\Big\|^2\\
&=\overline{\sigma}^2\Big\|\langle Q_J(h-\Pi_{\mathcal{H}_0}h),\psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_l^-\Big\|^2\lesssim s_J^{-2}\big(\|h-\Pi_{\mathcal{H}_0}h\|^2_{L^2(X)}+J^{-2p/d_x}\big)
\end{aligned}
$$

by the notation $V_i^J=(Y_i-\Pi_{\mathcal{H}_0}h(X_i))(G_b^{-1/2}SG^{-1/2})_l^-\widetilde{b}^K(W_i)$ and Lemma E.2. For term II we observe

$$
II=\sum_{j,j'=1}^J\mathrm{E}_h[V_{1j}V_{1j'}]^2-\Big(\sum_{j=1}^J\mathrm{E}_h[V_{1j}]^2\Big)^2\leq\sum_{j,j'=1}^J\mathrm{E}_h[V_{1j}V_{1j'}]^2=\mathrm{v}_J^2.
$$

Thus, the upper bounds derived for the terms I and II imply for all $n\geq 2$:

$$
\mathrm{E}_h\Big|\frac{1}{n(n-1)}\sum_{j=1}^J\sum_{i\neq i'}\big(V_{ij}V_{i'j}-\mathrm{E}_h[V_{1j}]^2\big)\Big|^2\lesssim\frac{\|h-\Pi_{\mathcal{H}_0}h\|^2_{L^2(X)}+J^{-2p/d_x}}{ns_J^2}+\frac{\mathrm{v}_J^2}{n^2}.\tag{E.5}
$$

Thus equality (E.3) implies the result by employing Lemma B.2 and Lemma E.1. □

Proof of Lemma A.1. By Lemma E.1 and the decomposition (E.2–E.3) we obtain

$$\mathrm{P}_{h_0}\left(\frac{n\widehat{D}_J(h_0)}{\mathrm{v}_J} > \eta_J(\alpha)\right) = \mathrm{P}_{h_0}\left(\frac{1}{\mathrm{v}_J(n-1)}\sum_{j=1}^{J}\sum_{i\neq i'} V_{ij}V_{i'j} > \eta_J(\alpha)\right) + o(1).$$

Using the martingale central limit theorem (see, e.g., Breunig [2020, Lemma A.3]) we obtain

$$\mathrm{P}_{h_0}\left(\frac{1}{\sqrt{2}\mathrm{v}_J(n-1)}\sum_{j=1}^{J}\sum_{i\neq i'} V_{ij}V_{i'j} > z_{1-\alpha}\right) = \alpha + o(1),$$

where $z_{1-\alpha}$ denotes the $(1-\alpha)$-quantile of the standard normal distribution. Further, Lemma B.4(i) implies $\mathrm{v}_J/\widehat{\mathrm{v}}_J = 1$ wpa1 uniformly for $h \in \mathcal{H}$ and since $\eta_J(\alpha)/\sqrt{2} = \frac{q(\alpha,J)-J}{\sqrt{2J}}$ converges to $z_{1-\alpha}$ as J tends to infinity, the result follows. □

Proof of Lemma B.1. Proof of (i): Using the notation $\widetilde{b}^K(\cdot) := G_b^{-1/2} b^K(\cdot)$, we observe for all $h \in \mathcal{H}$ that

$$\begin{aligned}
\|Q_J(h - \Pi_{\mathcal{H}_0}h)\|_{L^2(X)} &= \left\|(G_b^{-1/2}SG^{-1/2})_l^- \,\mathrm{E}[\widetilde{b}^K(W)(h - \Pi_{\mathcal{H}_0}h)(X)]\right\| \\
&\le \left\|(G_b^{-1/2}SG^{-1/2})_l^- \,\mathrm{E}[\widetilde{b}^K(W)(\Pi_J h - \Pi_J\Pi_{\mathcal{H}_0}h)(X)]\right\| \\
&\quad + \left\|(G_b^{-1/2}SG^{-1/2})_l^- \,\mathrm{E}[\widetilde{b}^K(W)((h - \Pi_{\mathcal{H}_0}h)(X) - (\Pi_J h - \Pi_J\Pi_{\mathcal{H}_0}h)(X))]\right\| \\
&\le \|\Pi_J h - \Pi_J\Pi_{\mathcal{H}_0}h\|_{L^2(X)} + s_J^{-1}\|\Pi_K T((h - \Pi_{\mathcal{H}_0}h) - (\Pi_J h - \Pi_J\Pi_{\mathcal{H}_0}h))\|_{L^2(W)} \\
&\le \|\Pi_J h - \Pi_J\Pi_{\mathcal{H}_0}h\|_{L^2(X)} + O\big(J^{-p/d_x}\big)
\end{aligned}$$

by Assumption 2(iv).

Proof of (ii): We observe $\|Q_J h - h\|_{L^2(X)} \le \|Q_J(h - \Pi_J h)\|_{L^2(X)} + \|\Pi_J h - h\|_{L^2(X)}$. The result thus follows by replacing $\Pi_{\mathcal{H}_0}h$ with $\Pi_J h$ in the derivation of (i). □

Proof of Lemma B.2. For any $J \times J$ matrix M it holds $\|M\|_F \le \sqrt{J}\|M\|$ and hence

$$\begin{aligned}
\mathrm{v}_J^2 &= \left\|\left(G_b^{-1/2}SG^{-1/2}\right)_l^- \mathrm{E}_h\left[(Y - h(X))^2\widetilde{b}^K(W)\widetilde{b}^K(W)'\right]\left(G_b^{-1/2}SG^{-1/2}\right)_l^-\right\|_F^2 \\
&\le J\left\|\left(G_b^{-1/2}SG^{-1/2}\right)_l^-\right\|^4 \left\|\mathrm{E}_h\left[(Y - h(X))^2\widetilde{b}^K(W)\widetilde{b}^K(W)'\right]\right\|^2.
\end{aligned}$$

The result now follows from $\|(G_b^{-1/2}SG^{-1/2})_l^-\| = s_J^{-1}$ and Lemma E.2. □

Proof of Lemma B.3. In the following, let e_j be the unit vector with 1 at the j–th position. Introduce a unitary matrix Q such that by Schur decomposition $Q'AG_bA'Q = \mathrm{diag}(s_1^{-2}, \dots, s_J^{-2})$. We make use of the notation $\widetilde{V}_i^J = (Y_i - h(X_i))Q'Ab^K(W_i)$. Now since the Frobenius norm is invariant under unitary matrix multiplication we have

$$v_J^2 = \sum_{j,j'=1}^J \mathrm{E}_h[\widetilde{V}_{1j}\widetilde{V}_{1j'}]^2 \geq \sum_{j=1}^J \mathrm{E}_h[\widetilde{V}_{1j}^2]^2 = \sum_{j=1}^J \big(\mathrm{E}_h\,|(Y-h(X))e_j'Q'Ab^K(W)|^2\big)^2.$$

Consequently, using the lower bound $\inf_{w\in\mathcal{W}}\inf_{h\in\mathcal{H}}\mathrm{E}_h[(Y-h(X))^2|W=w] \geq \underline{\sigma}^2$ by Assumption 1(i), we obtain uniformly for $h \in \mathcal{H}$:

$$\begin{aligned}
v_J^2 &\geq \underline{\sigma}^4 \sum_{j=1}^J \big(\mathrm{E}[e_j'Q'Ab^K(W)b^K(W)'A'Qe_j]\big)^2 = \underline{\sigma}^4 \sum_{j=1}^J \big(e_j'Q'AG_bA'Qe_j\big)^2 \\
&= \underline{\sigma}^4 \sum_{j=1}^J \big(e_j'\,\mathrm{diag}(s_1^{-2},\dots,s_J^{-2})e_j\big)^2 \geq \underline{\sigma}^4 \sum_{j=1}^J s_j^{-4},
\end{aligned}$$

which proves the result. □

Recall the definition $\mathcal{C}_h = \max_{\mathsf{e}\in\mathcal{S}^{K^\circ}} \int_0^1 \big(1+\log N_{[]}\big(\epsilon\|F_{h,\mathsf{e}}\|_{L^2(Z)}, \mathcal{F}_{h,\mathsf{e}}, L^2(Z)\big)\big)^{1/2} d\epsilon$.

Lemma E.3. *Let Assumptions 1(ii)-(iii), 2(i), 4(i)(iii), and 5(ii) hold. Then, for $J = J^\circ$, we have wpa1 uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:*

$$\left|\frac{1}{n(n-1)}\sum_{i\neq i'} U_i(\widehat{h}_J^{\mathrm{R}})U_{i'}(\widehat{h}_J^{\mathrm{R}})a_{J,ii'} - \mathrm{E}_h\left[U_i(\widehat{h}_J^{\mathrm{R}})U_{i'}(\widehat{h}_J^{\mathrm{R}})a_{J,ii'}\right]\right| \\ \lesssim n^{-1/2}s_J^{-1}\mathcal{C}_h\big(\|h-\mathcal{H}_0\|_{L^2(X)} + J^{-p/d_x}\big) + n^{-1}s_J^{-2}\sqrt{J},$$

where $U_i(\phi) = Y_i - \phi(X_i)$ and $a_{J,ii'} = b^K(W_i)'A'Ab^K(W_{i'})$.

Proof. For simplicity of notation, we write J instead of J° throughout the proof. We observe for all $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ that

$$\begin{aligned}
&\frac{1}{n(n-1)}\sum_{i\neq i'} U_i(\widehat{h}_J^{\mathrm{R}})U_{i'}(\widehat{h}_J^{\mathrm{R}})a_{J,ii'} - \mathrm{E}_h\left[U_i(\widehat{h}_J^{\mathrm{R}})U_{i'}(\widehat{h}_J^{\mathrm{R}})a_{J,ii'}\right] \\
&= \frac{1}{n(n-1)}\sum_{i\neq i'} U_i(\Pi_{\mathcal{H}_0}h)U_{i'}(\Pi_{\mathcal{H}_0}h)a_{J,ii'} - \mathrm{E}_h\left[U_i(\Pi_{\mathcal{H}_0}h)U_{i'}(\Pi_{\mathcal{H}_0}h)a_{J,ii'}\right] \\
&+ \frac{2}{n(n-1)}\sum_{i\neq i'} U_i(\Pi_{\mathcal{H}_0}h)(\Pi_{\mathcal{H}_0}h-\widehat{h}_J^{\mathrm{R}})(X_{i'})a_{J,ii'} - \mathrm{E}_h\left[U_i(\Pi_{\mathcal{H}_0}h)(\Pi_{\mathcal{H}_0}h-\widehat{h}_J^{\mathrm{R}})(X_{i'})a_{J,ii'}\right] \\
&+ \frac{1}{n(n-1)}\sum_{i\neq i'} (\Pi_{\mathcal{H}_0}h-\widehat{h}_J^{\mathrm{R}})(X_i)(\Pi_{\mathcal{H}_0}h-\widehat{h}_J^{\mathrm{R}})(X_{i'})a_{J,ii'} - \mathrm{E}_h\left[(\Pi_{\mathcal{H}_0}h-\widehat{h}_J^{\mathrm{R}})(X_i)(\Pi_{\mathcal{H}_0}h-\widehat{h}_J^{\mathrm{R}})(X_{i'})a_{J,ii'}\right] \\
&=: T_1 + 2\,T_2 + T_3.
\end{aligned}$$

From the proof of Theorem B.1 we conclude $\sup_{h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)}\mathrm{E}_h|T_1| \lesssim n^{-1}s_J^{-2}\sqrt{J}$. Consider T_2. Below, we let $a_i^J = Ab^K(W_i) = (G_b^{-1/2}SG^{-1/2})_\ell^{-}\widetilde{b}^K(W_i)$. By Assumption 5(ii) $\sup_{h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)}\mathrm{P}_h\big(\zeta_J\mathcal{C}_h\|\widehat{h}_J^{\mathrm{R}} - \Pi_{\mathcal{H}_0}h\|_{L^2(X)} > C\big) \to 0$ and consequently may assume that

$\widehat{h}_J^{\mathrm{R}} \in \mathcal{H}_{0,J}(h) := \{\|\phi - \Pi_{\mathcal{H}_0} h\|_{L^2(X)} \leq [\zeta_J \mathcal{C}_h]^{-1} : \phi \in \mathcal{H}_{0,J}\}$. We have for all $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ that the absolute value of T_2 is bounded by

$$\sup_{\phi\in\mathcal{H}_{0,J}(h)} \Big|\frac{1}{n(n-1)}\sum_{i\neq i'}\Big(U_i(\Pi_{\mathcal{H}_0}h)a_i^J - \mathrm{E}_h[U(\Pi_{\mathcal{H}_0}h)a^J]\Big)'\Big((\Pi_{\mathcal{H}_0}h-\phi)(X_{i'})a_{i'}^J - \mathrm{E}\left[(\Pi_{\mathcal{H}_0}h-\phi)(X)a^J\right]\Big)\Big|$$
$$+\Big|\frac{1}{n}\sum_i\Big(U_i(\Pi_{\mathcal{H}_0}h)a_i^J - \mathrm{E}_h[U(\Pi_{\mathcal{H}_0}h)a^J]\Big)'\,\mathrm{E}\left[(\Pi_{\mathcal{H}_0}h-\widehat{h}_J^{\mathrm{R}})(X)a^J\right]\Big)\Big|$$
$$+\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\frac{1}{n}\sum_i\Big((\Pi_{\mathcal{H}_0}h-\phi)(X_i)a_i^J - \mathrm{E}\left[(\Pi_{\mathcal{H}_0}h-\phi)(X)a^J\right]\Big)'\,\mathrm{E}_h[U(\Pi_{\mathcal{H}_0}h)a^J]\Big| =: T_{21}+T_{22}+T_{23}.$$

Below we let $a_{k,i} = b^K(W_i)'A'AG_b^{1/2}e_k$. Note that $\mathrm{E}\,\|a_{k,i}\|^2 \leq \|(G_b^{-1/2}SG^{-1/2})_\ell^-\|^4 = s_J^{-4}$ for all $k = 1,\dots,K$. We obtain uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ by van der Vaart and Wellner [2000, Theorem 2.14.2] that

$$\begin{aligned}
\mathrm{E}_h\,T_{21} &\leq \sum_{k=1}^K \mathrm{E}_h\Big|\frac{1}{n}\sum_i U_i(\Pi_{\mathcal{H}_0}h)a_{k,i} - \mathrm{E}_h[U(\Pi_{\mathcal{H}_0}h)a_k]\Big| \\
&\qquad\times \mathrm{E}_h \sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\frac{1}{n-1}\sum_{i'}(\Pi_{\mathcal{H}_0}h-\phi)(X_{i'})\widetilde{b}_k(W_{i'}) - \mathrm{E}_h\left[(\Pi_{\mathcal{H}_0}h-\phi)(X)\widetilde{b}_k(W)\right]\Big| \\
&\lesssim \frac{\mathcal{C}_h}{n}\sqrt{\sum_{k=1}^K \mathrm{E}_h\left[|U_i(\Pi_{\mathcal{H}_0}h)|^2\|G_b^{-1/2}A'Ab^K(W)\|^2\right]}\sqrt{\sum_{k=1}^K \mathrm{E}_h\sup_{\phi\in\mathcal{H}_{0,J}(h)}|(\Pi_{\mathcal{H}_0}h-\phi)(X)\widetilde{b}_k(W)|^2} \\
&\lesssim \frac{\mathcal{C}_h}{n}\overline{\sigma}s_J^{-2}\sqrt{J}\zeta_J\|\Pi_{\mathcal{H}_0}h - \Phi_J\|_{L^2(X)} \lesssim n^{-1}s_J^{-2}\sqrt{J}
\end{aligned}$$

for some $\Phi_J \in \mathcal{H}_{0,J}(h)$ and using that $\mathrm{E}_h[|U(\Pi_{\mathcal{H}_0}h)|^2|W] \leq \overline{\sigma}^2$ by Assumption 2(i). Further, we evaluate uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:

$$\begin{aligned}
\mathrm{E}_h\,T_{22} &= \overline{\sigma}n^{-1/2}\sqrt{\mathrm{E}\left|(a^J)'\,\mathrm{E}_h\left[(\Pi_{\mathcal{H}_0}h-\widehat{h}_J^{\mathrm{R}})(X)a^J\right]\right|^2} \\
&\leq \overline{\sigma}n^{-1/2}s_J^{-2}\sup_{\phi\in\mathcal{H}_{0,J}(h)}\|\Pi_K T(\Pi_{\mathcal{H}_0}h-\phi)\|_{L^2(W)} \lesssim n^{-1/2}s_J^{-1}\big(\|h-\mathcal{H}_0\|_{L^2(X)} + J^{-p/d_x}\big),
\end{aligned}$$

where in the last equation, we used Assumption 2(iv) and $\|h - \Pi_{\mathcal{H}_0}h\|_{L^2(X)} = \|h - \mathcal{H}_0\|_{L^2(X)}$. Consider T_{23}. Below, we make use of the relation $\mathrm{E}[U(\Pi_{\mathcal{H}_0}h)a^J]'a_i^J = \langle Q_J(h - \Pi_{\mathcal{H}_0}h), \psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_\ell^-\widetilde{b}^K(W_i)$ and obtain uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:

$$\begin{aligned}
\mathrm{E}_h\,T_{23} &\leq \big\|\langle Q_J(h-\Pi_{\mathcal{H}_0}h),\psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_\ell^-\big\| \\
&\qquad\times \mathrm{E}_h\sup_{\mathsf{e}\in\mathcal{S}^{K^\circ-1}}\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\frac{1}{n}\sum_i(\Pi_{\mathcal{H}_0}h-\phi)(X_i)\widetilde{b}^K(W_i)'\mathsf{e} - \mathrm{E}\left[(\Pi_{\mathcal{H}_0}h-\phi)(X)\widetilde{b}^K(W)'\mathsf{e}\right]\Big| \\
&\lesssim \big\|\langle Q_J(h-\Pi_{\mathcal{H}_0}h),\psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_\ell^-\big\| \times \mathcal{C}_h n^{-1/2}\zeta_J\|\Pi_{\mathcal{H}_0}h-\Phi_J\|_{L^2(X)} \\
&\lesssim \mathcal{C}_h n^{-1/2}s_J^{-1}\big(\|h-\mathcal{H}_0\|_{L^2(X)} + J^{-p/d_x}\big),
\end{aligned}$$

where we used that $\sup_w |\widetilde{b}^K(w)'\mathsf{e}| \leq \zeta_J$ for all $\mathsf{e} \in \mathcal{S}^{K^\circ}$. Consider T_3. We have

$$
\begin{aligned}
|T_3| \leq & \sup_{\phi\in\mathcal{H}_{0,J}(h)} \Big|\frac{1}{n(n-1)}\sum_{i\neq i'}\Big((\Pi_{\mathcal{H}_0}h-\phi)(X_i)a_i^J - \mathrm{E}\left[(\Pi_{\mathcal{H}_0}h-\phi)(X)a^J\right]\Big)' \\
&\qquad\qquad \times \Big((\Pi_{\mathcal{H}_0}h-\phi)(X_{i'})a_{i'}^J - \mathrm{E}\left[(\Pi_{\mathcal{H}_0}h-\phi)(X)a^J\right]\Big)\Big| \\
& + 2\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\frac{1}{n}\sum_i\Big((\Pi_{\mathcal{H}_0}h-\phi)(X_i)a_i^J - \mathrm{E}\left[(\Pi_{\mathcal{H}_0}h-\phi)(X)a^J\right]\Big)'\mathrm{E}\left[(\Pi_{\mathcal{H}_0}h-\phi)(X)a^J\right]\Big| \\
& =: T_{31} + T_{32}.
\end{aligned}
$$

We evaluate for the first term on the right hand side that uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:

$$
\begin{aligned}
\mathrm{E}\,T_{31} &\leq s_J^{-2}\sum_{k=1}^K\Big(\mathrm{E}\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\frac{1}{n}\sum_i(\Pi_{\mathcal{H}_0}h-\phi)(X_i)\widetilde{b}_k(W_i) - \mathrm{E}\left[(\Pi_{\mathcal{H}_0}h-\phi)(X)\widetilde{b}_k(W)\right]\Big|\Big)^2 \\
&\lesssim \frac{\mathcal{C}_h^2}{ns_J^2}\mathrm{E}\sup_{\phi\in\mathcal{H}_{0,J}(h)}\|(\Pi_{\mathcal{H}_0}h-\phi)(X)\widetilde{b}^K(W)\|^2 \lesssim \frac{\mathcal{C}_h^2}{ns_J^2}\zeta_J^2\|\Pi_{\mathcal{H}_0}h-\Phi_J\|_{L^2(X)}^2 \lesssim \frac{\sqrt{J}}{ns_J^2},
\end{aligned}
$$

for some $\Phi_J \in \mathcal{H}_{0,J}(h)$ and using that $\mathcal{C}_h^2 \lesssim \sqrt{J}$. Further, we have $\mathrm{E}[(\Pi_{\mathcal{H}_0}h-\phi)(X)a^J]'a_i^J = \langle Q_J(\Pi_{\mathcal{H}_0}h-\phi), \psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_\ell^-\widetilde{b}^K(W_i)$ and thus, following the derivation of the bound of T_{23}, we obtain

$$
\begin{aligned}
\mathrm{E}\,T_{32} \leq & \sup_{\phi\in\mathcal{H}_{0,J}(h)}\big\|\langle Q_J(\phi-\Pi_{\mathcal{H}_0}h), \psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_\ell^-\big\| \\
& \times \mathrm{E}\sup_{\mathsf{e}\in\mathcal{S}^{K^\circ}}\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\frac{1}{n}\sum_i(\phi-\Pi_{\mathcal{H}_0}h)(X_i)\widetilde{b}^K(W_i)'\mathsf{e} - \mathrm{E}\left[(\phi-\Pi_{\mathcal{H}_0}h)(X)\widetilde{b}^K(W)'\mathsf{e}\right]\Big| \\
\lesssim & \; n^{-1/2}s_J^{-1}\mathcal{C}_h\big(\|h-\Pi_{\mathcal{H}_0}h\|_{L^2(X)} + J^{-p/d_x}\big)
\end{aligned}
$$

uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$, where the last equation is due to Assumption 5(ii). Finally, the result follows from an application of Markov's inequality. □

Lemma E.4. *Let Assumptions 1(ii)-(iii), 2(i), 4(i)(iii), and 5(ii) hold. Then, for $J = J^\circ$, we have wpa1 uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:*

$$
\begin{aligned}
\frac{1}{n(n-1)}\sum_{i\neq i'}\big(Y_i - \widehat{h}_J^{\mathrm{R}}(X_i)\big)\big(Y_{i'} - \widehat{h}_J^{\mathrm{R}}(X_{i'})\big)b^K(W_i)'\Big(A'A - \widehat{A}'\widehat{A}\Big)b^K(W_{i'}) \\
\lesssim n^{-1/2}s_J^{-1}\mathcal{C}_h\big(\|h-\mathcal{H}_0\|_{L^2(X)} + J^{-p/d_x}\big) + n^{-1}s_J^{-2}\sqrt{J}.
\end{aligned}
$$

Proof. For simplicity of notation, we write J instead of J° throughout the proof. Following the proof of Lemma E.1 it is sufficient to control

$$
\begin{aligned}
&\mathrm{E}_h[(h-\widehat{h}^{\mathrm{R}}_J)(X)b^K(W)]'\Big(A'A-\widehat{A}'\widehat{A}\Big)\,\mathrm{E}_h[(h-\widehat{h}^{\mathrm{R}}_J)(X)b^K(W)]\\
&\quad=2\,\mathrm{E}_h[(h-\widehat{h}^{\mathrm{R}}_J)(X)b^K(W)]'A'(A-\widehat{A})\,\mathrm{E}_h[(h-\widehat{h}^{\mathrm{R}}_J)(X)b^K(W)]\\
&\qquad-\mathrm{E}_h[(h-\widehat{h}^{\mathrm{R}}_J)(X)b^K(W)]'(A-\widehat{A})'(A-\widehat{A})\,\mathrm{E}_h[(h-\widehat{h}^{\mathrm{R}}_J)(X)b^K(W)]=:2T_1-T_2,
\end{aligned}
$$

We first consider the term T_1 using the decomposition:

$$
\begin{aligned}
T_1&=\mathrm{E}_h[(h-\widehat{h}^{\mathrm{R}}_J)(X)b^K(W)]'A'(\widehat{A}-A)\,\mathrm{E}_h[\Pi_J(h-\widehat{h}^{\mathrm{R}}_J)(X)b^K(W)]\\
&\quad+\mathrm{E}_h[(h-\widehat{h}^{\mathrm{R}}_J)(X)b^K(W)]'A'(\widehat{A}-A)\,\mathrm{E}_h[(h-\widehat{h}^{\mathrm{R}}_J-\Pi_J(h-\widehat{h}^{\mathrm{R}}_J))(X)b^K(W)].
\end{aligned}
\tag{E.6}
$$

Consider the first summand on the right hand side of equation (E.6). By Assumption 5(ii) $\sup_{h\in\mathcal{H}_1(\delta^\circ \mathrm{r}_n)}\mathrm{P}_h\big(\zeta_J\mathcal{C}_h\|\widehat{h}^{\mathrm{R}}_J-\Pi_{\mathcal{H}_0}h\|_{L^2(X)}>C\big)\to 0$ and consequently may assume that $\widehat{h}^{\mathrm{R}}_J\in\mathcal{H}_{0,J}(h):=\{\phi\in\mathcal{H}_{0,J}:\|\phi-\Pi_{\mathcal{H}_0}h\|_{L^2(X)}\le[\zeta_J\mathcal{C}_h]^{-1}\}$. We calculate

$$
\begin{aligned}
&\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\big((G_b^{-1/2}S)_l^-\,\mathrm{E}[(h-\phi)(X)\widetilde{b}^K(W)]\big)'G\big((G_b^{-1/2}S)_l^- -(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}\big)\,\mathrm{E}[(h-\phi)(X)\widetilde{b}^K(W)]\Big|\\
&=\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\langle Q_J(h-\phi),\psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_l^-\Big(\frac{1}{n}\sum_i\Pi_J(h-\phi)(X_i)\widetilde{b}^K(W_i)-\mathrm{E}[\Pi_J(h-\phi)(X)\widetilde{b}^K(W)]\Big)\Big|\\
&\quad+\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\langle Q_J(h-\phi),\psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_l^-G_b^{-1/2}S\Big((\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-\Big)\\
&\qquad\qquad\times\Big(\frac{1}{n}\sum_i\Pi_J(h-\phi)(X_i)\widetilde{b}^K(W_i)-\mathrm{E}[\Pi_J(h-\phi)(X)\widetilde{b}^K(W)]\Big)\Big|=:T_{11}+T_{12}.
\end{aligned}
$$

Consider T_{11}, which coincides with the term T_{32} in the proof of Lemma E.3 and thus, we have $\mathrm{E}\,|T_{11}|\lesssim n^{-1/2}s_J^{-1}\mathcal{C}_h\big(\|h-\mathcal{H}_0\|_{L^2(X)}+J^{-p/d_x}\big)$. To establish an upper bound for T_{12} we infer from Chen and Christensen [2018, Lemma F.10(c)] that

$$
\begin{aligned}
|T_{12}|^2&\le\sup_{\phi\in\mathcal{H}_{0,J}(h)}\big\|\langle Q_J(h-\phi),\psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_l^-\big\|^2\\
&\quad\times\Big\|G_b^{-1/2}S\Big((\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-\Big)\Big\|^2\\
&\quad\times\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big\|\frac{1}{n}\sum_i\Pi_J(h-\phi)(X_i)b^K(W_i)-\mathrm{E}[\Pi_J(h-\phi)(X)b^K(W)]\Big\|^2\\
&\lesssim\sup_{\phi\in\mathcal{H}_{0,J}(h)}\big\|\langle Q_J(h-\phi),\psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_l^-\big\|^2\times n^{-1}s_J^{-2}\zeta_J^2(\log J)\times n^{-1}\zeta_J^2\mathcal{C}_h^2\\
&\lesssim n^{-1}s_J^{-2}\mathcal{C}_h^2\big(\|h-\Pi_{\mathcal{H}_0}h\|^2_{L^2(X)}+J^{-2p/d_x}\big)
\end{aligned}
$$

wpa1 uniformly for $h\in\mathcal{H}_1(\delta^\circ \mathrm{r}_n)$, where the last equation is due to $s_J^{-1}\zeta_J^2\sqrt{(\log J)/n}=O(1)$ from Assumption 4(i). Consider the second summand on the right hand side of equation

(E.6). Following the upper bound of T_{12} we obtain

$$\begin{aligned}
&\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\,\mathrm{E}[(h-\phi)(X)b^K(W)]'A'G(\widehat{A}-A)\,\mathrm{E}[(h-\phi-\Pi_J(h-\phi))(X)b^K(W)]\Big|^2\\
&\leq \sup_{\phi\in\mathcal{H}_{0,J}(h)}\big\|\langle Q_J(h-\phi),\psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_l^-\big\|^2\Big\|G_b^{-1/2}S'\Big((\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(G_b^{-1/2}S)_l^-\Big)\Big\|^2\\
&\qquad\times\sup_{\phi\in\mathcal{H}_{0,J}(h)}\big\|\langle T(h-\phi-\Pi_J(h-\phi)),\widetilde{b}^K\rangle_{L^2(W)}\big\|^2\\
&\lesssim \sup_{\phi\in\mathcal{H}_{0,J}(h)}\big\|\langle Q_J(h-\phi),\psi^J\rangle'_{L^2(X)}(G_b^{-1/2}S)_l^-\big\|^2\sup_{\phi\in\mathcal{H}_{0,J}(h)}\|\Pi_K T(h-\phi-\Pi_J(h-\phi))\|^2_{L^2(W)}\\
&\qquad\times n^{-1}s_J^{-2}\zeta_J^2(\log J)\\
&\lesssim n^{-1}s_J^{-2}\big(\|h-\Pi_{\mathcal{H}_0}h\|^2_{L^2(X)}+J^{-2p/d_x}\big)
\end{aligned}$$

wpa1 uniformly for $h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)$, using that $s_J^{-2}\|\Pi_K T(h-\Pi_{\mathcal{H}_0}h-\Pi_J(h-\Pi_{\mathcal{H}_0}h))\|^2_{L^2(W)}\lesssim\|h-\Pi_{\mathcal{H}_0}h-\Pi_J(h-\Pi_{\mathcal{H}_0}h)\|^2_{L^2(X)}$ by Assumption 4(i) and $\zeta_J^2(\log J)\|h-\Pi_J h\|^2_{L^2(X)}=O(1)$ by Assumption 4(iii).

We now consider the term T_2 using the decomposition

$$\begin{aligned}
T_2 &\leq 2\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\,\mathrm{E}[\Pi_J(h-\phi)(X)b^K(W)]'(\widehat{A}-A)'G(\widehat{A}-A)\,\mathrm{E}[\Pi_J(h-\phi)(X)b^K(W)]\Big|\\
&\quad+2\sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\,\mathrm{E}[\Pi_J^\perp(h-\phi)(X)b^K(W)]'(\widehat{A}-A)'G(\widehat{A}-A)\,\mathrm{E}[\Pi_J^\perp(h-\phi)(X)b^K(W)]\Big|\\
&=: 2T_{21}+2T_{22},
\end{aligned}$$

where $\Pi_J^\perp=\mathrm{id}-\Pi_J$ is the projection. We bound T_{21} as follows:

$$\begin{aligned}
T_{21} &\leq \sup_{\phi\in\mathcal{H}_{0,J}(h)}\Big|\langle\Pi_J(h-\phi),\psi^J\rangle'_{L^2(X)}\Big((\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}S-I_J\Big)'(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}\\
&\qquad\qquad\times\Big(\frac{1}{n}\sum_i\Pi_J(h-\phi)(X_i)\widetilde{b}^K(W_i)-\mathrm{E}[\Pi_J(h-\phi)(X)\widetilde{b}^K(W)]\Big)\Big|\\
&\leq \sup_{\phi\in\mathcal{H}_{0,J}(h)}\big\|\langle\Pi_J(h-\phi),\psi^J\rangle_{L^2(X)}\big\|\|S-\widehat{S}\|\Big\|(\widehat{G}_b^{-1/2}\widehat{S})_l^-\widehat{G}_b^{-1/2}\Big\|^2\\
&\qquad\qquad\times\Big\|\frac{1}{n}\sum_i\Pi_J(h-\phi)(X_i)\widetilde{b}^K(W_i)-\mathrm{E}[\Pi_J(h-\phi)(X)\widetilde{b}^K(W)]\Big\|\\
&\lesssim \sup_{\phi\in\mathcal{H}_{0,J}(h)}\|\Pi_J(h-\phi)\|_{L^2(X)}\times n^{-1/2}s_J^{-2}\zeta_J\sqrt{\log J}\times n^{-1/2}\zeta_J\mathcal{C}_h\\
&\lesssim n^{-1/2}s_J^{-1}\mathcal{C}_h\big(\|h-\Pi_{\mathcal{H}_0}h\|_{L^2(X)}+J^{-p/d_x}\big).
\end{aligned}$$

wpa1 uniformly for $h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)$. For T_{22}, we note that uniformly in $h\in\mathcal{H}$ and $\phi\in\mathcal{H}_{0,J}(h)$, $\|\,\mathrm{E}[\Pi_J^\perp(h-\phi)(X)\widetilde{b}^K(W)]\|=\|\Pi_K T\Pi_J^\perp(h-\phi)\|_{L^2(W)}\lesssim s_J J^{-p/d_x}$ by Assumption 2(iv). Thus, following the upper bound derivations of T_{21}, we obtain $T_{22}\lesssim n^{-1/2}s_J^{-1}J^{-p/d_x}$ wpa1 uniformly for $h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)$. □

Lemma E.5. *Let Assumptions 1(i)-(iii), 2(i), and 4 be satisfied. Then, using the notation* $S^o := G_b^{-1/2} S G^{-1/2}$*, we have for some constant* $C > 0$*:*

$$(i)\ \mathrm{P}\left(\max_{J\in\mathcal{I}_n}\left\{\frac{s_J^2\sqrt{n}}{\zeta_J\sqrt{\log J}}\left\|(\widehat{G}_b^{-1/2}\widehat{S}\widehat{G}^{-1/2})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(S^o)_l^-\right\|\right\}>C\right)=o(1),$$
$$(ii)\ \mathrm{P}\left(\max_{J\in\mathcal{I}_n}\left\{\frac{s_J^2\sqrt{n}}{\zeta_J\sqrt{\log J}}\left\|S^o\Big((\widehat{G}_b^{-1/2}\widehat{S}\widehat{G}^{-1/2})_l^-\widehat{G}_b^{-1/2}G_b^{1/2}-(S^o)_l^-\Big)\right\|\right\}>C\right)=o(1).$$

Proof. The results can be established by following the same proof from Chen et al. [2021, Lemma C.4] with their $(\tau_J, \sqrt{J})$ replaced by our (s_J^{-1}, ζ_J). □

Lemma E.6. *Let Assumptions 1(i)-(iii), 2(i) and 4(i) hold. Then, we have*

$$\mathrm{P}_h\Big(\max_{J\in\mathcal{I}_n}\Big|\frac{(\log\log J)^{-1/2}}{(n-1)\mathrm{v}_J}\sum_{i\neq i'}U_i(\Pi_{\mathcal{H}_0}h)U_{i'}(\Pi_{\mathcal{H}_0}h)b^K(W_i)'\big(A'A-\widehat{A}'\widehat{A}\big)b^K(W_{i'})\Big|>\frac{1-c_0}{8}\Big)=o(1)$$

uniformly for $h\in\mathcal{H}_0$*, where* $U_i(\phi)=Y_i-\phi(X_i)$ *and* c_0 *is as in the proof of Theorem 4.1.*

Proof. Let I_{s_J} denote the J dimensional identity matrix multiplied by the vector $C_0(s_1,\dots,s_J)'$ for some sufficiently large constant C_0 and where s_j^{-1}, $1\le j\le J$, are the nondecreasing singular values of $AG_b^{1/2}=(G_b^{-1/2}SG^{-1/2})_l^-$. There exists a unitary matrix Q such that

$$\begin{aligned}
&\sum_{i\neq i'}U_i(\Pi_{\mathcal{H}_0}h)U_{i'}(\Pi_{\mathcal{H}_0}h)b^K(W_i)'\big(A'A-\widehat{A}'\widehat{A}\big)b^K(W_{i'})\\
&\le\Big\|\sum_i U_i(\Pi_{\mathcal{H}_0}h)\widetilde{b}^K(W_i)'QI_{s_J}^{-1}\Big\|^2\big\|I_{s_J}Q'G_b^{1/2}(A'A-\widehat{A}'\widehat{A})G_b^{1/2}QI_{s_J}\big\|\\
&=\sum_{i\neq i'}U_i(\Pi_{\mathcal{H}_0}h)U_{i'}(\Pi_{\mathcal{H}_0}h)\widetilde{b}^K(W_i)'QI_{s_J}^{-2}Q'\widetilde{b}^K(W_{i'})\big\|I_{s_J}Q'G_b^{1/2}(A'A-\widehat{A}'\widehat{A})G_b^{1/2}QI_{s_J}\big\|\\
&\quad+\sum_i\Big\|U_i(\Pi_{\mathcal{H}_0}h)\widetilde{b}^K(W_i)QI_{s_J}^{-1}\Big\|^2\big\|I_{s_J}Q'G_b^{1/2}(A'A-\widehat{A}'\widehat{A})G_b^{1/2}QI_{s_J}\big\|.
\end{aligned}$$

The fourth moment condition imposed in Assumption 2(i) implies uniformly for $h\in\mathcal{H}_0$:

$$\begin{aligned}
&\mathrm{E}_h\max_{J\in\mathcal{I}_n}\Big|\frac{1}{n\mathrm{v}_J}\sum_i\Big(\|U_i(\Pi_{\mathcal{H}_0}h)\widetilde{b}^K(W_i)QI_{s_J}^{-1}\|^2-\mathrm{E}_h\|U(\Pi_{\mathcal{H}_0}h)\widetilde{b}^K(W)QI_{s_J}^{-1}\|^2\Big)\Big|^2\\
&\qquad\lesssim n^{-1}\zeta_J^2\sum_{J\in\mathcal{I}_n}\mathrm{v}_J^{-2}s_J^{-4}\lesssim n^{-1}\zeta_J^2\sum_{J\in\mathcal{I}_n}\Big(\sum_{j=1}^J s_J^4 s_j^{-4}\Big)^{-1}\lesssim n^{-1}\zeta_J^2\sum_{J\in\mathcal{I}_n}J^{-1}=o(1),
\end{aligned}$$

where the last inequality is due to Lemma B.3 and the definition of the index set $\mathcal{I}_n$. Consequently, due to the second moment condition imposed in Assumption 2(i) we obtain uniformly for $J \in \mathcal{I}_n$

$$n^{-1}\sum_i \|(Y_i - \Pi_{\mathcal{H}_0}h(X_i))\widetilde{b}^K(W_i)QI_{s_J}^{-1}\|^2 \leq \overline{\sigma}^2 c_0^{-1}\zeta_J\Big(\sum_{j=1}^J s_j^{-4}\Big)^{1/2} \leq \overline{\sigma}^2\underline{\sigma}^{-2}c_0^{-1}\zeta_J \mathrm{v}_J$$

with probability approaching one (under $h \in \mathcal{H}_0$), by making use of Lemma B.3. Further, we obtain uniformly for $h \in \mathcal{H}_0$:

$$\begin{aligned}
&\mathrm{P}_h\Big(\max_{J\in\mathcal{I}_n}\Big|\frac{(\log\log J)^{-1/2}}{(n-1)\mathrm{v}_J}\sum_{i,i'}U_i(\Pi_{\mathcal{H}_0}h)U_{i'}(\Pi_{\mathcal{H}_0}h)b^K(W_i)'\Big(A'A-\widehat{A}'\widehat{A}\Big)b^K(W_{i'})\Big| > \frac{1-c_0}{8}\Big)\\
&\leq \mathrm{P}_h\Big(\max_{J\in\mathcal{I}_n}\Big|\frac{(\log\log J)^{-1/2}}{(n-1)\mathrm{v}_J}\sum_{i\neq i'}U_i(\Pi_{\mathcal{H}_0}h)U_{i'}(\Pi_{\mathcal{H}_0}h)\widetilde{b}^K(W_i)'QI_{s_J}^{-2}Q'\widetilde{b}^K(W_{i'})\Big| > \frac{1-c_0}{8}\Big)\\
&\quad + \mathrm{P}_h\Big(\max_{J\in\mathcal{I}_n}\big(\|I_{s_J}QG_b^{1/2}(A'A-\widehat{A}'\widehat{A})G_b^{1/2}QI_{s_J}\|\big) > \frac{1-c_0}{16}\Big)\\
&\quad + \mathrm{P}_h\Big(\max_{J\in\mathcal{I}_n}\big(\overline{\sigma}^2\underline{\sigma}^{-2}c_0^{-1}\zeta_J(\log\log J)^{-1/2}\|I_{s_J}QG_b^{1/2}(A'A-\widehat{A}'\widehat{A})G_b^{1/2}Q'I_{s_J}\|\big) > \frac{1-c_0}{16}\Big) + o(1)\\
&=: T_1 + T_2 + T_3 + o(1).
\end{aligned}$$

Note that T_1 is arbitrarily small for C_0 sufficiently large by following step 1 in the proof of Theorem 4.1. Consider T_2. We make use of the inequality

$$\|I_{s_J}QG_b^{1/2}(\widehat{A}'\widehat{A}-A'A)G_b^{1/2}Q'I_{s_J}\| \leq 2\|I_{s_J}QG_b^{1/2}(\widehat{A}-A)'AG_b^{1/2}Q'I_{s_J}\| + \|(\widehat{A}-A)G_b^{1/2}QI_{s_J}\|^2.$$

It is sufficient to consider the first summand on the right hand side. Note that $\|AG_b^{1/2}Q'I_{s_J}\| \leq C_0^{-1}$. Consequently, from Lemma E.5(ii) we infer

$$\mathrm{P}\left(\max_{J\in\mathcal{I}_n}\Big\{\frac{s_J^2\sqrt{n}}{\zeta_J\sqrt{\log J}}\|I_{s_J}QG_b^{1/2}(\widehat{A}-A)'AG_b^{1/2}Q'I_{s_J}\|\Big\} > C\right) = o(1).$$

Consequently, Assumption 4(i), i.e., $s_J^{-1}\zeta_J^2\sqrt{(\log J)/n} = O(1)$ uniformly for $J \in \mathcal{I}_n$, implies $T_3 = o(1)$. □

Proof of Lemma B.4. It is sufficient to prove (ii). Let $\Sigma = \mathrm{E}_h[(Y-h(X))^2 b^{K(J)}(W)b^{K(J)}(W)']$ and $\widehat{\Sigma} = n^{-1}\sum_i\big(Y_i - \widehat{h}_J(X_i)\big)^2 b^{K(J)}(W_i)b^{K(J)}(W_i)'$. Then $\mathrm{v}_J = \|A\Sigma A'\|_F$ and $\widehat{\mathrm{v}}_J = \|\widehat{A}\widehat{\Sigma}\widehat{A}'\|_F$. For all $J \in \mathcal{I}_n$ the triangular inequality implies

$$|\widehat{\mathrm{v}}_J - \mathrm{v}_J| \leq \|\widehat{A}\widehat{\Sigma}\widehat{A}' - A\Sigma A'\|_F \leq 2\|(\widehat{A}-A)\widehat{\Sigma}A'\|_F + \|(\widehat{A}-A)\widehat{\Sigma}^{1/2}\|_F^2 + \|A(\widehat{\Sigma}-\Sigma)A'\|_F.$$

In the remainder of this proof, it is sufficient to consider $\|(\widehat{A}-A)\Sigma A'\|_F + \|A(\widehat{\Sigma}-\Sigma)A'\|_F =: T_1 + T_2$. Consider T_1. By Lemma E.2 we have the upper bound $\|G_b^{-1/2}\Sigma G_b^{-1/2}\| \leq \overline{\sigma}$. Below we make use of the inequality $\|m_1 m_2\|_F \leq \|m_1\|\|m_2\|_F$ for matrices m_1 and m_2. Since the

Frobenius norm is invariant under rotation, we calculate uniformly for $J \in \mathcal{I}_n$ that

$$\begin{aligned} T_1 &= \|(G_b^{1/2} S G^{1/2})(\widehat{A} - A)\Sigma A' A G_b^{1/2}\| \\ &\leq \|(G_b^{1/2} S G^{1/2})(\widehat{A} - A) G_b^{1/2}\| \|G_b^{-1/2} \Sigma G_b^{-1/2}\| \|(G_b^{1/2} S G^{1/2})_l^{-2}\|_F \lesssim \frac{\zeta_J}{s_J}\Big(\frac{\log(J)}{n}\sum_{j=1}^J s_j^{-4}\Big)^{1/2} \end{aligned}$$

wpa1 uniformly or $h \in \mathcal{H}$, by making use of Lemma E.5(i) and the Schur decomposition as in the proof of Lemma B.3. From Assumption 4(i), i.e., $s_J^{-1}\zeta_J^2\sqrt{(\log J)/n} = O(1)$, uniformly for $J \in \mathcal{I}_n$ we infer $T_1/v_J = J^{-1/2}(\sum_{j=1}^J s_j^{-4})^{1/2}/v_J \to 0$ wpa1 uniformly or $h \in \mathcal{H}$, where the last equation is due to Lemma B.3. Consider T_2. Again using Lemma B.3 we obtain $T_2 \leq \underline{\sigma}^{-2}\|G_b^{-1/2}(\widehat{\Sigma} - \Sigma)G_b^{-1/2}\|$ by using the upper bound as derived for T_1. Further, evaluate

$$\begin{aligned} \|G_b^{-1/2}(\widehat{\Sigma} - \Sigma)G_b^{-1/2}\| &= \Big\|\frac{1}{n}\sum_i \big((Y_i - \widehat{h}_J(X_i))^2 - (Y_i - h(X_i))^2\big)\widetilde{b}^K(W_i)\widetilde{b}^K(W_i)'\Big\| \\ &\leq \Big\|\frac{1}{n}\sum_i \big(\widehat{h}_J(X_i) - h(X_i)\big)^2 \widetilde{b}^K(W_i)\widetilde{b}^K(W_i)'\Big\| \\ &\quad + 2\Big\|\frac{1}{n}\sum_i \big(\widehat{h}_J(X_i) - h(X_i)\big)(Y_i - h(X_i))\widetilde{b}^K(W_i)\widetilde{b}^K(W_i)'\Big\| =: T_{21} + T_{22}. \end{aligned}$$

Consider T_{21}. The definition of the unrestricted sieve NPIV estimator in (2.6) implies uniformly for $J \in \mathcal{I}_n$

$$\begin{aligned} T_{21} &\leq \Big\|\frac{1}{n}\sum_i \big(\widehat{h}_J(X_i) - Q_J h(X_i)\big)^2 \widetilde{b}^K(W_i)\widetilde{b}^K(W_i)'\Big\| + \Big\|\frac{1}{n}\sum_i \big(Q_J h(X_i) - h(X_i)\big)^2 \widetilde{b}^K(W_i)\widetilde{b}^K(W_i)'\Big\| \\ &\leq \zeta_J^2 \Big\|\widehat{A}\frac{1}{n}\sum_i Y_i b^K(W_i) - A\,\mathrm{E}_h[Y b^K(W)]\Big\|^2 \times \Big\|\frac{1}{n}\sum_i \psi^J(X_i)\psi^J(X_i)'\Big\| \\ &\quad + \zeta_J^2 \Big\|\frac{1}{n}\sum_i \big(Q_J h(X_i) - h(X_i)\big)^2\Big\| \lesssim \zeta_J^4 s_J^{-2} n^{-1} + \max_{J \in \mathcal{I}_n}\big\{\zeta_J^2 \|Q_J h - h\|_{L^2(X)}\big\} \end{aligned}$$

wpa1 uniformly for $h \in \mathcal{H}$, where the right hand side tends to zero. This follows by the rate condition imposed in Assumption 4(i) and that $\|Q_J h - h\|_{L^2(X)} = O(J^{-p/d_x})$ uniformly for $J \in \mathcal{I}_n$ and $h \in \mathcal{H}$ by Lemma B.1(ii). Analogously, we obtain that $\max_{J \in \mathcal{I}_n} T_{22}$ vanishes wpa1 uniformly for $h \in \mathcal{H}$. □

Proof of Lemma B.5. We first prove the lower bound. By the definition of the RES index set $\widehat{\mathcal{I}}_n$ we have that any element $J \in \widehat{\mathcal{I}}_n$, tends slowly to infinity as $n \to \infty$. Let $\widehat{j}_{\max} \leq j_{\max}$ be the largest integer such that $\underline{J} 2^{\widehat{j}_{\max}} \leq \widehat{J}_{\max}$. Consequently, the definition of the RES index set implies for all $J \in \widehat{\mathcal{I}}_n$ that

$$\log(J) \leq \log(\underline{J} 2^{\widehat{j}_{\max}}) = \widehat{j}_{\max}\log(2) + \log(\underline{J}) \leq \widehat{j}_{\max} + 1 = \#(\widehat{\mathcal{I}}_n)$$

for n sufficiently large. From the lower bounds for quantiles of the chi-squared distribution established in Inglot [2010, Theorem 5.2] we deduce for all $J \in \widehat{\mathcal{I}}_n$ and n sufficiently large:

$$\begin{aligned}\widehat{\eta}_J(\alpha) &= \frac{q\big(\alpha/\#(\widehat{\mathcal{I}}_n), J\big) - J}{\sqrt{J}} \geq \frac{q\big(\alpha/(\log J), J\big) - J}{\sqrt{J}} \\ &\geq \frac{\sqrt{\log\big((\log J)/\alpha\big)}}{4} + \frac{2\log\big((\log J)/\alpha\big)}{\sqrt{J}} \geq \frac{\sqrt{\log\log(J) - \log(\alpha)}}{4}\end{aligned}$$

using the lower bounds for quantiles of the chi-squared distribution established in Inglot [2010, Theorem 5.2]. We now consider the upper bound. From the definition of $\#(\widehat{\mathcal{I}}_n)$ we infer $\#(\widehat{\mathcal{I}}_n) = \widehat{j}_{\max} + 1 \leq \lceil \log_2(n^{1/3}/\underline{J})\rceil + 1 \leq \log(n^{1/3}/\underline{J}) + 1$ and thus $\#(\widehat{\mathcal{I}}_n) \leq \log(n)$. Consequently, we calculate for all $J \in \widehat{\mathcal{I}}_n$ and n sufficiently large:

$$\begin{aligned}\widehat{\eta}_J(\alpha) &\leq \frac{q\big(\alpha/(\log n), J\big) - J}{\sqrt{J}} \leq 2\sqrt{\log\big((\log n)/\alpha\big)} + \frac{2\log\big((\log n)/\alpha\big)}{\sqrt{J}} \\ &\leq 2\sqrt{\log\big((\log n)/\alpha\big)}(1 + o(1)) \leq 4\sqrt{\log\log(n) - \log(\alpha)},\end{aligned}$$

where the second inequality is due to Laurent and Massart [2000, Lemma 1]. □

Proof of Lemma B.6. Result B.6(i) directly follows from Houdré and Reynaud-Bouret [2003, Theorem 3.4]; see also Gine and Nickl [2016, Theorem 3.4.8]. We next prove the bounds on Λ_1, Λ_2, Λ_3, Λ_4 for Result B.6(ii).
For the bound on Λ_1, we recall the notation $V_i^J = U_i A b^K(W_i)$ with $U_i = Y_i - h(X_i)$ for $h \in \mathcal{H}_0$. Then, under $\mathcal{H}_0$ we have:

$$\begin{aligned}\mathrm{E}_h[R_1^2(Z_1, Z_2)] &\leq \mathrm{E}_h\left|U_1 b^K(W_1)' A' A b^K(W_2) U_2\right|^2 = \mathrm{E}_h\left[(V^J)' \,\mathrm{E}_h\left[V^J (V^J)'\right] V^J\right] \\ &= \sum_{j,j'=1}^{J} \mathrm{E}_h[V_j V_{j'}]^2 = \mathrm{v}_J^2.\end{aligned}$$

For the bound on Λ_2, for any function ν and κ with $\|\nu\|_{L^2(Z)} \leq 1$ and $\|\kappa\|_{L^2(Z)} \leq 1$, respectively, we obtain

$$\begin{aligned}|\,\mathrm{E}_h[R_1(Z_1, Z_2)\nu(Z_1)\kappa(Z_2)]| \leq & \left|\,\mathrm{E}_h[U\mathbb{1}_M b^K(W)'\nu(Z)]A'A\,\mathrm{E}_h[U\mathbb{1}_M b^K(W)\kappa(Z)]\right| \\ \leq & \|A\,\mathrm{E}_h[U\mathbb{1}_M b^K(W)\kappa(Z)]\|\,\|A\,\mathrm{E}_h[U\mathbb{1}_M b^K(W)\nu(Z)]\| \\ \leq & \|AG_b^{1/2}\|^2\sqrt{\mathrm{E}\left[|\,\mathrm{E}_h[U\mathbb{1}_M\kappa(Z)|W]|^2\right]} \times \sqrt{\mathrm{E}\left[|\,\mathrm{E}_h[U\mathbb{1}_M\nu(Z)|W]|^2\right]}\end{aligned}$$

Now observe $\mathrm{E}\left[|\,\mathrm{E}_h[U\mathbb{1}_M\kappa(Z)|W]|^2\right] \leq \mathrm{E}\left[\,\mathrm{E}_h[U^2|W]\kappa^2(Z)\right] \leq \overline{\sigma}^2$ by Assumption 2(i) and using that $\|\kappa\|_{L^2(Z)} \leq 1$, which yields the upper bound by using $\|AG_b^{1/2}\| = s_J^{-1}$.

For the bound on Λ_3, observe that for any $z = (u, w)$

$$\begin{aligned}\big|\operatorname{E}_h[R_1^2(Z_1, z)]\big| &\leq \operatorname{E}_h \Big| U \mathbb{1}\{|U| \leq M_n\} b^K(W)' A' A b^K(w) u \mathbb{1}\{|u| \leq M_n\}\Big|^2 \\ &\leq \|A b^K(w) u \mathbb{1}\{|u| \leq M_n\}\|^2 \operatorname{E}_h \|A b^K(W) U\|^2 \leq \overline{\sigma}^2 M_n^2 \zeta_{b,K}^2 \|A G_b^{1/2}\|^4,\end{aligned}$$

again by using Assumption 2(i) and hence the upper bound on Λ_3 follows.
For the bound on Λ_4, observe that for any $z_1 = (u_1, w_1)$ and $z_2 = (u_2, w_2)$ we get

$$\begin{aligned}\big|R_1(z_1, z_2)\big| &\leq \Big|u_1 \mathbb{1}\{|u_1| \leq M_n\} b^K(w_1)' A' A b^K(w_2) u_2 \mathbb{1}\{|u_2| \leq M_n\}\Big| \\ &\leq \sup_{u,w} \big\|A b^K(w) u \mathbb{1}\{|u| \leq M_n\}\big\|^2 \leq M_n^2 \zeta_{b,K}^2 \|A G_b^{1/2}\|^2,\end{aligned}$$

which completes the proof. □

Proof of Lemma B.7. It suffices to prove (ii) for a simple null $\mathcal{H}_0 = \{h_0\}$. For any $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$, we denote $B_J = (\|\operatorname{E}_h[V^J]\| - \|h - h_0\|_{L^2(X)})^2$. Applying $\|\operatorname{E}_h[V^{J^*}]\|^2 = \|Q_{J^*}(h - h_0)\|_{L^2(X)}^2$ and Lemma B.1(i) we obtain: $B_{J^*} = \big(\|Q_{J^*}(h - h_0)\|_{L^2(X)} - \|h - h_0\|_{L^2(X)}\big)^2 \leq C_B\, \mathsf{r}_n^2$ for some constant C_B. By the inequality $\|\operatorname{E}_h[V^{J^*}]\|^2 \geq \|h - h_0\|_{L^2(X)}^2/2 - B_{J^*}$ we have uniformly for $h \in \mathcal{H}_1(\delta^\circ r_{n,J})$:

$$\begin{aligned}&\operatorname{P}_h\Big(n\,\widehat{D}_{J^*}(h_0) \leq 2c_1\sqrt{\log\log n}\,\mathrm{v}_{J^*}\Big) = \operatorname{P}_h\Big(\|\operatorname{E}_h[V^{J^*}]\|^2 - \widehat{D}_{J^*}(h_0) > \|\operatorname{E}_h[V^{J^*}]\|^2 - \frac{2c_1\sqrt{\log\log n}\,\mathrm{v}_{J^*}}{n}\Big) \\ &\leq \operatorname{P}_h\Big(\Big|\frac{4}{n(n-1)}\sum_{j=1}^{J^*}\sum_{i<i'}\big(V_{ij}V_{i'j} - \operatorname{E}_h[V_{1j}]^2\big)\Big| > \rho_h\Big) \\ &+ \operatorname{P}_h\Big(\Big|\frac{4}{n(n-1)}\sum_{i<i'}(Y_i - h_0(X_i))(Y_{i'} - h_0(X_{i'}))b^{K^*}(W_i)'\big(A'A - \widehat{A}'\widehat{A}\big)b^{K^*}(W_{i'})\Big| > \rho_h\Big) = T_1 + T_2,\end{aligned}$$

where $\rho_h = \|h - h_0\|_{L^2(X)}^2/2 - 2c_1 n^{-1}\sqrt{\log\log n}\mathrm{v}_{J^*} - B_{J^*}$. To bound term T_1, we apply inequality (E.5) and Markov's inequality:

$$T_1 \lesssim n^{-1} s_{J^*}^{-2}\rho_h^{-2}\big(\|h - h_0\|_{L^2(X)}^2 + (J^*)^{-2p/d_x}\big) + n^{-2}\mathrm{v}_{J^*}^2\rho_h^{-2}. \tag{E.7}$$

In the following, we distinguish between two cases. First, consider the case where $n^{-2}\mathrm{v}_{J^*}^2\rho_h^{-2}$ dominates the right hand side. For any $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ we have $\|h - h_0\|_{L^2(X)} \geq \delta^\circ \mathsf{r}_n$ and hence, we obtain the lower bound

$$\rho_h = \|h - h_0\|_{L^2(X)}^2/2 - 2c_1 n^{-1}\sqrt{\log\log n}\mathrm{v}_{J^*} - B_{J^*} \geq \kappa_0 \mathsf{r}_n^2 \tag{E.8}$$

where $\kappa_0 := (\delta^\circ)^2/2 - C - C_B$ for some constant $C > 0$ and $\kappa_0 > 0$ whenever $\delta^\circ > \sqrt{2(C + C_B)}$. From inequality (E.7) we infer $T_1 \lesssim n^{-2}\mathrm{v}_{J^*}^2(J^*)^{4p/d_x} = o(1)$. Second, consider the case where $n^{-1}s_{J^*}^{-2}\rho_h^{-2}\big(\|h-h_0\|_{L^2(X)}^2+(J^*)^{-2p/d_x}\big)$ dominates. For any $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$

we have $\|h-h_0\|^2_{L^2(X)} \geq (\delta^\circ)^2 \mathsf{r}_n^2 \geq 5c_1 n^{-1} v_{J^*}\sqrt{\log\log n}$ for δ° sufficiently large and hence, we obtain $\rho_h \geq \kappa_1 \|h-h_0\|^2_{L^2(X)}$ for some constant $\kappa_1 := 1/5 - C_B/(\delta^\circ)^2$ which is positive for any $\delta^\circ > \sqrt{5C_B}$. Hence, inequality (E.7) yields uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ that

$$T_1 \lesssim n^{-1} s_{J^*}^{-2}\Big(\|h-h_0\|^{-2}_{L^2(X)} + \|h-h_0\|^{-4}_{L^2(X)}(J^*)^{-2p/d}\Big) \lesssim n^{-1} s_{J^*}^{-2}\mathsf{r}_n^{-2} = o(1).$$

Finally, $T_2 = o(1)$ uniformly for $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$ by making use of Lemma E.1. $\square$

Proof of Lemma B.8. Recall the definition of $\overline{J} = \sup\{J : \zeta^2(J)\sqrt{(\log J)/n} \leq \overline{c}\, s_J\}$. Following the proof of Chen et al. [2021, Lemma C.6], using Weyl's inequality (see e.g. Chen and Christensen [2018, Lemma F.1]) together with Chen and Christensen [2018, Lemma F.7] we obtain that $|\widehat{s}_J - s_J| \leq c_0 s_J$ uniformly in $J \in \mathcal{I}_n$ for some $0 < c_0 < 1$ with probability approaching one uniformly for $h \in \mathcal{H}$.
Proof of (i). By making use of the definition of $\widehat{J}_{\max}$ given in (2.12), we obtain uniformly for $h \in \mathcal{H}$:

$$\mathrm{P}_h\Big(\widehat{J}_{\max} > \overline{J}\Big) \leq \mathrm{P}_h\Big(\zeta^2(\overline{J})\sqrt{\log(\overline{J})/n} < \frac{3}{2}\widehat{s}_{\overline{J}}\Big) \leq \mathrm{P}_h\Big(\zeta^2(\overline{J})\sqrt{\log(\overline{J})/n} < \frac{3}{2}(1+c_0)s_{\overline{J}}\Big) + o(1)$$

The upper bound imposed on the growth of $\overline{J}$ is determined by a sufficiently large constant $\overline{c} > 0$ and hence, there exists a constant $\underline{c} \geq 3(1+c_0)/2$ such that $s_{\overline{J}}^{-1}\zeta^2(\overline{J})\sqrt{\log(\overline{J})/n} \geq \underline{c}$. Consequently, we obtain

$$\mathrm{P}_h\Big(\widehat{J}_{\max} > \overline{J}\Big) \leq \mathrm{P}_h\Big(s_{\overline{J}}^{-1}\zeta^2(\overline{J})\sqrt{\log(\overline{J})/n} < \frac{3}{2}(1+c_0)\Big) + o(1) = o(1).$$

Proof of (ii). From the definition of J° given in (4.3) we infer as above for some constant $0 < c_0 < 1$ and uniformly for $h \in \mathcal{H}$:

$$\mathrm{P}_h\Big(J^\circ > \widehat{J}_{\max}\Big) \leq \mathrm{P}_h\Big((1-c_0)n^{-1}\sqrt{\log\log n}\,\widehat{J}_{\max}^{2p/d_x+1/2} \leq \widehat{s}^2_{\widehat{J}_{\max}}\Big) + o(1).$$

Consider the case $\zeta(J) = \sqrt{J}$. The definition of $\widehat{J}_{\max}$ in (2.12) yields uniformly for $h \in \mathcal{H}$:

$$\begin{aligned}\mathrm{P}_h\Big(J^\circ > \widehat{J}_{\max}\Big) &\leq \mathrm{P}_h\Big((1-c_0)\sqrt{\log\log n}\,\widehat{J}_{\max}^{2p/d_x-3/2} \leq (\log \overline{J})\Big) + o(1)\\ &\leq \mathrm{P}_h\Big((1-c_0)\widehat{s}_{\widehat{J}_{\max}}\sqrt{n} \leq \frac{2}{3}\sqrt{\log\overline{J}}\Big(\frac{\log\overline{J}}{\sqrt{\log\log n}}\Big)^{1/(2p/d_x-3/2)}\Big) + o(1)\\ &\leq \mathrm{P}_h\Big((1-c_0)^2 s_{\overline{J}}\sqrt{n} \leq \frac{2}{3}\sqrt{\log\overline{J}}\Big(\frac{\log\overline{J}}{\sqrt{\log\log n}}\Big)^{1/(2p/d_x-3/2)}\Big) + o(1)\\ &\leq \mathrm{P}_h\Big(\frac{(1-c_0)^2}{\overline{c}}\overline{J} \leq \frac{2}{3}\Big(\frac{\log\overline{J}}{\sqrt{\log\log n}}\Big)^{1/(2p/d_x-3/2)}\Big) + o(1),\end{aligned}$$

where the last inequality follows from the definition of $\overline{J}$, i.e., $s_{\overline{J}} \geq \overline{c}^{-1}\overline{J}\sqrt{\log(\overline{J})/n}$. From Assumption 4(iii), i.e., $p \geq 3d_x/4$, we infer $\mathrm{P}_h(J^\circ > \widehat{J}_{\max}) = o(1)$ and, in particular,

$\mathrm{P}_h(2J^\circ > \widehat{J}_{\max}) = o(1)$ uniformly for $h \in \mathcal{H}$. The proof of $\zeta(J) = J$ follows analogously using the condition $p \geq 7d_x/4$. $\square$

References

H. J. Bierens. A consistent conditional moment test of functional form. *Econometrica*, 58 (6):1443–1458, 1990.

C. Breunig. Specification testing in nonparametric instrumental quantile regression. *Econometric Theory*, 36(4):583–625, 2020.

C. Breunig and X. Chen. Adaptive, rate-optimal testing in instrumental variables models. *arXiv preprint arXiv:2006.09587v1*, 2020.

C. Breunig and X. Chen. Adaptive, rate-optimal hypothesis testing in nonparametric iv models. *arXiv preprint arXiv:2006.09587v2*, 2021.

X. Chen and T. M. Christensen. Optimal sup-norm rates and uniform inference on nonlinear functionals of nonparametric iv regression. *Quantitative Economics*, 9(1):39–84, 2018.

X. Chen, T. Christensen, and S. Kankanala. Adaptive estimation and uniform confidence bands for nonparametric structural functions and elasticities. *arXiv preprint arXiv:2107.11869*, 2021.

D. Chetverikov and D. Wilhelm. Nonparametric instrumental variable estimation under monotonicity. *Econometrica*, 85:1303–1320, 2017.

E. Gine and R. Nickl. *Mathematical Foundations of Infinite-Dimensional Statistical Models*. Cambridge University Press, 2016.

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

C. Houdré and P. Reynaud-Bouret. Exponential inequalities, with constants, for u-statistics of order two. In *Stochastic inequalities and applications*, pages 55–69. Springer, 2003.

T. Inglot. Inequalities for quantiles of the chi-square distribution. *Probability and Mathematical Statistics*, 30(2):339–351, 2010.

B. Laurent and P. Massart. Adaptive estimation of a quadratic functional by model selection. *The Annals of Statistics*, pages 1302–1338, 2000.

A. van der Vaart and J. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics (Springer Series in Statistics)*. Springer, 2000.