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

Christoph Breunig[†] Xiaohong Chen[‡]

First version: August 2018; Latest revision: July 30, 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 two-stage least squares 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 the supremum of type I error over the composite null and the supremum of type II error over nonparametric alternative models cannot be minimized by any other tests for NPIV models of unknown regularities. Confidence sets in L^2 are obtained by inverting the adaptive test. Simulations confirm that, across different strength of instruments and sample sizes, 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: Conditional instrumental variables; Shape restrictions; Composite hypothesis; Nonparametric alternatives; Minimax rate of testing; Adaptive hypothesis testing; Power; Random exponential scan; Sieve U statistics; Sieve two-stage least squares.

*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 (composite) 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, supermodularity, quasi-concavity). Our new test builds on a simple data-driven choice of tuning parameter that ensures asymptotic size control and non-trivial power uniformly against a large class of nonparametric alternatives.

Let $L^2(X)$ denote the space of square integrable function of X. Our new test is designed to test a composite null hypothesis $\mathcal{H}_0$ that is a closed, convex strict subset of $L^2(X)$ satisfying the NPIV model (1.1). Before presenting the theoretical properties of our new test, we derive the *minimax rate of testing* r_n in L^2, which is the fastest rate of separation in root-mean squared distance between the null hypothesis $\mathcal{H}_0$ and the class of nonparametric alternative NPIV functions $\mathcal{H}_1(\delta r_n)$ that enables consistent testing uniformly over the latter, with the rate r_n shrinking to zero as the sample size n goes to infinity and $\delta > 0$ being a finite constant independent of n. We establish the minimax result in two steps: First, we derive, uniformly over all possible tests, a lower bound for the sum of the supremum of type I error over $\mathcal{H}_0$ and the supremum of type II error over $\mathcal{H}_1(\delta r_n)$ separated from the null hypothesis by a rate r_n. 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 (i.e., sieve two-stage least squares) 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). This test is called 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 leave-one-out 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 within 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 our adaptive minimax rate of testing in L^2 is to derive a sharp upper bound on the convergence rate of a leave-one-out sieve estimator of a quadratic functional of a NPIV function, which is proved using an exponential inequality for U-statistics with increasing dimensions. We show that our adaptive test has asymptotic size control under a composite null by deriving a tight, slowly divergent lower bound for Bonferroni corrected chi-squared critical value. By inverting our adaptive tests we obtain L^2-confidence sets on restricted NPIV 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.

Our adaptive minimax L^2 rate of testing decreases to zero strictly faster than the optimal L^2 rate of estimation (with known smoothness) for mildly ill-posed NPIV models, and coincides with the optimal L^2 rate of estimation for severely ill-posed NPIV models. In the existing literature on testing for parametric, semiparametric or shape NPIV restrictions against nonparametric alternatives, all of the non-adaptive tests achieve their asymptotic size controls by choosing some deterministic tuning parameters such that the L^2 estimation bias for h is of a smaller order than the L^2 standard deviation (aka, under-smoothing), which leads to a L^2 separation rate of testing shrinking to zero strictly slower than the optimal L^2 rate of estimation, and hence strictly slower than our adaptive minimax L^2 rate of testing for both mildly and severely ill-posed NPIV models. In particular, among all of the existing NPIV tests that have asymptotic size controls, our new adaptive test is asymptotically more powerful, uniformly over a larger class of nonparametric alternatives.

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

[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).

from others' work. The simulations reveal the following patterns of our adaptive test in comparison to recent non-adaptive tests: First, while the competing tests can be over-sized at the boundary of the null hypothesis, our test delivers adequate size control under different composite null hypotheses, across different sample sizes and for varying strengths of instruments. Second, our test is as powerful as the competing tests when alternative functions are relatively simple, and is more powerful when alternatives are more nonlinear/complex. The great power gains of our adaptive test are present even for relatively weak strength of instruments or small sample sizes. These findings highlight the importance of our data-driven choice of the sieve dimension to simultaneously ensure size control and powerful performance uniformly against a larger class of nonparametric alternative NPIV functions. Finally, unlike many NPIV tests using bootstrapped critical values, our powerful adaptive test uses simple Bonferroni corrected chi-squared critical values and hence is fast to compute.

We present two empirical applications of our adaptive test. The first is to test the connected substitutes shape restrictions in demand for differential products using market level data (e.g., Berry and Haile (2014)). The second is to test for monotonicity, convexity and parametric forms in Engel curves (e.g., Blundell, Chen, and Kristensen (2007)).

There are growing number of papers on testing equality and inequality (shape) restrictions in NPIV type models. See, e.g., Horowitz (2006), Santos (2012), Breunig (2015), Chen and Pouzo (2015), Chernozhukov, Newey, and Santos (2015), Zhu (2020), Fang and Seo (2021) and references therein.[2] Most of these papers assume that some non-random sequences of key tuning (regularization) parameters satisfy some theoretical rate conditions with known smoothness of NPIV functions. None of the published work achieves the adaptive minimax L^2 rate of testing for NPIV models. Our paper makes an important contribution by providing the first data-driven choice of a key tuning parameter that leads to a new minimax rate-adaptive and powerful test for equality and inequality (shape) restrictions in NPIV models. Our paper also complements a concurrent work by Chen, Christensen, and Kankanala (2024), 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).

The rest of the paper is 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 new test. Section 5 presents simulation studies and Section 6 provides empirical illustrations. Appendices A and B present proofs of Theorems 3.1, 3.2, 4.1 and 4.2. The online supplementary appendices

[2]There are also papers on NPIV estimation by directly imposing shape restrictions; see, e.g., Horowitz and Lee (2012), Blundell, Horowitz, and Parey (2017), Chetverikov and Wilhelm (2017) and Freyberger and Reeves (2019). See Chetverikov, Santos, and Shaikh (2018) for a review on shape restrictions and Chetverikov (2019) for adaptive kernel testing for monotonicity of a regression without endogeneity.

include Appendix C for additional simulation results, Appendix D for proofs of Corollaries 4.1 and 4.2, and Appendix E for additional lemmas and their proofs.

Basic notation. For a random variable X, we let $L^2(X)$ denote the Hilbert space of real-valued measurable functions ϕ of X with finite second moment, with the norm $\|\phi\|_{L^2(X)} := \sqrt{\mathrm{E}[\phi^2(X)]}$ and the inner product $\langle\cdot,\cdot\rangle_X$. Let $\|\phi\|_\infty := \sup_x |\phi(x)|$ be the sup-norm and $L^\infty = \{\phi : \|\phi\|_\infty < \infty\}$. For a matrix M, let M' be its transpose and M^- be 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 $\|\cdot\|$ be 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.

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. We can then equivalently express the NPIV model (2.1) as $\mathrm{E}_h[Y|W] = (Th)(W)$ for $h \in \mathcal{H}$. For ease of 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). We assume that $\mathcal{H}_0$ is a *nonempty, closed and convex, strict subset* of $\mathcal{H}$. For any $h \in \mathcal{H}$ there is 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 testing 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 $\mathcal{H}$.*

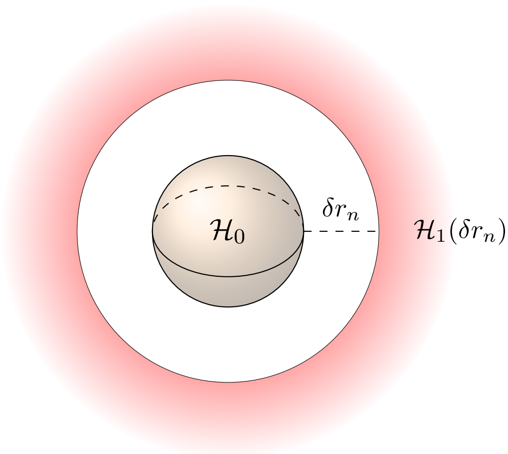
Example 2.2 (Semiparametric restrictions). *Let $F(\cdot;\theta,g)$ be a known function up to unknown (θ,g) and $\mathcal{H}_0 = \{h \in \mathcal{H} : h(\cdot) = F(\cdot;\theta,g) \text{ for some } \theta \in \Theta \text{ and } g \in \mathcal{G}\}$, for a finite-dimensional, convex compact parameter 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 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

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

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

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

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 conditional expectation operator $T : L^2(X) \mapsto L^2(W)$. Both are unknown in practice. 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. Precisely, we propose a feasible test statistic $\widehat{\mathtt{T}}_n$ with data-driven tuning parameters in Section 2.2. We show that $\widehat{\mathtt{T}}_n$ attains the minimax rate of testing r_n^* within a $\sqrt{\log\log(n)}$ multiplicative factor, has asymptotic size control over the composite null, and is consistent uniformly over the class of nonparametric alternatives in Theorem 4.2. We call our 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 $K = K(J) \geq J$ 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

$$\|Q_J(h - \Pi_{\mathcal{H}_0}h)\|^2_{L^2(X)} = \left\| A\, \mathrm{E}_h\left[b^K(W)(Y - \Pi_{\mathcal{H}_0}h(X))\right]\right\|^2, \tag{2.4}$$

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

For each sieve dimension J, we construct a test based on an estimated quadratic distance $\|Q_J(h - \Pi_{\mathcal{H}_0}h)\|^2_{L^2(X)}$ 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, Chen, and Kristensen (2007)):

$$\begin{aligned}\widehat{h}_J &= \arg\min_{\phi \in \Psi_J} \sum_{1 \le i,i' \le n} \big(Y_i - \phi(X_i)\big) b^{K(J)}(W_i)' \widehat{A}' \widehat{A} b^{K(J)}(W_{i'}) \big(Y_{i'} - \phi(X_{i'})\big) \\ &= \psi^J(\cdot)'[\Psi' P_B \Psi]^- \Psi' P_B \mathrm{Y} \ . \end{aligned} \tag{2.5}$$

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

$$\widehat{h}^{\mathrm{R}}_J = \arg\min_{\phi \in \mathcal{H}_{0,J}} \sum_{1 \le i,i' \le n} \big(Y_i - \phi(X_i)\big) b^{K(J)}(W_i)' \widehat{A}' \widehat{A} b^{K(J)}(W_{i'}) \big(Y_{i'} - \phi(X_{i'})\big). \tag{2.6}$$

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 $n\widehat{D}_J/\widehat{V}_J$, which is a standardized, centered (or leave-one-out) version of the sample analog of (2.4):

$$\widehat{D}_J = \frac{2}{n(n-1)} \sum_{1 \le i < i' \le n} \big(Y_i - \widehat{h}^{\mathrm{R}}_J(X_i)\big) b^{K(J)}(W_i)' \widehat{A}' \widehat{A} b^{K(J)}(W_{i'}) \big(Y_{i'} - \widehat{h}^{\mathrm{R}}_J(X_{i'})\big), \tag{2.7}$$

$$\widehat{V}_J = \left\| \widehat{A} \left(\frac{1}{n} \sum_{i=1}^{n} (Y_i - \widehat{h}_J(X_i))^2 b^{K(J)}(W_i) b^{K(J)}(W_i)' \right) \widehat{A}' \right\|_F , \tag{2.8}$$

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

$$V_J = \big\| A\, \mathrm{E}_h[(Y - h(X))^2 b^{K(J)}(W) b^{K(J)}(W)'] A' \big\|_F , \tag{2.9}$$

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

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:

$$\widehat{\mathcal{I}}_n := \left\{ J \le \widehat{J}_{\max} : \quad J = \underline{J} 2^j \text{ where } j = 0, 1, \ldots, j_{\max} \right\} \tag{2.10}$$

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

$$\widehat{J}_{\max} := \min \left\{ J > \underline{J} : \quad 1.5\, [\zeta(J)]^2 \sqrt{(\log J)/n} \ge \widehat{s}_J \right\}, \tag{2.11}$$

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

$$\widehat{\mathtt{T}}_n := \mathbb{1} \left\{ \text{there exists } J \in \widehat{\mathcal{I}}_n \text{ such that } \quad \widehat{\mathcal{W}}_J(\alpha) > 1 \right\} \tag{2.12}$$

where $\mathbb{1}\{\cdot\}$ is the indicator function. Under the nominal level $\alpha \in (0, 1)$, $\widehat{\mathtt{T}}_n = 1$ indicates rejection of the null $\mathcal{H}_0$ and $\widehat{\mathtt{T}}_n = 0$ indicates a failure to reject the null.

Remark 2.1 (Index set for J)**.** *The RES index set $\widehat{\mathcal{I}}_n$ 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 simulations and empirical applications where we have used quadratic B-splines, 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. For other sieve bases one might need to use a different constant to ensure a sufficiently large index set.*

Remark 2.2 (Choice of K)**.** *Our adaptive testing procedure lets $K := K(J)$ be any deterministic function of J satisfying $\lim_{J\to\infty} \frac{K(J)}{J} = c \in [1, \infty)$, and simply optimizes over $J \in \widehat{\mathcal{I}}_n$. Our theoretical results, including the asymptotic size control, are valid for any finite constant $c \geq 1$. In simulation studies and real data applications we let $K(J) = cJ$. Since a larger $c > 1$ implies more over-identification restrictions in a sieve NPIV (2SLS) regression, we expect that a larger $c > 1$ would lead to better power in finite samples. We have tried $K(J) \in \{2J,\ 4J,\ 8J\}$ in simulation studies in various designs. The simulation results show that (i) our adaptive test indeed has size control regardless of sample sizes, strength of instruments and even when $K(J) = 8J$; (ii) while our adaptive test with $K(J) = 8J$ has better empirical power for small sample sizes and weak instruments, the empirical powers are not sensitive to the choice of K for moderate to large sample sizes or strong instruments. These findings are 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.*

Remark 2.3 (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.*

3. The Minimax Rate of Testing

This section derives the minimax separation rate of hypothesis testing in NPIV models, when $\mathcal{H}$ is a relative compact subset of $L^2(X)$. For simplicity, we assume in this paper that

$\mathcal{H}$ is a standard Sobolev ellipsoid of smoothness $p > 0$, which can be expressed as

$$\mathcal{H} = \left\{ h \in L^2(X) : \sum_{j=1}^{\infty} j^{2p/d_x} \langle h, \widetilde{\psi}_j \rangle_X^2 \le C_{\mathcal{H}}^2 \right\}, \quad \text{for a finite constant } C_{\mathcal{H}} > 0,$$

where $\{\widetilde{\psi}_j\}_{j=1}^{\infty}$ is the orthonormal basis for $L^2(X)$ that is constructed from the basis $\{\psi_j\}_{j=1}^{\infty}$ (using the Gram-Schmidt procedure). Assuming the smoothness p is known, we first establish a lower bound for the L^2–rate of testing in Subsection 3.1, and then show that the lower bound can be achieved by a sieve test if the sieve dimension J can be chosen optimally in Subsection 3.2.

3.1. The Lower Bound

Before we state the lower bound for the rate of testing, we introduce the main assumptions.

Assumption 1. *(i)* $\inf_{w \in \mathcal{W}} \inf_{h \in \mathcal{H}} \mathbb{V}ar_h(Y - h(X)|W = w) \ge \underline{\sigma}^2 > 0$*; (ii) for any* $h \in \mathcal{H}$, $Th = 0$ *implies that* $\|h\|_{L^2(X)}^2 = 0$*; (iii) the densities of* X *and* W *are uniformly bounded below from zero and from above on their supports, which are Cartesian product of bounded intervals; (iv) there are a finite constant* $C > 0$ *and a positive decreasing function* ν *with* $\nu_j := \nu(j)$ *such that* $\|Th\|_{L^2(W)}^2 \le C \sum_{j \ge 1} \nu_j^2 \langle h, \widetilde{\psi}_j \rangle_X^2$ *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), the faster the sequence ν_j in Assumption 1(iv) decreases to zero, and the harder it is to detect properties of the NPIV function in the $L^2(X)$ 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 hold. Consider testing a closed convex null* $\mathcal{H}_0$ *versus* $\mathcal{H}_1(\delta r_n) = \{h \in \mathcal{H} : \|h - \mathcal{H}_0\|_{L^2(X)} \ge \delta r_n\}$ *for some constant* $\delta > 0$ *and a separation rate*

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

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

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

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*

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

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

(2) Severely ill-posed ($\nu_j = \exp(-j^{a/d_x}/2)$) *case:* $r_n \sim (\log n)^{-p/a}$.

According to Theorem 3.1, the lower bound of the L^2-rate of testing is $n^{-2p/(4(p+a)+d_x)}$ in the mildly ill-posed case, which goes to zero faster than the lower bound $n^{-p/(2(p+a)+d_x)}$ of the L^2-rate of estimation (Hall and Horowitz (2005) and Chen and Reiß (2011)). For the severely ill-posed NPIV models, the lower bound of the L^2-rate of testing is $(\log n)^{-p/a}$, which coincides with the lower bound of estimation in both the L^2 norm (Chen and Reiß (2011)) and the sup-norm (Chen and Christensen (2018)).

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:

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

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, Chen, and Kristensen (2007)). Since the conditional expectation operator T is compact under very mild conditions (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 a Simple Null Hypothesis

For a simple null $\mathcal{H}_0 = \{h_0\}$, we redefine $\widehat{D}_J$ in (2.7) with $\widehat{h}_J^{\mathrm{R}} = h_0$ as

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

We also redefine our test statistic $\widehat{\mathtt{T}}_n$ with a singleton RES index set $\{J\}$ as

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

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 $\Pi_K m(\cdot) = b^K(\cdot)'G_b^{-1}\,\mathrm{E}[b^K(W)m(W)]$ for $m \in L^2(W)$. Further, let $s_J = \inf_{h\in\Psi_J} \|\Pi_K Th\|_{L^2(W)}/\|h\|_{L^2(X)}$, i.e., s_J coincides with the minimal singular value of $G_b^{-1/2}SG^{-1/2}$. 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)* $\sup_{w\in\mathcal{W}}\sup_{h\in\mathcal{H}} \mathrm{E}_h[(Y - \widetilde{h}(X))^2|W = w] \leq \overline{\sigma}^2 < \infty$ *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)* $s_J^{-1}\|\Pi_K T(\Pi_J h - h)\|_{L^2(W)} \leq C_T\|\Pi_J h - h\|_{L^2(X)}$ *for a constant* $C_T > 0$*, uniformly for* $h \in \mathcal{H}$*.*

Let $\Psi_{J,1} := \{h \in \Psi_J : \|h\|_{L^2(X)} = 1\}$. 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, Chen, and Kristensen (2007)). We have $s_J \leq \tau_J^{-1}$ by definition.

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 unconditional 4th moment. We note that the DGP displayed in our proof of Theorem 3.1 already satisfies this assumption, it has no effect on our lower bound result. Assumptions 2(ii)(iii)(iv) are imposed since our test statistic involves linear sieve estimated operator T to achieve the separation rate. Assumptions 2(ii)(iii) impose restrictions on the sieve dimension J, which are satisfied by J_* given in (3.2) of Theorem 3.1. Assumption 2(iv) imposes an upper bound on the smoothing properties of the conditional expectation operator T. It is akin to the L^2 stability condition used in sieve NPIV estimation and is satisfied by Riesz bases (see Blundell, Chen, and Kristensen (2007, Assumption 6)). 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 asymptotically equivalent:

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

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

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

Further, $s_J \sim \tau_J^{-1} \leq \|T\widetilde{\psi}_J\|_{L^2(W)} \leq \sqrt{C}\nu_J$ under Assumption 1(iv) and $\{\widetilde{\psi}_j\}_j$ being an orthonormal basis in $L^2(X)$, and Assumption 2(iv) is satisfied under Assumptions 1(iv) and 3. Therefore, Assumptions 2 and 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 hold. 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*

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

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 $J_{*0} := \max\big\{J : n^{-1/2}\nu_J^{-1} J^{1/4} \leq J^{-p/d_x}\big\}$. 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* for 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 $r_{n,J_{*0}} \sim n^{-2p/(4(p+a)+d_x)}$ than the optimal L^2 rate of estimation $n^{-p/(2(p+a)+d_x)}$ in the mildly ill-posed case, the optimal L^2-rate of sieve testing $r_{n,J_{*0}} \sim (\log n)^{-p/a}$ in the severely ill-posed case is the same as the optimal rate of sieve estimation in both the L^2-norm (Chen and Reiß (2011)) and the sup-norm (Chen and Christensen (2018)). This is because, in the severely ill-posed case, the bias term dominates the standard deviation term for the optimally chosen sieve dimension in both sieve testing and estimation.

4. Adaptive Inference

This section establishes theoretical properties of our test $\widehat{\mathtt{T}}_n$ defined in (2.12). We show that it adapts to the unknown smoothness $p > 0$ of the functions in $\mathcal{H}$. Subsection 4.1 establishes the rate optimality of our 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 a Simple Null Hypothesis

Under the simple null hypothesis $\mathcal{H}_0 = \{h_0\}$ with a known function h_0 satisfying (1.1), our test $\widehat{\mathtt{T}}_n$ given in (2.12) simplifies to

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

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

Recall that the RES index set $\widehat{\mathcal{I}}_n$, given in (2.10), depends on an upper bound $\widehat{J}_{\max}$ given in (2.11). 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) $s_J^{-4} J \lesssim \sum_{j=1}^J s_j^{-4}$ uniformly for all $J \in \mathcal{I}_n$; (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.*

Assumptions 4(i)(iii) strengthen Assumption 2(ii)(iii)(iv) to hold uniformly over the deterministic index set $\mathcal{I}_n$. They are used to establish Lemma B.8. Assumption 4(i) restricts the growth of the deterministic upper bound $\overline{J}$ of the RES index set $\widehat{\mathcal{I}}_n$. Assumption 4(ii) is satisfied if $\{s_j\}$ is regularly varying, which is implied by Assumptions 1(iv) and 3 with $\{\nu_j\}$ regularly varying. We note that Assumptions 4(ii), 2(i) and 1(i) together imply that $V_J \sim s_J^{-2}\sqrt{J}$ uniformly for $h \in \mathcal{H}$ and $J \in \mathcal{I}_n$ (see Lemmas B.2 and B.3).

Theorem 4.1. *Let Assumptions 1(i)-(iii), 2(i), 3, and 4 hold with $\{\nu_j\}$ regularly varying. 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*

$$\mathsf{r}_n = (J^\circ)^{-p/d_x}, \ \textit{where } J^\circ := \max\left\{J : n^{-1/2} \nu_J^{-1} (J \log\log n)^{1/4} \leq J^{-p/d_x}\right\}. \tag{4.3}$$

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

$$\limsup_{n \to \infty} \mathrm{P}_{h_0}(\widehat{\mathtt{T}}_n = 1) \leq \alpha \qquad \textit{and} \qquad \lim_{n\to\infty} \sup_{h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)} \mathrm{P}_h(\widehat{\mathtt{T}}_n = 0) = 0. \tag{4.4}$$

(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 establish the 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 $\mathcal{H}$. 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 $g \in \mathcal{G}$.

Below we impose some conditions on the complexity of the closed and convex null class of functions $\mathcal{H}_0$. Let $\mathcal{S}^K = \{\mathsf{e} \in \mathbb{R}^K : \mathsf{e}_1^2 + \ldots + \mathsf{e}_K^2 = 1\}$ denote the $(K-1)$-dimensional unit sphere. Let $K^\circ = K(J^\circ)$, $\widetilde{b}^K(\cdot) = G_b^{-1/2} b^K(\cdot)$ and $Z := (X', W')'$. For any $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$, we consider the following class of functions

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

with its envelope function denoted by $F_{h,\mathsf{e}}$. Let $N_{[]}(\epsilon, \mathcal{F}, L^2(Z))$ be the $L^2(Z)$- covering number with bracketing for $\mathcal{F}$, which is the minimal number of ϵ-brackets, in $L^2(Z)$ sense, needed to cover $\mathcal{F}$. We let $\mathcal{C}_h := \max_{\mathsf{e}\in\mathcal{S}^{K^\circ}} \int_0^1 \left(1 + \log N_{[]}\left(\epsilon \|F_{h,\mathsf{e}}\|_{L^2(Z)}, \mathcal{F}_{h,\mathsf{e}}, L^2(Z)\right)\right)^{1/2} d\epsilon$.

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

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 $\mathcal{H}_1(\delta^\circ \mathsf{r}_n)$. Note that Assumption 5 imposes estimation rate conditions on $\widehat{h}_J^{\mathrm{R}}$ 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 $\mathcal{H}_0$.*

(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*

$$\max_{J\in\mathcal{I}_n} \frac{\zeta_J\|\widehat{h}_J - h\|_{L^2(X)}}{c_J} \leq const. \times \max_{J\in\mathcal{I}_n} \left\{ \frac{\zeta_J\sqrt{J}}{\sqrt{n}s_J c_J} + \frac{\zeta_J\|\Pi_J^{\mathcal{I}_n} h - h\|_{L^2(X)}}{c_J} \right\} \tag{4.5}$$

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 (4.5) 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* $h \in \mathcal{H}_0$*. Let* p_0 *denote the smoothness and* d_0 *the dimension of the nonparametric component under* $\mathcal{H}_0$*. If* $\|\Pi_J^{\mathcal{I}_n} h - h\|_{L^2(X)} = O(J^{-\mathrm{p}_0/\mathrm{d}_0})$ *and* $\zeta_J = O(\sqrt{J})$*, the second summand of the right hand side of (4.5) uniformly converges to zero if* $\mathrm{p}_0/\mathrm{d}_0 \geq 1/2$*. Since the class* $\mathcal{H}_0$ *is a less complex subset of* $\mathcal{H}$*, it is reasonable to assume that* $\mathrm{p}_0/\mathrm{d}_0 \geq p/d_x$ *and thus* $\mathrm{p}_0/\mathrm{d}_0 \geq 1/2$ *is automatically satisfied given Assumption 4(iii).*

Remark 4.2 (Primitive Conditions for Assumption 5(ii))**.** *Assumption 5(ii) restricts the complexity of* $\mathcal{H}_0$ *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* $\log N_{[]}(\epsilon, \mathcal{F}_{h,\mathsf{e}}, L^2(Z)) \leq \log N_{[]}(\epsilon, \mathcal{H}_{0,J^\circ}, L^\infty) \lesssim \epsilon^{-d_x/p}$ *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 (2.12).

Theorem 4.2. *Let Assumptions 1(i)-(iii), 2(i), 3, 4, and 5 hold with* $\{\nu_j\}$ *regularly varying. Consider testing the composite null* $\mathcal{H}_0$ *versus* $\mathcal{H}_1(\delta^\circ \mathrm{r}_n) = \{h \in \mathcal{H} : \|h - \mathcal{H}_0\|_{L^2(X)} \geq \delta^\circ \mathrm{r}_n\}$

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

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_0} \mathrm{P}_h(\widehat{\mathtt{T}}_n = 1) \le \alpha \qquad \text{and} \qquad \lim_{n\to\infty} \sup_{h\in\mathcal{H}_1(\delta^\circ \mathsf{r}_n)} \mathrm{P}_h(\widehat{\mathtt{T}}_n = 0) = 0. \tag{4.6}$$

(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. Moreover, (4.6) shows that $\widehat{\mathtt{T}}_n$ simultaneously has asymptotic size control over the composite null, and is consistent uniformly over the largest class of nonparametric alternatives $\mathcal{H}_1(\delta^\circ \mathsf{r}_n)$. The asymptotic size control is established by controlling the sieve approximation error uniformly over the index set $\widehat{\mathcal{I}}_n$ under the null, due to a projection property built in the construction of our test $\widehat{\mathtt{T}}_n$; see Lemma B.9. Theorem 4.2 is applicable to any composite null hypothesis $\mathcal{H}_0$ that is a closed convex strict subset of $\mathcal{H}$, including closed convex cone null restrictions as special cases.

Theorem 4.2 shows that our adaptive test has asymptotic size control and non-trivial power against a large class of nonparametric NPIV alternatives without using under-smoothed choice of sieve dimensions in testing. This is different from the existing non-adaptive tests for semiparametric or shape NPIV restrictions, which achieve asymptotic size controls via under-smoothed choice of tuning parameters in L^2 estimation. For instance, in their bootstrap test for convex cone restrictions of a NPIV function, Fang and Seo (2021) estimate the unrestricted NPIV function by a sieve 2SLS estimator assuming known smoothness, and chose the sieve dimension J deterministically such that the estimation bias J^{-p/d_x} is of a smaller order than the standard deviation $n^{-1/2}s_J^{-1}J^{1/2}$ in L^2 estimation, which leads to a non-adaptive rate of testing $n^{-1/2}s_J^{-1}J^{1/2}$ that is suboptimal for L^2 testing of NPIV models.

Remark 4.3. *Our adaptive minimax L^2 rate of testing $\left(\sqrt{\log\log n}/n\right)^{2p/(4(p+a)+d_x)}$ decreases to zero strictly faster than the optimal L^2 rate of estimation $n^{-p/(2(p+a)+d_x)}$ (even assuming known smoothness) for mildly ill-posed NPIV models, and coincides with the optimal L^2 rate of estimation $(\log n)^{-p/a}$ for severely ill-posed NPIV models. Therefore, any test statistic based on a tuning parameter chosen for the under-smoothed L^2 rate of NPIV estimation will not be as powerful as our new test uniformly over a large class of nonparametric alternatives.*

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 $\mathcal{H}_1(r_n) := \Big\{g \in \mathcal{G} : \|g - \mathcal{G}_0\|_{L^2(X_2)} \geq r_n\Big\}$ 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 $\mathcal{H} = \Theta \oplus \mathcal{G}$. The maintained partial linear structure can be easily enforced in the sieve space used to estimate the unconstrained NPIV function. For instance, we impose a partial linear structure $\mathcal{H}$ in our empirical illustration on demand for differential products in Section 6.1. 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.

4.3. Confidence Sets in L^2

One can construct L^2-confidence sets for a NPIV function by inverting our adaptive test. For any small $\alpha > 0$, the $(1-\alpha)-$confidence set for a NPIV function h belonging to a restricted nonparametric class $\mathcal{H}_0$ is given by

$$\mathcal{C}_n(\alpha) = \Big\{h \in \mathcal{H}_0 : \frac{n\widehat{D}_J(h)}{\widehat{V}_J} \leq \widehat{\eta}_J(\alpha) \quad \text{for all } J \in \widehat{\mathcal{I}}_n\Big\}. \tag{4.7}$$

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), 3, 4 and 5 hold. Let* $\mathrm{r}_n = (J^\circ)^{-p/d_x}$ *be the adaptive rate of testing given in Theorem 4.2. Then, for any* $\alpha \in (0,1)$ *it holds*

$$\limsup_{n\to\infty} \sup_{h\in\mathcal{H}_0} \mathrm{P}_h\left(h \notin \mathcal{C}_n(\alpha)\right) \leq \alpha \tag{4.8}$$

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

$$\lim_{n\to\infty} \inf_{h\in\mathcal{H}_1(\delta^\circ \mathrm{r}_n)} \mathrm{P}_h\left(h \notin \mathcal{C}_n(\alpha)\right) = 1. \tag{4.9}$$

Corollary 4.1 result (4.8) shows that the L^2-confidence set $\mathcal{C}_n(\alpha)$ controls size uniformly over the class of functions $\mathcal{H}_0$. Moreover, result (4.9) 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, $\mathsf{diam}(\mathcal{C}_n(\alpha)) = \sup\left\{\|h_1 - h_2\|_{L^2(X)} : h_1, h_2 \in \mathcal{C}_n(\alpha)\right\}$, depends on the degree of ill-posedness captured by the singular value s_{J°.

Corollary 4.2. *Let Assumptions 1(i)-(iii), 2(i), 3, 4, and 5 hold. Then, for any* $\alpha \in (0,1)$ *we have* $\sup_{h\in\mathcal{H}_0} \mathrm{P}_h\left(\mathit{diam}(\mathcal{C}_n(\alpha)) \geq C\mathsf{r}_n\right) = o(1)$, *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 $\mathcal{H}$ (the class of unrestricted NPIV functions).

5. Monte Carlo Studies

This section presents Monte Carlo performance of our adaptive test for monotonicity and parametric form of an NPIV function using simulation designs based on Chernozhukov, Newey, and Santos (2015). See the online Appendix C for additional simulation results using other designs. All the simulation results reported here are based on 5000 Monte Carlo replications for each experimental design and at $\alpha = 0.05$ nominal level. The simulation results clearly indicate that our simple adaptive test has size-control and finite-sample non-trivial power uniformly against a large class of NPIV alternatives, even for models with relatively weak instruments. In addition, simulation and real data application results reported in Breunig and Chen (2020), but not here due to the lack of space, have demonstrated that our adaptive test and its bootstrapped version perform similarly well in both finite-sample size and power.

For all the 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, Newey, and Santos (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 using (2.1) and (5.1) with h from the Chernozhukov, Newey, and Santos (2015) design:

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

This function $h(x)$ is decreasing in x, where c_0 captures the degree of monotonicity. We note that $c_0 = 0$ corresponds to $h(x) \equiv 0$ (the boundary case); $h(x) \approx 0$ 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 (2.12) 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 $K(J) \in \{2J,\ 4J,\ 8J\}$. Implementation of the restricted sieve NPIV estimator $\widehat{h}_J^{\mathrm{R}}$ is straightforward using the R package `coneproj`. We compare our adaptive test to the nonadaptive test of Fang and Seo (2021), which involves approximately computing $[n^{-1/2} s_J^{-1} J^{1/2}]^{-1} \min_{h \in \mathcal{H}_0} \|\widehat{h}_J - h\|_{L^2(X)}$ for a deterministic choice of sieve dimensions J and $K \geq J$ in their B-spline 2SLS estimate $\widehat{h}_J$. Their 2019 arXiv preprint presents a simulation study with $J = 3$, $K \in \{3,\ 4,\ 5\}$, and other tuning parameter choices $c_n = (\log n)^{-1}$ and $\gamma_n = 0.01/\log n$, such that their test achieves approximately empirical size control with a sample size $n = 500$. Below we use FS to denote their test with $J = 3$ and $K = 5$ (as $K = 5$ yields the best empirical power in their simulation), which is computed using R language translation of their Matlab program code. To study the sensitivity to the choice of K we also implement their test with $K = 12,\ 24$. In our simulations we implement their test using 200 bootstrap iterations.

Size. Table 1 presents the average data-driven choice of tuning parameter J, denoted by $\widehat{J}$. 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

n	c_0	ξ	$\widehat{\mathtt{T}}_n$	$\widehat{J}$	$\widehat{\mathtt{T}}_n$	$\widehat{J}$	$\widehat{\mathtt{T}}_n$	$\widehat{J}$	FS	FS	FS
			$K(J) = 2J$		$K(J) = 4J$		$K(J) = 8J$		$K = 5$	$K = 12$	$K = 24$
500	boundary	0.3	0.007	3.00	0.023	3.03	0.040	3.25	0.009	0.045	0.113
		0.5	0.020	3.29	0.025	3.35	0.039	3.41	0.041	0.059	0.095
		0.7	0.030	3.56	0.035	3.56	0.040	3.73	0.057	0.066	0.093
	0.01	0.3	0.006	3.00	0.021	3.03	0.038	3.25	0.008	0.040	0.103
		0.5	0.019	3.30	0.023	3.36	0.036	3.41	0.039	0.055	0.086
		0.7	0.029	3.57	0.033	3.58	0.037	3.75	0.046	0.057	0.080
	0.1	0.3	0.005	3.00	0.016	3.03	0.022	3.25	0.004	0.023	0.050
		0.5	0.013	3.33	0.018	3.38	0.025	3.43	0.019	0.026	0.038
		0.7	0.019	3.65	0.023	3.65	0.026	3.82	0.014	0.017	0.022
1000	boundary	0.3	0.009	3.01	0.019	3.06	0.032	3.30	0.013	0.037	0.079
		0.5	0.017	3.47	0.023	3.44	0.031	3.44	0.040	0.049	0.066
		0.7	0.029	3.84	0.034	3.93	0.040	3.95	0.052	0.058	0.075
	0.01	0.3	0.009	3.01	0.019	3.06	0.029	3.30	0.014	0.033	0.075
		0.5	0.017	3.48	0.023	3.45	0.030	3.44	0.038	0.045	0.060
		0.7	0.026	3.88	0.030	3.96	0.036	3.98	0.041	0.050	0.061
	0.1	0.3	0.006	3.02	0.013	3.06	0.019	3.30	0.008	0.019	0.038
		0.5	0.012	3.54	0.016	3.49	0.022	3.48	0.016	0.018	0.022
		0.7	0.017	4.02	0.019	4.09	0.024	4.10	0.008	0.008	0.010
5000	boundary	0.3	0.021	3.36	0.025	3.38	0.029	3.38	0.038	0.046	0.056
		0.5	0.033	3.54	0.034	3.60	0.041	3.79	0.051	0.055	0.057
		0.7	0.041	4.11	0.044	4.10	0.044	4.07	0.052	0.055	0.057
	0.01	0.3	0.020	3.36	0.024	3.39	0.028	3.39	0.037	0.043	0.052
		0.5	0.031	3.56	0.033	3.62	0.038	3.80	0.038	0.044	0.046
		0.7	0.038	4.18	0.039	4.17	0.039	4.14	0.034	0.036	0.039
	0.1	0.3	0.016	3.39	0.018	3.41	0.020	3.39	0.019	0.021	0.025
		0.5	0.022	3.68	0.022	3.73	0.027	3.91	0.008	0.009	0.008
		0.7	0.023	4.46	0.025	4.44	0.026	4.40	0.001	0.001	0.001

Table 1: Testing Monotonicity – Empirical size of our adaptive test $\widehat{\mathtt{T}}_n$ and of the FS test (with $J = 3$). Monte Carlo average value $\widehat{J}$. Nominal level $\alpha = 0.05$. True DGP from Section 5.1 using NPIV function (5.2). Instrument strength increases in ξ.

reports empirical rejection probabilities under the null hypothesis using our adaptive test $\widehat{\mathtt{T}}_n$ and the FS test. Our adaptive test is slightly under-sized across different sample sizes $n \in \{500, 1000, 5000\}$, different instrument strength $\xi \in \{0.3, 0.5, 0.7\}$, different degrees of monotonicity $c_0 \in \{0, 0.01, 0.1\}$ with $c_0 = 0$ being the "boundary" case. Table 1 shows that our adaptive test has empirical size control for all $K(J) \in \{2J,\ 4J,\ 8J\}$, which is in line with our theoretical results establishing asymptotic size control for any deterministic relation of $K \sim cJ$ for some fixed constant $c \geq 1$. The difference between the empirical size (of our adaptive test) for different choice of $K(J)$ is small when n is large or $\xi = 0.7$. While the FS test with $K = 5$ has empirical size control, the FS test with $K = 12$ can be slightly over-sized, and with $K = 24$ can be heavily over-sized for all ξ and $n = 500$, especially so for functions at or close to the boundary.[3]

[3]In our previous version (arXiv:2006.09587v4), we implemented what we called a *nonadaptive bootstrap test* $T^B_{n,3}$ of Fang and Seo (2021), which is essentially their test, but uses empirical root-mean squared metric instead of their trapezoid rule approximated $\|\widehat{h}_J - h\|_{L^2(X)}$, and a cone projection onto a J-dimensional sieve space instead of their optimization over grid points (for x). The "nonadaptive bootstrap test" $T^B_{n,3}$ has an empirical size closer to that of our adaptive test.

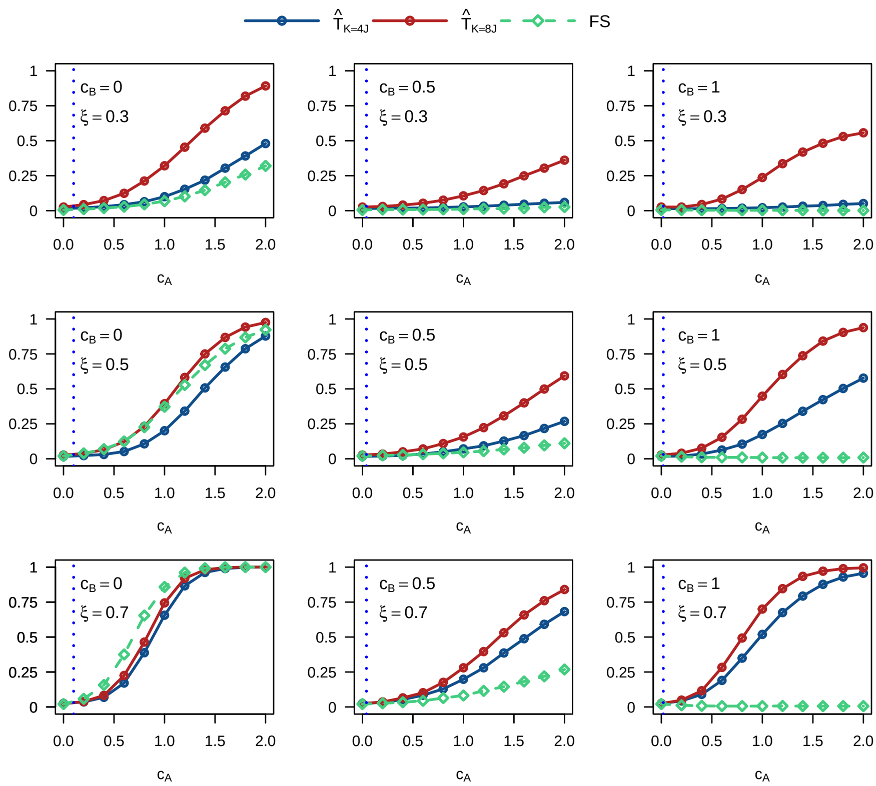


Figure 1: Testing Monotonicity – Empirical power of our adaptive test $\widehat{\mathtt{T}}_n$ with $K(J) = 4J$ (blue solid lines) and $K(J) = 8J$ (red solid lines) and the FS test (with $J = 3$, $K = 5$, green dashed lines). True DGP from Section 5.1 using NPIV function (5.3) with $n = 500$. The vertical 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.

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

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

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 $c_A \le 0.1$. When $c_B = 0.5$ the null hypothesis is satisfied only if $c_A \le 0.1/(1 + \pi/2) \approx 0.04$. When $c_B = 1$ the null is satisfied only if $c_A \le 0.1/(1 + \pi) \approx 0.02$.

Figure 1 depicts the empirical power function of our adaptive test $\widehat{\mathtt{T}}_n$ (blue solid lines for $K(J) = 4J$ and red solid lines for $K(J) = 8J$), and of the FS test (green dashed lines, $J = 3$, $K = 5$), under the 5% nominal level for different instrument strengths $\xi \in \{0.3, 0.5, 0.7\}$, and sample size $n = 500$.[4] 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 instrument strength ξ and the sample size n increase.

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

For weak instrument strength $\xi = 0.3$ and a small sample size (i.e., $n = 500$), our adaptive test with a larger $K(J) = 8J$ is more powerful.

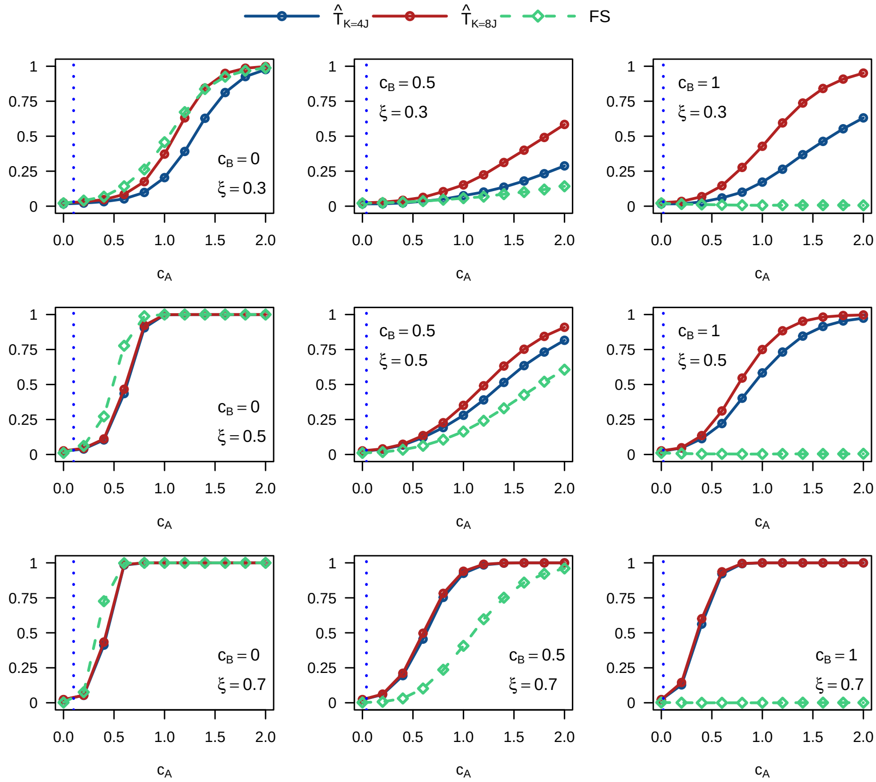


Figure 2: Testing Monotonicity – Replication of Figure 1 with $n = 5000$.

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., $c_B = 0$), there is little difference between our adaptive test $\widehat{\mathrm{T}}_n$ and the FS test. But, as the alternative becomes more nonlinear when $c_B > 0$ increases, the FS test becomes much less powerful than our adaptive test. This shows that a test with a tuning parameter J that is a deterministic nondecreasing function of n can be powerful in a certain direction but not for other nonlinear deviations.

In Online Appendix C we present another simulation design, which is based on an NPIV monotonicity design of Chetverikov and Wilhelm (2017). Simulation results using that design reveal that the empirical size and power of our adaptive test have patterns very similar to the ones reported in this subsection.

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 $c_B \in \{0, 0.5\}$. The null hypothesis is h being linear (i.e., $c_A = c_B = 0$).

We implement our adaptive test $\widehat{\mathtt{T}}_n$ given in (2.12) 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 $K(J) \in \{2J, 4J, 8J\}$. We compare our adaptive test to the asymptotic t-test and the test by Horowitz (2006) (denoted by JH).[5] 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 minimizing mean squared error of estimating f_{XW}.

n	ξ	$\widehat{\mathtt{T}}_n$, $K(J) = 2J$	$\widehat{J}$	$\widehat{\mathtt{T}}_n$, $K(J) = 4J$	$\widehat{J}$	$\widehat{\mathtt{T}}_n$, $K(J) = 8J$	$\widehat{J}$	t-test	JH test
500	0.3	0.008	3.00	0.021	3.03	0.040	3.29	0.001	0.049
	0.5	0.022	3.32	0.024	3.40	0.037	3.46	0.027	0.054
	0.7	0.036	3.61	0.037	3.63	0.035	3.81	0.045	0.058
1000	0.3	0.014	3.01	0.024	3.08	0.032	3.33	0.006	0.060
	0.5	0.025	3.52	0.033	3.49	0.033	3.48	0.043	0.060
	0.7	0.036	3.91	0.039	4.03	0.042	4.06	0.046	0.053
5000	0.3	0.022	3.38	0.029	3.41	0.037	3.43	0.032	0.057
	0.5	0.043	3.58	0.048	3.65	0.045	3.85	0.050	0.061
	0.7	0.050	4.17	0.051	4.15	0.050	4.14	0.049	0.055

Table 2: Testing Parametric Form – Empirical size of our adaptive test $\widehat{\mathtt{T}}_n$, the t-test and JH test. Monte Carlo average value $\widehat{J}$. Nominal level $\alpha = 0.05$. 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 of several tests under the null hypothesis of linearity of h. Results are presented under different sample sizes $n \in \{500, 1000, 5000\}$ and instrument strength $\xi \in \{0.3, 0.5, 0.7\}$. It also reports our adaptive test with different $K(J)$ and $\widehat{J}$ (which is defined the same way as that in Table 1). We note that $\widehat{J}$ is again weakly increasing with sample size and with instrument strength. While the JH test can be slightly over-sized, our adaptive test $\widehat{\mathtt{T}}_n$ provides adequate size control across different sample size n, different instrument strength ξ and different $K(J)$. The difference in empirical size of our adaptive test with different $K(J)$ is again small for large n, which is consistent with our theory.

Power. Figure 3 provides empirical power curves for the 5% level tests with sample sizes $n \in \{500, 5000\}$. From this figure, we see that our adaptive test $\widehat{\mathtt{T}}_n$ (blue solid lines with $K(J) = 4J$ and red solid lines with $K(J) = 8J$) has power similar to the asymptotic t-test (yellow 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 nonlinear/complex with $c_B = 0.5$, our adaptive test becomes more powerful than the JH test. This is theoretically sensible since Horowitz (2006) test is designed to have power

[5] 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.

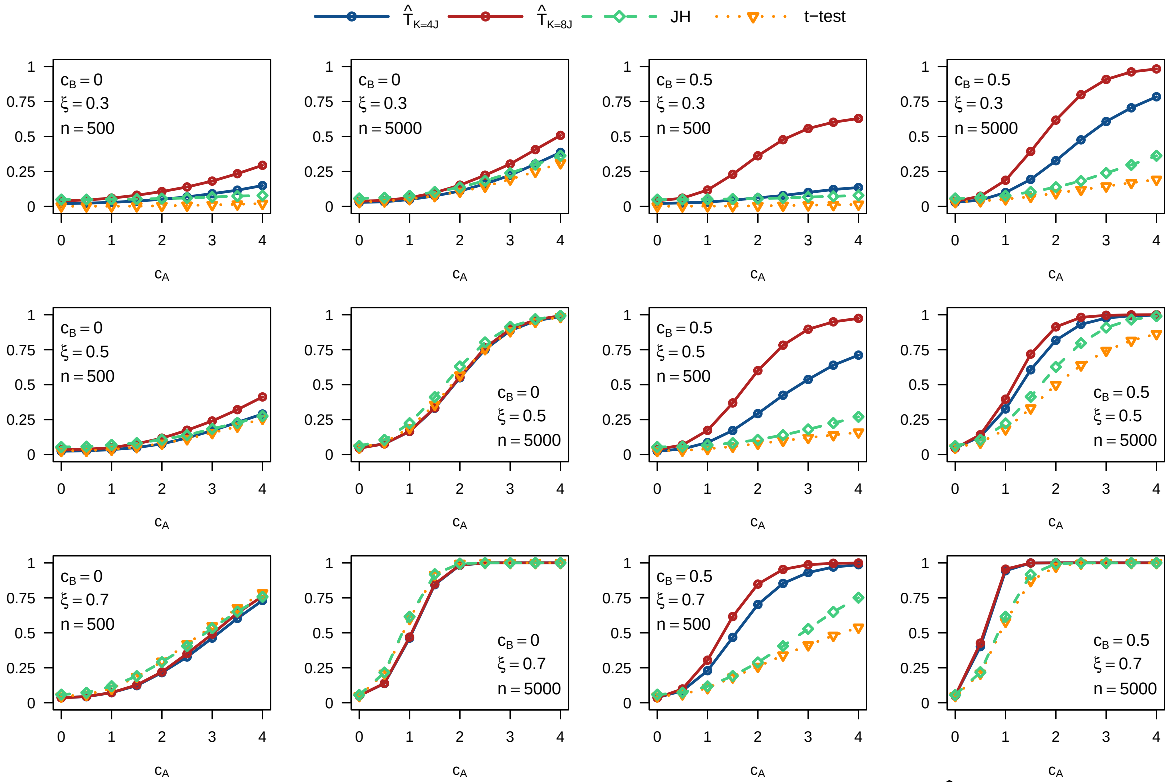


Figure 3: Testing Parametric Form – Empirical power of our adaptive test $\widehat{\mathtt{T}}_n$ with $K(J) = 4J$ (blue solid lines) and $K(J) = 8J$ (red solid lines), of JH test (green dashed lines) and of t-test (yellow dotted lines). True DGP from Section 5.2 using NPIV function (5.3). Alternatives are quadratic when $c_B = 0$ and more complex for $c_B = 0.5$.

against $n^{-1/2}$ smooth alternative only. Since our adaptive test is slightly undersized for small sample sizes or for weak instrument strength, the size-adjusted empirical power of our test is even better (see our arXiv:2006.09587v3 version). 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. We observe that our adaptive test $\widehat{\mathtt{T}}_n$ again has 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. The applications demonstrate that our simple adaptive test is powerful to detect economic shape restrictions.

In both empirical applications, we implement our adaptive test $\widehat{\mathtt{T}}_n$ given in (2.12) 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). Let $\widehat{\mathcal{J}}$ be $\{J \in \widehat{\mathcal{I}}_n : \widehat{\mathcal{W}}_J(\alpha) > 1\}$ when our test rejects the null, and be $\arg\max_{J \in \widehat{\mathcal{I}}_n} \widehat{\mathcal{W}}_J(\alpha)$ when our test fails to reject the null. Let $\widehat{J}$ be the minimal integer of $\widehat{\mathcal{J}} \subset \widehat{\mathcal{I}}_n$. Tables in this section report $\widehat{\mathcal{J}}$ and $\widehat{\mathcal{W}}_{\widehat{J}}$. 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

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 (2022) 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 (2022)'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 (2022) directly imposes "connected substitutes" restriction in his sieve NPIV estimation of inverse demand, we want to test this restriction. Following Compiani (2022) 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 (2022) 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 (2022, 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 use the data set of Compiani (2022)[6], 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 RES index set $\widehat{\mathcal{I}}_n = \{22, 28, 34\}$.

H_0	$\widehat{\mathcal{W}}_{\widehat{J}}$	p val.	reject H_0?	$\widehat{J}$
$\partial h/\partial p_o \geq 0$	0.854	0.031	no	$\{34\}$
$\partial h/\partial p_o \leq 0$	3.154	0.000	yes	$\{28, 34\}$
$\partial h/\partial s_o \geq 0$	0.661	0.057	no	$\{34\}$
$\partial h/\partial s_o \leq 0$	2.022	0.001	yes	$\{22, 28, 34\}$
$\partial h/\partial s_{no} \geq 0$	-0.115	0.471	no	$\{22\}$
$\partial h/\partial s_{no} \leq 0$	-0.238	0.734	no	$\{22\}$
$\partial h/\partial s_o \geq \partial h/\partial s_{no}$	0.663	0.057	no	$\{34\}$
$\partial h/\partial s_o \leq \partial h/\partial s_{no}$	2.022	0.001	yes	$\{22, 28, 34\}$

Table 3: Adaptive testing for the shape of h (the inverse demand for organic produce).

According to Table 3, at the nominal level $\alpha = 0.05$, our adaptive test fails to reject that h is weakly increasing in the own price (but rejects $\partial h/\partial p_o \leq 0$), and fails to reject that h is weakly increasing in the own share (but rejects $\partial h/\partial s_o \leq 0$). Our test fails to reject that h is weakly increasing or decreasing in the non-organic share (i.e., fails to reject a constant partial effect of h with respect to the non-organic share). Our test also fails to reject the diagonal dominance (but rejects $\partial h/\partial s_o \leq \partial h/\partial s_{no}$). In summary, our adaptive test provides strong empirical evidence for the connected substitutes 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

[6]For details on the construction of the data and descriptive statistics, see Compiani (2022, Appendix F).

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, Blundell, and Lewbel (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.

Blundell, Chen, and Kristensen (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 RES 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	$\widehat{\mathcal{W}}_{\hat{\mathcal{J}}}$	p value	reject H_0?	$\hat{\mathcal{J}}$	$\widehat{\mathcal{W}}_{\hat{\mathcal{J}}}$	p value	reject H_0?	$\hat{\mathcal{J}}$
"food in"	2.871	0.000	yes	$\{3\}$	-0.324	0.852	no	$\{4\}$
"fuel"	8.192	0.000	yes	$\{3, 4, 5\}$	0.547	0.072	no	$\{3\}$
"travel"	2.527	0.000	yes	$\{3, 4\}$	0.381	0.124	no	$\{3\}$
"leisure"	0.299	0.165	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	$\widehat{\mathcal{W}}_{\hat{\mathcal{J}}}$	p value	reject H_0?	$\hat{\mathcal{J}}$	$\widehat{\mathcal{W}}_{\hat{\mathcal{J}}}$	p value	reject H_0?	$\hat{\mathcal{J}}$
"food in"	-0.287	0.791	no	$\{4\}$	-0.324	0.853	no	$\{3\}$
"fuel"	-0.325	0.844	no	$\{3\}$	1.621	0.001	yes	$\{3\}$
"travel"	1.188	0.007	yes	$\{3\}$	-0.322	0.837	no	$\{5\}$
"leisure"	-0.197	0.656	no	$\{5\}$	0.691	0.047	no	$\{4\}$

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

	H_0: h is linear				H_0: h is quadratic			
Goods	$\widehat{\mathcal{W}}_{\hat{\mathcal{J}}}$	p value	reject H_0?	$\hat{\mathcal{J}}$	$\widehat{\mathcal{W}}_{\hat{\mathcal{J}}}$	p value	reject H_0?	$\hat{\mathcal{J}}$
"food in"	-0.273	0.781	no	$\{3\}$	0.125	0.272	no	$\{3\}$
"fuel"	1.623	0.001	yes	$\{3\}$	-0.120	0.540	no	$\{5\}$
"travel"	1.210	0.006	yes	$\{3\}$	-0.014	0.407	no	$\{4\}$
"leisure"	0.691	0.047	no	$\{4\}$	0.513	0.086	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 functions together with their non-adaptive 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, statistically 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, Blundell, and Lewbel (1997) using Student t-test for linear against quadratic forms of Engel curves.

References

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

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

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

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

Blundell, R., X. Chen, and D. Kristensen (2007): "Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves," *Econometrica*, 75(6), 1613–1669.

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

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

Breunig, C., and X. Chen (2020): "Adaptive, Rate-Optimal Testing in Instrumental Variables Models," *arXiv preprint arXiv:2006.09587v1*.

Chen, X., T. Christensen, and S. Kankanala (2024): "Adaptive Estimation and Uniform Confidence Bands for Nonparametric Structural Functions and Elasticities," *The Review of Economic Studies, forthcoming*.

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

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

Chen, X., and M. Reiss (2011): "On Rate Optimality for Ill-Posed Inverse Problems in Econometrics," *Econometric Theory*, 27(03), 497–521.

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

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

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

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

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

Compiani, G. (2022): "Market Counterfactuals and the Specification of Multi-Product Demand: A Nonparametric Approach," *Quantitative Economics*, 13, 545–591.

Deaton, A., and J. Muellbauer (1980): "An almost ideal demand system," *American Economic Review*, pp. 312–326.

Fang, Z., and J. Seo (2021): "A Projection Framework for Testing Shape Restrictions That Form Convex Cones," *Econometrica*, 89(5), 2439–2458.

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

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

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

Horowitz, J. L. (2006): "Testing a Parametric Model Against a Nonparametric Alternative with Identification Through Instrumental Variables," *Econometrica*, 74(2), 521–538.

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

Horowitz, J. L., and V. G. Spokoiny (2001): "An Adaptive, Rate-Optimal Test of a Parametric Mean-Regression Model against a Nonparametric Alternative," *Econometrica*, 69(3), 599–631.

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

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

Santos, A. (2012): "Inference in Nonparametric Instrumental Variables With Partial Identification," *Econometrica*, 80(1), 213–275.

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

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

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

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

A. Proofs of Theorems 3.1 and 3.2 in Section 3

Proof of Theorem 3.1. We first derive the lower bound for testing a simple null hypothesis $\mathcal{H}_0 = \{h_0\}$. 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 $h_0 = 0$ and $\mathcal{H}_0 = \{0\}$. We introduce $\theta = (\theta_j)_{j\geq 1}$ with $\theta_j \in \{-1, 1\}$ and introduce the test function

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

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

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

Therefore, orthonormality of the basis functions $\{\widetilde{\psi}_j\}_{j\geq 1}$ in $L^2(X)$ together with the Cauchy-Schwarz inequality implies for any $\theta \in \{\pm 1\}^J$ with any $J \geq J_*$:

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

for all $\delta_* \in (0, 1]$, and thus, we conclude that $h_\theta \in \mathcal{H}$ by the definition of the Sobolev ellipsoid $\mathcal{H}$. For any $\theta \in \{\pm 1\}^{J_*}$ we have

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

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 $\{\pm 1\}^{J_*}$ and P_0 the probability distribution when $h_\theta = 0$. From the proof of Collier, Comminges, and Tsybakov (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 $\theta \in \{\pm 1\}^{J_*}$, we thus evaluate

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

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

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

By Tsybakov (2009, Section 2.7.5) there exists a constant $c_1 > 0$ such that $\exp(-n\gamma_k^2/\sigma^2) + \exp(n\gamma_k^2/\sigma^2) \le 2\exp\left(c_1 n^2 \gamma_k^4\right)$. Assumption 1(iv) implies for some constant $c > 0$ that $\sum_{k=1}^{J_*} \langle T\phi_{J_*}, \widetilde{b}_k\rangle_{\mathcal{W}}^4 \le c \sum_{j\ge 1} \nu_j^4 \langle \phi_{J_*}, \widetilde{\psi}_j\rangle_X^4 = c$ and we thus obtain by the definition of γ_k:

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

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 $J_* \sim \max\left\{J : n^{-1/2} J^{1/4} \nu_J^{-1} \le J^{-p/d_x}\right\}$, we note that inequality(A.2) holds within a constant and we have $r_n = n^{-1/2}\left(\sum_{j=1}^{J_*} \nu_j^{-4}\right)^{1/4} \sim n^{-1/2} J_*^{1/4} \nu_{J_*}^{-1} \sim J_*^{-p/d_x}$. 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 $r_n \sim (\log n)^{-p/a}$, which completes the proof for the simple null $\mathcal{H}_0 = \{0\}$ case.

We now turn to the lower bound for testing a closed convex composite null hypothesis. Consider the test function given in equation (A.1). Since $\mathcal{H}_0$ 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

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

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 $\|h_\theta - \mathcal{H}_0\|_{L^2(X)} \ge \|h_{\theta_*}\|_{L^2(X)} \ge \delta_* r_n$, by following inequality (A.3). Therefore, we may proceed with the proof of the lower bound as for the simple null case. □

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

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

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): $\limsup_{n\to\infty} \mathrm{P}_{h_0}(\mathtt{T}_{n,J} = 1) = \limsup_{n\to\infty} \mathrm{P}_{h_0}\left(n\widehat{D}_J(h_0) > \eta_J(\alpha)\widehat{V}_J\right) \le \alpha$. To control the type II error, we have uniformly for $h \in \mathcal{H}_1(\delta^\circ r_{n,J})$,

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

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. Note that $\nu_j^{-2}\ge cs_J^{-2}$ by Assumption 3. Given the definition of J_{*0}, the final 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. □

B. Proofs of Theorems 4.1 and 4.2 in Section 4

We first introduce additional notation. For a $r\times c$ matrix M with $r\le c$ 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 $A=(G_b^{-1/2}SG^{-1/2})_l^-G_b^{-1/2}$. Then $\|AG_b^{1/2}\|=\|(G_b^{-1/2}SG^{-1/2})_l^-\|=s_J^{-1}$ with $s_J=s_{\min}(G_b^{-1/2}SG^{-1/2})>0$. 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:

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

We next present Theorem B.1 and eight lemmas (Lemma B.1 - Lemma B.8) 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) $\|Q_Jh-h\|_{L^2(X)}=O(J^{-p/d_x})$.*

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

Lemma B.3. *Let Assumption 1(i) hold. Then: $J\le\sum_{j=1}^J s_j^{-4}\le\underline{\sigma}^{-4}V_J^2$ 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

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

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

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

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

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

$R_1(Z_i, Z_{i'}) := R(Z_i, Z_{i'}, M_i)$ 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

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

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

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:*

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

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

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

Lemma B.7. *(i) Under the conditions of Theorem 3.2 we have for some constant $c_0 > 0$ that $\mathrm{P}_h\big(n\widehat{D}_J(h_0) \leq (1 + c_0)\eta_J(\alpha)V_J\big) = o(1)$ uniformly for $h \in \mathcal{H}_1(\delta^\circ r_{n,J})$.*

(ii) Under the conditions of Theorem 4.1 we have $\mathrm{P}_h\big(n\widehat{D}_{J^*}(h_0) \leq 2c_1\sqrt{\log\log n}\, V_{J^*}\big) = o(1)$ *uniformly for* $h \in \mathcal{H}_1(\delta^\circ \mathrm{r}_n)$*, 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* (2.11) *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 $\widetilde{\mathtt{T}}_n = \mathbb{1}\big\{\max_{J\in\mathcal{I}_n}\big(n\widehat{D}_J(h_0)/(\eta_J'(\alpha)V_J)\big) > 1\big\}$, $\eta_J'(\alpha) := (1-c_0)\sqrt{\log\log J - \log\alpha}/4$ 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 $\mathcal{U}_{J,l} = \frac{2}{n(n-1)}\sum_{1\leq i<i'\leq n} R_l(Z_i, Z_{i'})$ for $l = 1, 2$ and $U_i = Y_i - h_0(X_i)$:

$$\mathrm{P}_{h_0}\big(\widetilde{\mathtt{T}}_n = 1\big) \leq \mathrm{P}_{h_0}\Big(\max_{J\in\mathcal{I}_n}\Big|\frac{1}{\eta_J'(\alpha)V_J(n-1)}\sum_{j=1}^{J}\sum_{i\neq i'} U_{ij}U_{i'j}\Big| + \max_{J\in\mathcal{I}_n}\Big|\frac{1}{\eta_J'(\alpha)V_J(n-1)}\sum_{i\neq i'} U_iU_{i'}b^K(W_i)'\Big(A'A - \widehat{A}'\widehat{A}\Big)b^K(W_{i'})\Big| > 1\Big) \leq I + II + III,$$

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

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

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 $J \in \mathcal{I}_n$,

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

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:

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

by the definition of $\eta_J'(\alpha)$. Given $s_J^{-2}\sqrt{J} \sim V_J$ by Assumption 4(ii), Lemmas B.2 and B.3

imply uniformly for $h \in \mathcal{H}$ and $J \in \mathcal{I}_n$: $V_L/V_J \lesssim s_L^{-2} s_J^2 \sqrt{L/J} = o(1)$ for all $L = o(J)$. Thus, for all $J \in \mathcal{I}_n$ and for n sufficiently large: $\Lambda(u_J, L(J)) \leq C_{R_1} \frac{n-1}{8} V_J \eta_J'(\alpha)$ 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:

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

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

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

where the fourth moment of $U = Y - h_0(X)$ is bounded under Assumption 2(i). Lemma B.3 implies $s_J^{-2} \leq \underline{\sigma}^{-2} V_J$. By the definition of $M_n = \sqrt{n}\, \zeta_{\overline{J}}^{-1} (\log\log \overline{J})^{-3/4}$ 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 $\overline{J} = \sup\{J : s_J^{-1} \zeta_J^2 \sqrt{(\log J)/n} \leq \bar{c}\}$ we infer that the dimension parameter J° given in (4.3) satisfies $\underline{J} \leq J^\circ \leq \overline{J}/2$ for $\bar{c}$ sufficiently large by Assumption 3 and 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 $h \in \mathcal{H}_1(\delta^\circ \mathtt{r}_n)$

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

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

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

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 $h \in \mathcal{H}_1(\delta^\circ \mathsf{r}_n)$:

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

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

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 $\widetilde{\mathtt{T}}_n = \mathbb{1}\big\{\max_{J\in\mathcal{I}_n}\big\{n\widehat{D}_J/(\eta_J'(\alpha)V_J)\big\} > 1\big\}$, 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

$$
\begin{aligned}
n(n-1)\,\widehat{D}_J &= \sum_{i\neq i'} \big(Y_i - \widehat{h}^{\mathrm{R}}_J(X_i)\big)\big(Y_{i'} - \widehat{h}^{\mathrm{R}}_J(X_{i'})\big) b^K(W_i)'\widehat{A}'\widehat{A} b^K(W_{i'}) \\
&= \Big\| \sum_i \big(Y_i - \widehat{h}^{\mathrm{R}}_J(X_i)\big)\widehat{A} b^K(W_i)\Big\|^2 - \sum_i \Big\|\big(Y_i - \widehat{h}^{\mathrm{R}}_J(X_i)\big)\widehat{A} b^K(W_i)\Big\|^2.
\end{aligned}
$$

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 (2.6) yields for all $h \in \mathcal{H}_0$:

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

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

$$
\begin{aligned}
\frac{n\widehat{D}_J}{\eta'_J(\alpha)V_J} - \frac{n\widehat{D}_J(h)}{\eta'_J(\alpha)V_J} &\lesssim (V_J\sqrt{(\log\log J)/J})^{-1/2} n^{-1} \sum_i \big(Y_i - h(X_i)\big) b^K(W_i)'\widehat{A}'\widehat{A} b^K(W_i)\big(\widehat{h}^{\mathrm{R}}_J - h\big)(X_i) \\
&\quad + (V_J\sqrt{(\log\log J)/J})^{-1/2} \Big\| \frac{1}{\sqrt{n}} \sum_i \big(Y_i - h(X_i)\big)\widehat{A} b^K(W_i)\Big\| \\
&=: (V_J\sqrt{(\log\log J)/J})^{-1/2}\big(T_{1,J} + 2T_{2,J}\big)
\end{aligned}
$$

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

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

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

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

Consider $T_{11,J}$. We first observe by the Cauchy–Schwarz 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

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

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

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

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., $\mathrm{P}_h(\max_{J\in\mathcal{I}_n}\|\widehat{h}^{\mathrm{R}}_J-h\|_{L^2(X)}\zeta_J/(\log\log J)^{1/4} > \varepsilon) \to 0$ uniformly for $h\in\mathcal{H}_0$ for any $\varepsilon>0$. Similarly, $\max_{J\in\mathcal{I}_n} T_{12,J}/(V_J\sqrt{\log\log J})$ 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

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

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

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

uniformly for $h\in\mathcal{H}_0$. Concerning the second summand $T_{22,J}$, by another application of Lemma E.5, $\max_{J\in\mathcal{I}_n}T_{22,J}/(V_J\sqrt{J(\log\log J)})$ 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 $h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)$ that

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

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

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

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

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

where $\rho_h = \|h-\mathcal{H}_0\|^2_{L^2(X)}/2 - 2c_1 n^{-1}\sqrt{\log\log n}V_{J^*} - C_B\mathsf{r}_n^2$. 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 $\|h-\mathcal{H}_0\|^2_{L^2(X)} \geq (\delta^\circ)^2\mathsf{r}_n^2 \geq 5c_1n^{-1}V_{J^*}\sqrt{\log\log n}$ 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

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

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 $h\in\mathcal{H}_1(\delta^\circ\mathsf{r}_n)$: $\mathrm{P}_h\left(\widehat{\mathtt{T}}_n = 0\right) \leq \mathrm{P}_h\big(n\widehat{D}_{J^*} \leq \eta''(\alpha)V_{J^*}\big) + \mathrm{P}_h\left(\Big|\widehat{V}_{J^*}/V_{J^*} - 1\Big| > c_0\right) + o(1) =$

$o(1)$, 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*

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

Proof of Lemma B.9. 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

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

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 $\mathrm{P}_h(\max_{J\in\mathcal{I}_n}T_{31,J}/(V_J\sqrt{\log\log J})>\varepsilon)=o(1)$ uniformly for $h\in\mathcal{H}_0$. Consider $T_{32,J}$. For any $\varepsilon>0$, we have $\mathrm{P}_h(\max_{J\in\mathcal{I}_n}T_{32,J}/(V_J\sqrt{\log\log J})>\varepsilon)=o(1)$ 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 $\mathrm{P}_h\big(\max_{J\in\mathcal{I}_n}T_{42,J}/(V_J\sqrt{\log\log J})>\varepsilon\big)=o(1)$ uniformly for $h\in\mathcal{H}_0$ and thus it is sufficient to consider $T_{41,J}$. We have

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

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 $s_J^{-2}\le\underline{\sigma}^{-2}V_J$ (by Lemma B.3). $\square$