ADAPTIVE ESTIMATION AND UNIFORM CONFIDENCE BANDS FOR NONPARAMETRIC IV

By

Xiaohong Chen, Timothy Christensen, and Sid Kankanala

July 2021

COWLES FOUNDATION DISCUSSION PAPER NO. 2292

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.yale.edu/

Adaptive Estimation and Uniform Confidence Bands for Nonparametric IV*

Xiaohong Chen† Timothy Christensen‡ Sid Kankanala§

July 24, 2021

Abstract

We introduce computationally simple, data-driven procedures for estimation and inference on a structural function h_0 and its derivatives in nonparametric models using instrumental variables. Our first procedure is a bootstrap-based, data-driven choice of sieve dimension for sieve nonparametric instrumental variables (NPIV) estimators. When implemented with this data-driven choice, sieve NPIV estimators of h_0 and its derivatives are adaptive: they converge at the best possible (i.e., minimax) sup-norm rate, without having to know the smoothness of h_0, degree of endogeneity of the regressors, or instrument strength. Our second procedure is a data-driven approach for constructing honest and adaptive uniform confidence bands (UCBs) for h_0 and its derivatives. Our data-driven UCBs guarantee coverage for h_0 and its derivatives uniformly over a generic class of data-generating processes (honesty) and contract at, or within a logarithmic factor of, the minimax sup-norm rate (adaptivity). As such, our data-driven UCBs deliver asymptotic efficiency gains relative to UCBs constructed via the usual approach of undersmoothing. In addition, both our procedures apply to nonparametric regression as a special case. We use our procedures to estimate and perform inference on a nonparametric gravity equation for the intensive margin of firm exports and find evidence against common parameterizations of the distribution of unobserved firm productivity.

Keywords: Honest and adaptive uniform confidence bands, minimax sup-norm rate-adaptive estimation, nonparametric instrumental variables, bootstrap.

JEL codes: C13, C14, C36

*We are grateful to Enno Mammen, Richard Nickl, and Yixiao Sun for helpful suggestions, and to Rodrigo Adão, Costas Arkolakis, and Sharat Ganapati for sharing their data. The data-driven choice of sieve dimension in this paper is based on and supersedes Section 3 of the preprint `arXiv:1508.03365v1` (Chen and Christensen, 2015a). The research is partially supported by the Cowles Foundation Research Funds (Chen) and the National Science Foundation under Grant No. SES-1919034 (Christensen).

†Cowles Foundation for Research in Economics, Yale University. `xiaohong.chen@yale.edu`

‡Department of Economics, New York University. `timothy.christensen@nyu.edu`

§Department of Economics, Yale University. `sid.kankanala@yale.edu`

1 Introduction

In this paper, we propose computationally simple, data-driven procedures for choosing tuning parameters when estimating and constructing uniform confidence bands (UCBs) for a nonparametric structural function h_0 satisfying

$$Y = h_0(X) + u, \qquad \mathbb{E}[u|W] = 0 \text{ (almost surely)}, \tag{1}$$

where X is a vector of regressors, W is a vector of (conditional) instrumental variables, and the conditional distribution of X given W is unspecified. We allow for the possibility that some elements of X are endogenous and hence that $\mathbb{E}[u|X] \neq 0$ with positive probability. Model (1) nests nonparametric regression as a special case with $W = X$ and $h_0(x) = \mathbb{E}[Y|X = x]$.

As endogenous regressors are frequently encountered in applied work in economics, there are already many theoretical results on nonparametric instrumental variables (NPIV) estimation of h_0 in model (1).[1] To implement any NPIV estimator and construct UCBs for h_0 in practice, researchers must choose various tuning (or regularization) parameters. To date there is no work on sup-norm rate-adaptive procedures for data-driven choice of tuning parameters for any nonparametric estimator of h_0 or its derivatives, nor are there data-driven procedures for constructing UCBs for h_0 or its derivatives. In this paper, we fill these gaps for sieve NPIV estimators of h_0, which are simply two-stage least-squares estimators applied to basis functions of X and W. Sieve NPIV estimators are very easy to compute and have been used in empirical work across several fields of economics.[2]

Our first procedure is a bootstrapped-based, data-driven choice for the number of basis functions J used to approximate h_0 (i.e., the sieve dimension), which is the key tuning parameter to be chosen for sieve NPIV estimators $\hat{h}_J$. When implemented with our data-driven choice $\tilde{J}$, the resulting estimators $\hat{h}_{\tilde{J}}$ and their (partial) derivatives $\partial^a \hat{h}_{\tilde{J}}$ converge at the minimax sup-norm rates. That is, the maximal estimation errors over the support of X:

$$\sup_x |\hat{h}_{\tilde{J}}(x) - h_0(x)| \quad \text{and} \quad \sup_x |\partial^a \hat{h}_{\tilde{J}}(x) - \partial^a h_0(x)|\,,$$

vanish as fast as possible—among *all* estimators of h_0 and its derivatives $\partial^a h_0$—as the sample size increases, uniformly over a class of data-generating processes, for both nonparametric regression and NPIV models. Following the statistics literature, we refer to this data-driven procedure as *sup-norm rate-adaptive*: the estimators adapt to features of the data-generating

[1] Early publications on NPIV estimation include Newey and Powell (2003), Hall and Horowitz (2005), Blundell, Chen, and Kristensen (2007), Darolles, Fan, Florens, and Renault (2011), and Horowitz (2011).

[2] Some examples include the analysis of household demand (Blundell et al., 2007; Blundell, Horowitz, and Parey, 2017), demand for differentiated products (Compiani, 2020), and international trade (Adão, Arkolakis, and Ganapati, 2020).

process that the researcher does not know ex ante, including the smoothness of h_0 and the strength of the instruments, to converge at the optimal sup-norm rate. Simulation studies reveal that the resulting sieve estimator $\hat{h}_{\tilde{J}}$ is accurate even when h_0 is highly nonlinear.

Sup-norm rate-adaptive procedures for choosing tuning parameters are particularly important in NPIV estimation, as the performance of nonparametric estimators of h_0 in model (1) is known to be more sensitive to tuning parameters than in nonparametric regression. It is therefore attractive to practitioners that one single data-driven choice of sieve dimension $\tilde{J}$ attains the best possible sup-norm convergence rates when estimating both h_0 and $\partial^a h_0$ using $\hat{h}_{\tilde{J}}$ and $\partial^a \hat{h}_{\tilde{J}}$. Sup-norm rate-adaptive procedures are also useful for several inference problems. For instance, they are inputs to our data-driven UCBs. They are also useful for ensuring that remainder terms are asymptotically negligible when performing inference on nonlinear functionals of h_0 and its derivatives via sample splitting and other two-step inference methods.

Our second data-driven procedure is for constructing UCBs for the structural function h_0 and its derivatives. Here we use the term "uniform" to indicate that the entire function lies within the confidence bands with the desired asymptotic coverage probability. Our data-driven UCBs are centered at the adaptive estimators $\hat{h}_{\tilde{J}}$ and $\partial^a \hat{h}_{\tilde{J}}$ and have their widths determined by a critical value that accounts for bias and sampling uncertainty. We show that our data-driven UCBs are *honest*, in the sense that they guarantee coverage for h_0 and its derivatives uniformly over a generic class of data-generating processes, and *adaptive*, in the sense that they contract at, or within a logarithmic factor of, the minimax sup-norm rate. Moreover, though not the focus of our paper, our UCBs for derivatives of h_0 provide an alternative, fully data-driven approach to testing certain shape restrictions. For instance, (strict) monotonicity can be tested by checking whether the UCB for the derivative of h_0 lies uniformly above or below zero.

A recent literature on (non data-driven) UCBs for h_0 and functionals thereof in NPIV models relies on *undersmoothing* to guarantee asymptotically valid inference.[3] That is, tuning parameters are assumed to be chosen (deterministically) in the hope that bias is of smaller order than sampling uncertainty. However, undersmoothing requires prior knowledge of model features such as the smoothness of h_0 and instrument strength, which are typically unknown in real data applications. Undersmoothing is also asymptotically inefficient, in the sense that undersmoothed UCBs are unnecessarily wide, as it uses a sub-optimal choice of tuning parameter. This issue is particularly important for NPIV models, as the variance of sieve NPIV estimators increases much faster in the sieve dimension J than the variance of sieve nonparametric regression estimators. Indeed, the maximal width of undersmoothed UCBs relative to ours will become infinite as the sample size goes to infinity for the class of models for which our bands contract at the minimax rate. Simulation studies demonstrate the efficiency gains

[3]See, for example, Horowitz and Lee (2012), Chen and Christensen (2018) and Babii (2020).

of our data-driven UCBs relative to undersmoothed UCBs in finite samples.

To illustrate the practical use of our data-driven procedures, we estimate a nonparametric gravity equation of Adão, Arkolakis, and Ganapati (2020) for the intensive margin of firm exports and construct UCBs for the function and its elasticity. Our estimates and UCBs provide evidence against common parameterizations of the distribution of unobserved firm productivity (Chaney, 2008; Eaton, Kortum, and Kramarz, 2011; Head, Mayer, and Thoenig, 2014; Melitz and Redding, 2015) in monopolistic competition models of international trade. Though we do not impose monotonicity a priori, our estimated function is monotone and our UCBs for the function and its elasticity are narrow and informative. In particular, we reject a constant elasticity specification for the intensive margin of firm exports.

Related Literature. Our first procedure is inspired by the bootstrap-based implementation of Lepski's method of Chernozhukov, Chetverikov, and Kato (2014) for density estimation and Spokoiny and Willrich (2019) for linear regression with Gaussian errors. However, our procedure does not follow easily from these existing procedures due to several challenges present in nonparametric models with endogeneity. In particular, the structural function h_0 in (1) is identified by inverting the conditional moment restriction

$$\mathbb{E}[Y|W] = \mathbb{E}[h_0(X)|W] \text{ (almost surely).}$$

The degree of difficulty of inverting $\mathbb{E}[h_0(X)|W]$ to recover h_0 is a nonparametric notion of instrument strength and plays an important role in determining the best possible convergence rates for estimators of h_0 and $\partial^a h_0$.[4] While adaptive procedures for nonparametric density estimation or regression deal only with unknown smoothness of the estimand, our data-driven procedures must also deal with uncertainty about the degree of difficulty of the inversion problem. This is important, as the NPIV literature has typically classified the difficulty of the inversion problem into "mild" and "severe" regimes, and the properties of NPIV models differ fundamentally across these cases. Minimax convergence rates in the mild regime are achieved by a choice of sieve dimension that balances bias and sampling uncertainty, much like in standard nonparametric problems such as density estimation and regression. By contrast, minimax rates in the severe regime are obtained by a bias-dominating choice of sieve dimension.

Our procedure for data-driven choice of sieve dimension delivers the minimax sup-norm rate for h_0 and its derivatives across the whole spectrum of models, from nonparametric regression to NPIV models in the severe regime. As we shall explain in Section 3, this adaptivity is achieved through several novel modifications to existing bootstrap-based implementations

[4] See Hall and Horowitz (2005) and Chen and Reiss (2011) for minimax L^2-norm rates, and Chen and Christensen (2013, 2018) for minimax sup-norm rates for NPIV models. When the conditional density of X given W is continuous, each of these minimax rates are slower than the corresponding rates for nonparametric regression (Stone, 1982).

of Lepski's method. Our procedure improves significantly on and supersedes an (unpublished) earlier procedure from Section 3 of the working paper Chen and Christensen (2015a) on sup-norm rate-adaptive estimation of NPIV models.[5] In particular, it uses the bootstrap to avoid selection of several constants, it performs much better in practice, and its minimax sup-norm rate-adaptive guarantees encompass both nonparametric regression and NPIV models in both mild and severe regimes.

Our second procedure for data-driven UCBs builds on the statistics literature on honest, adaptive UCBs for nonparametric density estimation (Giné and Nickl, 2010; Chernozhukov et al., 2014) and Gaussian white noise models (Bull, 2012; Giné and Nickl, 2016). However, none of these works allows for nonparametric models with endogeneity, and our procedures do not follow easily from these existing methods due to the above-mentioned complications that arise in NPIV models. In fact, our data-driven UCBs for h_0 and its partial derivatives apply to nonparametric regression with non-Gaussian, heteroskedastic errors as a special case.[6]

Finally, our work also compliments several recent papers on (non data-driven) estimation and inference for sieve NPIV models with shape constraints; see for example Chetverikov and Wilhelm (2017), Chernozhukov, Newey, and Santos (2015), Blundell et al. (2017), Freyberger and Reeves (2019), Zhu (2020), and Fang and Seo (2021). Each of these works assumes a deterministic sequence of sieve tuning parameters satisfying regularity conditions that depend on unknown model features. An exception is Breunig and Chen (2020) who study data-driven, L^2-norm rate-adaptive testing of NPIV models. Note, however, that in nonparametric models a data-driven choice of tuning parameters for L^2-norm rate-adaptive testing does not, in general, lead to the minimax rate of estimation or testing in sup-norm (see, e.g., Chapter 8.1 of Giné and Nickl (2016)). In theory the adaptive test of Breunig and Chen (2020) could be inverted to construct a L^2-norm confidence ball for h_0 or its derivatives, though doing so in practice may be computationally challenging. It is also difficult to interpret L^2-norm confidence balls in nonparametric models. By contrast, our UCBs for h_0 and its derivatives are easy to compute, plot, and interpret.

The remainder of our paper is structured as follows. Section 2 presents a motivating illustration in the context of Engel curve estimation and an empirical application to international trade. Section 3 introduces our computationally simple data-driven procedures. Section 4 presents the main theoretical results on sup-norm rate adaptive estimation and honest and adaptive UCBs. Additional simulation evidence is presented in Section 5 while Section 6 concludes. Appendix A presents details on basis functions. Appendix B contains technical results and proofs of main

[5]See Horowitz (2014), Liu and Tao (2014), Breunig and Johannes (2016) and Jansson and Pouzo (2020) for other data-driven procedures for choosing tuning parameters for sieve NPIV estimators. None of these papers considered sup-norm rate adaptivity, which is required for constructing honest and adaptive UCBs.

[6]Although a by-product of our main results, this appears to be a new addition to the literature on honest, adaptive UCBs for nonparametric regression and its derivatives.

results in Sections 4.2 and 4.3. Supplemental Appendix C presents proofs of all the technical results in Appendix B and proofs of the theorems in Section 4.4.

2 Motivating Illustration and Empirical Application

2.1 Motivating Illustration: Engel Curve Estimation

We first present an empirically relevant simulation exercise to illustrate why an adaptive choice of sieve dimension J is important for sieve NPIV estimation and UCB construction. Our Monte Carlo design mimics the British Family Expenditure Survey data set used by Blundell et al. (2007) to nonparametrically estimate household Engel curves with endogenous log total expenditure (X), using log gross earnings of the household head as an instrument (W). We simulate data from a kernel estimate of the joint density of (X, W).[7] For each (X, W) draw, we set

$$Y = h_0(X) + u\,, \qquad u = \mathbb{E}[h_0(X)|W] - h_0(X) + v\,, \qquad v \sim N(0, 0.01)\,,$$

with v independent of (X, W). The structural function we use is $h_0(x) = \Phi(4.864x - 1.256)$, similar to the simulation design 1 in Blundell et al. (2007). For each simulated data set, we compute the estimator $\hat{h}_{\tilde{J}}$ using the data-driven choice of sieve dimension $\tilde{J}$ described in Section 3.1 as well as data-driven UCBs for h_0 formed as described in display (8).[8] We then compare our data-driven procedures to "undersmoothed" estimators and UCBs based on a value of J that is larger than the (unknown) optimal choice, in the hope that bias is of smaller order than sampling uncertainty. We construct the undersmoothed value J^u as described in Section 5.2 so that it diverges faster than $\tilde{J}$ by a $\log n$ factor. We compute the undersmoothed estimator $\hat{h}_{J^u}$ and undersmoothed UCBs that are centered at $\hat{h}_{J^u}$ and whose widths are determined using the approach of Chen and Christensen (2018), which uses a less conservative critical value that only accounts for sampling uncertainty conditional on the choice of sieve dimension. Results are presented in Table 1; additional results are presented in Table 2 in Section 5.

Columns 2-5 of Table 1 show that the maximal error for the undersmoothed estimator is several multiples larger than that of the data-driven estimator $\hat{h}_{\tilde{J}}$. In most Monte Carlo replications our data-driven choice was $\tilde{J} = 3$,[9] in which case $J^u = 6$ for the smaller sample sizes $n = 1250$ and 2500, and $J^u = 10$ for the larger sample sizes $n = 5000$ and 10000. Evidently, modest increases in J above $\tilde{J}$ can produce a significant deterioration in the properties of $\hat{h}_J$.

[7] We use data for families without children, as described in Blundell et al. (2007). We transform the raw data by first standardizing the log total expenditure and log gross earnings of the household head by their sample means and standard deviations, then taking a standard normal cumulative density function transformation of the standardized data, so that the transformed data take values in $[0, 1]$.

[8] We use B-spline sieve constructed as described in Appendix A.1 with $r = 3$ and $q = 2$.

[9] Interestingly, our procedure chooses $\tilde{J} = 3$ (nonlinear function) over $\tilde{J} = 2$ (linear function).

NPIV	$\|\hat{h}_{\tilde{J}} - h_0\|_\infty$		$\|\hat{h}_{J^u} - h_0\|_\infty$		Coverage		RMW	
n	mean	med	mean	med	90%	95%	mean	med
1250	0.184	0.149	0.576	0.496	0.980	0.986	1.701	1.584
2500	0.163	0.134	0.564	0.474	0.990	0.993	2.226	2.098
5000	0.152	0.129	0.662	0.565	0.981	0.989	3.961	3.801
10000	0.146	0.137	0.659	0.573	0.946	0.959	5.522	5.330

Table 1: Mean and median maximal estimation errors of $\hat{h}_{\tilde{J}}$ (data-driven) and $\hat{h}_{J^u}$ (undersmoothed) (columns 2-5), coverage of data-driven 90% and 95% UCBs for h_0 (columns 6-7), and mean and median relative maximal width (RMW) of undersmoothed 95% UCBs to data-driven 95% UCBs (columns 8-9) across 1000 Monte Carlo simulations for the Engel curve design, with 1000 bootstrap replications per simulation.

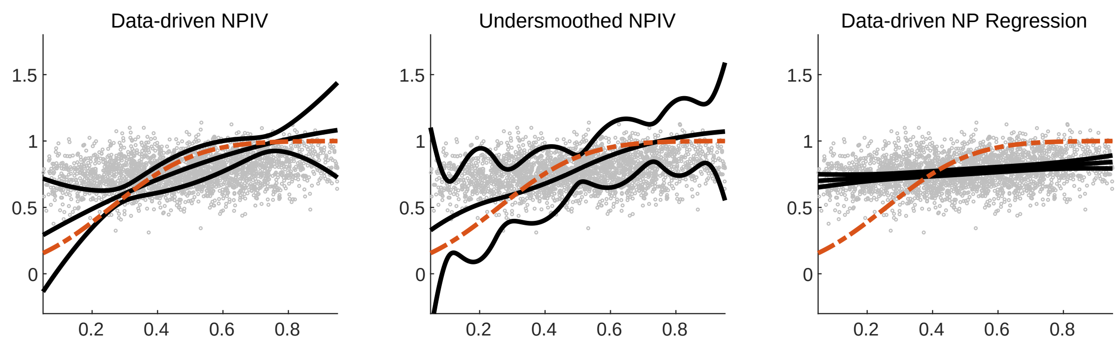

Figure 1: Estimated structural function h_0 and 95% UCBs for h_0 (solid lines) using our data-driven procedure (left panel) and undersmoothing (center panel) for a sample of size 2500 (grey dots). The true function h_0 is also shown (orange dashed lines). The right panel shows the estimated conditional mean function $\mathbb{E}[Y|X = x]$ and 95% UCBs for the conditional mean function.

Turning to columns 8-9 of Table 1, we see that the maximal widths of undersmoothed 95% UCBs are overall larger than the maximal widths of our data-driven 95% UCBs, and the relative width is increasing with sample size. For instance, our data-driven 95% UCBs are roughly 50% the width of undersmoothed bands for $n = 2500$ and 20% of the width of undersmoothed bands for $n = 10000$. Importantly, this reduction in width is not at the cost of coverage: the results in columns 6-7 of Table 1 show that our data-driven UCBs are slightly conservative. Note that some conservativeness is to be expected, as our UCBs deliver uniform coverage guarantees over a class of data-generating processes.

Figure 1 illustrates the improvement in terms of width of our data-driven UCBs relative to undersmoothed UCBs for a sample of size $n = 2500$. In this sample, our data-driven procedure chooses sieve dimension $\tilde{J} = 3$ while the undersmoothed estimator uses $J^u = 6$. For comparison, in Figure 1 we also plot a sieve nonparametric estimate of the conditional mean function $\mathbb{E}[Y|X = x]$ and 95% UCBs for it using our data-driven procedures. While the UCB for the conditional mean function is much narrower than the UCBs for the true structural function h_0, it excludes the true structural function over almost all the support of X. The true conditional mean function falls entirely within these bands for this sample, however.

2.2 Empirical Application: International Trade

Adão et al. (2020) derive two semiparametric gravity equations for the extensive and intensive margins of firm exports in a monopolistic competition model of international trade without parametric restrictions on the distribution of firm heterogeneity. The (nonparametric) functions identified by these equations characterize the elasticity of firm exports to changes in bilateral trade costs. Adão et al. (2020) estimate the elasticities from aggregate bilateral trade data using a sieve NPIV approach, approximating the unknown function with cubic splines and using (functions of) cost shifters as instruments. In this section, we apply our procedure to determine the sieve dimension in a data-driven manner and to construct data-driven UCBs for the function characterizing the intensive margin and its elasticity.

Model and Data. We base our estimation on the gravity equation for the intensive margin of firm exports derived by Adão et al. (2020), which may be expressed as

$$y_{ij} = \log \rho(\log \nu_{ij}) + \zeta_i + \delta_j + u_{ij}\,,$$

where $y_{ij} = \log \bar{x}_{ij} + \tilde{\sigma}\kappa^{\tau} z_{ij}$ where $\bar{x}_{ij}$ is the average sales of firms of country i selling in country j and z_{ij} is a cost shifter,[10] ν_{ij} is the share of such firms, ζ_i and δ_j are origin and destination fixed effects, and u_{ij} is an idiosyncratic error term. Our goal is to estimate the function $\log \rho$, as its derivative $\frac{\partial \log \rho(\log \nu)}{\partial \log \nu}$ characterizes the elasticity of the intensive margin of firm exports to changes in bilateral trade costs. We use the same data as Adão et al. (2020) use for their baseline estimates, which is a sample of 1522 country pairs for the year 2012.

Implementation. We simplify the empirical implementation of Adão et al. (2020) in a number of respects so that our data-driven procedures can be applied in a transparent manner. First, we maintain their assumption that the instrumental variable z_{ij} and origin and destination fixed effects are exogenous, but we further assume that $\mathbb{E}[\log \rho(\log \nu_{ij})|z_{ij}, \zeta_i, \delta_j] = \mathbb{E}[\log \rho(\log \nu_{ij})|z_{ij}]$ (almost surely). Under this assumption, the reduced form for y_{ij} may then be expressed as a nonparametric regression model with the same origin and destination fixed effects:

$$y_{ij} = g(z_{ij}) + \zeta_i + \delta_j + e_{ij}\,, \tag{2}$$

where $g(z_{ij}) = \mathbb{E}[\log \rho(\log \nu_{ij})|z_{ij}]$ and $\mathbb{E}[e_{ij}|z_{ij}, \zeta_i, \delta_j] = 0$. We estimate ζ_i and δ_j from (2) by regressing y_{ij} on origin and destination dummies and the basis functions $b_{K1}, \ldots, b_{KK}$ of z_{ij} at dimension $K(\hat{J}_{\max})$ (these notations are defined in Section 3). We then apply our procedures

[10] Adão et al. (2020) construct y_{ij} based on external estimates of $\tilde{\sigma}$ and κ^{τ}.

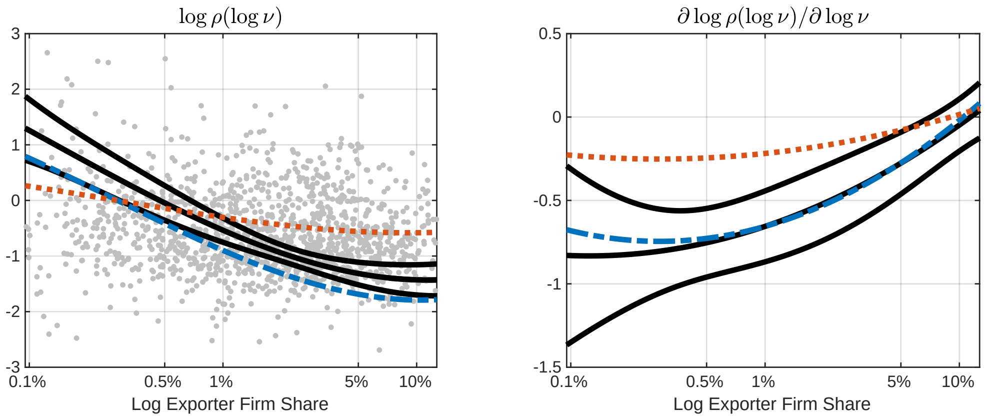

Figure 2: Left panel: estimated structural function $\log \rho$ and 95% UCBs for $\log \rho$ (solid lines) using our data-driven procedure. Right panel: estimated elasticity of ρ and 95% UCBs for the elasticity (solid lines) using our data driven procedure. Nonparametric regression estimates (orange dotted lines) and estimates with $\log \rho$ and the fixed effects jointly estimated in the second stage (blue dot-dashed lines) are also shown.

using $Y_{ij} = y_{ij} - \hat{\zeta}_i - \hat{\delta}_j$ as the outcome variable (Y), $\log \nu_{ij}$ as the endogenous regressor (X), and z_{ij} as the instrumental variable (W). We also estimate just the single equation (2) on its own rather than both semiparametric gravity equations as a system. In addition, we use different basis functions to approximate h_0 and to estimate the reduced form. Specifically, we use a B-spline sieve constructed as described in Appendix A.1 with $r = 4$ and $q = 2$. Finally, rather than restricting $\log \rho(\log \nu)$ to be linear in the tails of $\log \nu$, we instead compute confidence bands over the 5th to 95th percentiles of $\log \nu$.

Results. Figure 2 plots the estimated function $\log \rho$ and the elasticity of ρ, together with 95% UCBs constructed as in displays (8) and (13), respectively. The estimated function $\log \rho$ appears monotone and the UCBs for $\log \rho$ and the elasticity of ρ are both narrow and informative. The estimated elasticity is similar to (but more upwards-sloping than) the estimate reported in Adão et al. (2020), though our empirical implementation is different and so our results are not directly comparable. Interestingly, a flat line does not fit between our UCBs for the elasticity because the upper limit of the lower band exceeds the lower limit of the upper band. As such, our UCBs for the elasticity of ρ can be interpreted as providing evidence against the Pareto specification for unobserved firm productivity used e.g. by Chaney (2008), under which the elasticity is constant. In addition, the estimated elasticity appears to be *increasing* in the exporter firm share whereas Figure 1 of Adão et al. (2020) shows that several conventional parameterizations (Eaton et al., 2011; Head et al., 2014; Melitz and Redding, 2015) of the distribution of unobserved firm productivity all imply a *decreasing* elasticity. Indeed, decreasing elasticities would necessarily fall outside our UCBs for the elasticity of ρ.

Recall that our estimation approach eliminates the fixed effects from the reduced form and estimates $\log \rho$ instrumenting with (transformations of) z_{ij} whereas Adão et al. (2020) estimate $\log \rho$ and the fixed effects jointly, instrumenting with (transformations of) z_{ij} and the origin and destination dummies. As a robustness check of our approach, we estimate $\log \rho$ and the fixed effects jointly, using the data-driven choice $\tilde{J}$ computed using our procedure, instrumenting with $b_{K(\tilde{J})1}(z_{ij}), \ldots, b_{K(\tilde{J})K(\tilde{J})}(z_{ij})$ (these notations are defined in Section 3) and the origin and destination dummies, and using y_{ij} as the dependent variable. Estimates using this approach are also shown in Figure 2 and are similar to those obtained for our original implementation. In particular, the estimated elasticity—which is the focus of the analysis of Adão et al. (2020)—are almost identical to our baseline elasticity estimates for most of the support of $\log \nu$.

Finally, to clarify the importance of using an IV approach in this application, we also plot nonparametric regression estimates of $\log \rho$ and the elasticity of ρ using y_{ij} as the dependent variable and regressing on the same basis functions of $\log \nu_{ij}$ used for our NPIV estimator (with dimension $\tilde{J}$) and origin and destination dummies. The regression estimates are clearly very different from those we estimate using NPIV, and lie outside the UCBs for large regions of the support of $\log \nu$.

We note in closing that our procedures can equally be applied to other IV-based nonparametric analyses in international trade; see, e.g., Adao, Costinot, and Donaldson (2017).

3 Data-driven Procedures

Our data-driven estimation and UCB procedures are based on sieve NPIV estimators, which are two-stage least-squares (TSLS) estimators applied to basis functions of X and W. Given a random sample $(X_i, Y_i, W_i)_{i=1}^n$ from model (1), the sieve NPIV estimator of h_0 is

$$\hat{h}_{J,K}(x) = \psi_x^{J\prime} \hat{c}_{J,K}\,, \quad \hat{c}_{J,K} = \left[(\Psi_J' B_K)(B_K' B_K)^{-1} B_K' \Psi_J\right]^{-1} (\Psi_J' B_K)(B_K' B_K)^{-1} B_K' Y\,, \tag{3}$$

where $'$ denotes transpose, $Y = (Y_1, \ldots, Y_n)'$, and

$$\psi_x^J = (\psi_{J1}(x), \ldots, \psi_{JJ}(x))'\,, \qquad \Psi_J = (\psi_{X_1}^J, \ldots, \psi_{X_n}^J)'\,,$$
$$b_w^K = (b_{K1}(w), \ldots, b_{KK}(w))'\,, \qquad B_K = (b_{W_1}^K, \ldots, b_{W_n}^K)'\,,$$

where ψ_x^J and b_w^K are vectors of basis functions of dimensions J and K, respectively. The TSLS interpretation regards ψ_x^J as the J dimensional vector of endogenous variables and b_w^K as the K dimensional vector of (unconditional) instruments, and hence $K \geq J$ as the default relation (Blundell et al., 2007; Chen and Christensen, 2018). As the choice of K will be determined by J in our data-driven procedure, in what follows we write $K = K(J)$, $\hat{h}_J = \hat{h}_{J,K(J)}$, and $\hat{c}_J = \hat{c}_{J,K(J)}$. The estimator (3) nests a series least-squares regression estimator as a special case when $W = X$, $J = K$, and $b_w^k = \psi_x^J$.

Sieve NPIV estimators using B-spline bases and Cohen–Daubechies–Vial (CDV) wavelet bases for h_0 are known to attain the minimax sup-norm rates for h_0 and its derivatives under a theoretically optimal choice of tuning parameters (Chen and Christensen, 2018). As we are concerned with sup-norm rate-adaptive estimation and UCBs, we therefore restrict our attention to these bases in what follows. We refer the reader to Appendix A for further details on these bases.

3.1 Data-driven Choice of Sieve Dimension

Appendix A describes the set $\mathcal{T}$ denoting all possible integer values of sieve dimension J for the bases we use and the relation $K(J)$ linking the sieve dimensions for the instrumental variables and regressors. Our data-driven choice of sieve dimension J is computed in three simple steps.

Step 1. Compute a feasible index set for J, namely

$$\hat{\mathcal{J}} = \{J \in \mathcal{T} : 0.1(\log \hat{J}_{\max})^2 \leq J \leq \hat{J}_{\max}\}, \quad \text{with}$$

$$\hat{J}_{\max} = \min \Big\{ J \in \mathcal{T} : J\sqrt{\log J}\big[(0.1 \log n)^4 \vee \hat{s}_J^{-1}\big] \leq 10\sqrt{n} < J^+\sqrt{\log J^+}\big[(0.1 \log n)^4 \vee \hat{s}_{J^+}^{-1}\big] \Big\}, \quad (4)$$

where $a \vee b = \max\{a, b\}$, $J^+ = \min\{j \in \mathcal{T} : j > J\}$, and $\hat{s}_J$ is the smallest singular value of $\widehat{G}_{b,J}^{-1/2} \widehat{S}_J \widehat{G}_{\psi,J}^{-1/2}$ where $\widehat{G}_{b,J} = B'_{K(J)} B_{K(J)}/n$, $\widehat{G}_{\psi,J} = \Psi'_J \Psi_J/n$, and $\widehat{S}_J = B'_{K(J)} \Psi_J/n$.

The $(0.1 \log n)^4$ terms in (4) are included to accommodate nonparametric regression and can be omitted for NPIV models. Note that $\hat{s}_J = 1$ and the index set $\hat{\mathcal{J}}$ is deterministic for nonparametric regression (where $W = X$). However, for NPIV models, $\hat{s}_J$ decreases to zero as J increases, and the index set $\hat{\mathcal{J}}$ is random as $\hat{J}_{\max}$ depends on $\hat{s}_J^{-1}$ (the estimated difficulty of inversion).

Step 2. Let $\hat{\alpha} = \min\{0.5, \hat{J}_{\max}^{-1}\}$. Compute a bootstrap-based critical value $\theta^*(\hat{\alpha})$, namely the $(1 - \hat{\alpha})$ quantile of the sup statistic

$$\sup_{(x,J,J_2) \in \hat{\mathcal{S}}} |\mathbb{Z}_n^*(x, J, J_2)|,$$

where $\hat{\mathcal{S}} = \{(x, J, J_2) \in \mathcal{X} \times \hat{\mathcal{J}} \times \hat{\mathcal{J}} : J_2 > J\}$ with $\mathcal{X}$ denoting the support of X, and

$$\mathbb{Z}_n^*(x, J, J_2) = \frac{1}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\hat{L}_{J,x} b_{W_i}^{K(J)} \hat{u}_{i,J} - \hat{L}_{J_2,x} b_{W_i}^{K(J_2)} \hat{u}_{i,J_2} \right) \varpi_i \right), \tag{5}$$

where $(\varpi_i)_{i=1}^n$ are IID $N(0,1)$ draws that are independent of the data, and for $J, J_2 \in \hat{\mathcal{J}}$,

$$\|\hat{\sigma}_{x,J,J_2}\|_{sd}^2 = \|\hat{\sigma}_{x,J}\|_{sd}^2 + \|\hat{\sigma}_{x,J_2}\|_{sd}^2 - 2\hat{\sigma}_{x,J,J_2}$$

with $\|\hat{\sigma}_{x,J}\|_{sd}^2 = \hat{\sigma}_{x,J,J}$ and $\hat{\sigma}_{x,J,J_2} = \hat{L}_{J,x} \widehat{\Omega}_{J,J_2} (\hat{L}_{J_2,x})'$ where

$$\hat{L}_{J,x} = (\psi_x^J)' [\widehat{S}_J' \widehat{G}_{b,J}^{-1} \widehat{S}_J]^{-1} \widehat{S}_J' \widehat{G}_{b,J}^{-1}, \quad \widehat{\Omega}_{J,J_2} = \frac{1}{n} \sum_{i=1}^{n} \hat{u}_{i,J} \hat{u}_{i,J_2} b_{W_i}^{K(J)} (b_{W_i}^{K(J_2)})',$$

and $\hat{u}_{i,J} = Y_i - \hat{h}_J(X_i)$. Note that the variance term $\|\hat{\sigma}_{x,J,J_2}\|_{sd}^2$ is no harder to compute than standard errors for the difference of two TSLS estimators.

The critical value $\theta^*(\hat{\alpha})$ is calculated by computing $\sup_{(x,J,J_2) \in \hat{\mathcal{S}}} |\mathbb{Z}_n^*(x, J, J_2)|$ across a large number of independent draws of $(\varpi_i)_{i=1}^n$, then taking the $(1 - \hat{\alpha})$ quantile across the draws. This bootstrap is simple to implement, as it avoids resampling the data and recomputing the estimators across each iteration. In practice, the sup over x can be replaced by the maximum over a fine grid.

Step 3. Compute our data-driven optimal choice of J as

$$\tilde{J} = \min\{\hat{J}, \hat{J}_n\}, \tag{6}$$

where $\hat{J}_n = \max\{j \in \mathcal{T} : j < \hat{J}_{\max}\}$ is the largest value of J that is smaller than $\hat{J}_{\max}$, and

$$\hat{J} = \min\left\{ J \in \hat{\mathcal{J}} : \sup_{(x,J_2) \in \mathcal{X} \times \hat{\mathcal{J}} : J_2 > J} \frac{\sqrt{n} |\hat{h}_J(x) - \hat{h}_{J_2}(x)|}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} \leq 1.1 \theta^*(\hat{\alpha}) \right\}.$$

Theorem 4.1 and Corollary 4.1 below establish that the data-driven choice $\tilde{J}$ leads to estimators of h_0 and its derivatives that attain minimax sup-norm rates, and the performance is guaranteed across nonparametric regression models and NPIV models in both the mild and severe regimes. Section 5 presents additional simulation results which illustrate the performance of our data-driven choice $\tilde{J}$ in a highly nonlinear nonparametric regression design and an additional nonparametric IV design.

Remark 3.1 *Our procedure differs from bootstrap-based implementations of Lepski's method for density estimation and nonparametric regression without endogeneity in a number of ways. For instance, it uses a search over a data-dependent index set $\hat{\mathcal{J}}$ (see Step 1). This is an* *important modification for NPIV models, as the quality of the estimator $\hat{h}_J$ deteriorates more rapidly as J increases when the inversion problem is more difficult. As the degree of difficulty depends on the conditional density of X given W and is therefore unknown, the approximate degree of difficulty is inferred from data through $\hat{s}_J^{-1}$. The dependence on $\hat{s}_J^{-1}$ ensures $\hat{\mathcal{J}}$ is smaller for more severe problems and larger for milder problems. Moreover, our final choice $\tilde{J}$ (in Step 3) truncates the bootstrap-based Lepski estimator $\hat{J}$ at $\hat{J}_n$. While the truncation is typically not binding for nonparametric regression and NPIV models in the mild regime, this modification is important to achieve adaptivity for NPIV models in the severe regime.*

3.2 Data-driven Uniform Confidence Bands

Our UCBs for h_0 are centered at the data-driven estimator $\hat{h}_{\tilde{J}}$ and use a bootstrap-based critical value. Recall $\hat{L}_{J,x}$, $\hat{u}_{i,J}$, $\|\hat{\sigma}_{x,J}\|_{sd}$, and $(\varpi_i)_{i=1}^n$ from Step 2 above and let

$$\mathbb{Z}_n^*(x,J) = \frac{1}{\|\hat{\sigma}_{x,J}\|_{sd}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{L}_{J,x} b_{W_i}^{K(J)} \hat{u}_{i,J} \varpi_i \right). \tag{7}$$

The critical value $z_{1-\alpha}^*$ ($\alpha = 0.05$ for a 95% UCB) is calculated by repeatedly computing $\sup_{(x,J)\in\mathcal{X}\times\hat{\mathcal{J}}} |\mathbb{Z}_n^*(x,J)|$ across a large number of independent draws of $(\varpi_i)_{i=1}^n$, then taking the $(1-\alpha)$ quantile. This bootstrap is as simple to compute as that in Step 2 above.

Given the critical value $z_{1-\alpha}^*$, our first UCB is

$$C_n(x) = \left[\hat{h}_{\tilde{J}}(x) - (z_{1-\alpha}^* + A_n \theta^*(\hat{\alpha})) \frac{\|\hat{\sigma}_{x,\tilde{J}}\|_{sd}}{\sqrt{n}},\ \hat{h}_{\tilde{J}}(x) + (z_{1-\alpha}^* + A_n \theta^*(\hat{\alpha})) \frac{\|\hat{\sigma}_{x,\tilde{J}}\|_{sd}}{\sqrt{n}} \right] \tag{8}$$

with $A_n = 0.25 \log\log n$. Theorem 4.2 below establishes the coverage properties of this UCB and its contraction rate in the mild regime. We recommend to use this UCB whenever $\tilde{J} = \hat{J}$ from Step 3. If $\tilde{J} = \hat{J}_n$, we recommend using the UCB

$$C_n(x) = \left[\hat{h}_{\tilde{J}}(x) - z_{1-\alpha}^* \frac{\|\hat{\sigma}_{x,\tilde{J}}\|_{sd}}{\sqrt{n}} - \hat{A}_n^*,\ \hat{h}_{\tilde{J}}(x) + z_{1-\alpha}^* \frac{\|\hat{\sigma}_{x,\tilde{J}}\|_{sd}}{\sqrt{n}} + \hat{A}_n^* \right], \tag{9}$$

where

$$\hat{A}_n^* = A_n \max\left\{ \theta^*(\hat{\alpha}) \frac{\|\hat{\sigma}_{x,\tilde{J}}\|_{sd}}{\sqrt{n}}\ ,\ \tilde{J}^{-\underline{p}/d} \right\}$$

and $\underline{p}$ is the minimal degree of smoothness assumed for h_0 (e.g., $\underline{p} = 1$ if h_0 is assumed to be at least Lipschitz continuous). Theorem 4.3 below establishes the coverage properties of this UCB and its contraction rate in the severe regime.

Lower and upper versions of (8) are given by

$$C_{L,n}(x) = \left[\hat{h}_{\tilde{J}}(x) - (z^*_{L,1-\alpha} + A_n \theta^*(\hat{\alpha})) \frac{\|\hat{\sigma}_{x,\tilde{J}}\|_{sd}}{\sqrt{n}} \, , \, \infty \right), \quad \text{and} \tag{10}$$

$$C_{U,n}(x) = \left(-\infty, \, \hat{h}_{\tilde{J}}(x) + (z^*_{U,1-\alpha} + A_n \theta^*(\hat{\alpha})) \frac{\|\hat{\sigma}_{x,\tilde{J}}\|_{sd}}{\sqrt{n}} \right], \tag{11}$$

where $z^*_{L,1-\alpha}$ is the $(1-\alpha)$ quantile of $\sup_{(x,J)\in\mathcal{X}\times\hat{\mathcal{J}}} \mathbb{Z}^*_n(x,J)$ and $-z^*_{U,1-\alpha}$ is the α quantile of $\inf_{(x,J)\in\mathcal{X}\times\hat{\mathcal{J}}} \mathbb{Z}^*_n(x,J)$. One-sided versions of (9) are constructed similarly, again using $\hat{A}^*_n$ in place of $A_n\theta^*(\hat{\alpha})\|\hat{\sigma}_{x,\tilde{J}}\|_{sd}/\sqrt{n}$.

Section 5 presents additional simulation results which illustrate the performance of our data-driven UCBs in a highly nonlinear nonparametric regression design and an additional nonparametric IV design, and their efficiency relative to undersmoothing. Finally, we note that while we used Gaussian draws $(\varpi_i)_{i=1}^n$ in the bootstrap, it is also possible to use IID draws from any distribution that has mean zero, unit variance, and finite third moment (e.g. Rademacher draws or draws from the two-point distribution of Mammen (1993)).

4 Theory

In this section, we first outline the main regularity conditions before presenting the theoretical results about adaptive estimation and uniform confidence bands.

4.1 Assumptions

We first state and then discuss the assumptions that we impose on the model and sieve space. We require these to hold for some constants $a_f, \underline{c}, \overline{C}, C_T, C_Q, \underline{\sigma}, \overline{\sigma} > 0$ and $\gamma \in (0,1)$.

For the first assumption, let $L^2(X)$ and $L^2(W)$ denote the classes of measurable functions of X and W, respectively, with finite second moments, and let $T : L^2(X) \to L^2(W)$ denote the conditional expectation operator $Th(w) = \mathbb{E}[h(X)|W=w]$. In nonparametric regression models where $W = X$, the conditional expectation T reduces to the identity operator.

Assumption 1 *(i) X has support $\mathcal{X} = [0,1]^d$ and its distribution has Lebesgue density f_X which satisfies $a_f^{-1} < f_X(x) < a_f$ on $\mathcal{X}$;*
(ii) W has support $\mathcal{W} = [0,1]^{d_w}$ and its distribution has Lebesgue density f_W which satisfies $a_f^{-1} < f_W(w) < a_f$ on $\mathcal{W}$;
(iii) $T : L^2(X) \to L^2(W)$ is an injective operator.

Assumption 2 *(i) $\mathbb{P}\big(\mathbb{E}[u^4|W] \le \overline{\sigma}^2\big) = 1$;*
(ii) $\mathbb{P}\big(\mathbb{E}[u^2|W] \ge \underline{\sigma}^2\big) = 1$.

For the next assumption, let Ψ_J and B_K be the closed linear subspaces of $L^2(X)$ and $L^2(W)$ spanned by $\psi_{J1}, \ldots, \psi_{JJ}$ and $b_{K1}, \ldots, b_{KK}$, respectively. The *sieve measure of ill-posedness* is

$$\tau_J = \sup_{h \in \Psi_J : \|h\|_{L^2(X)} \neq 0} \frac{\|h\|_{L^2(X)}}{\|Th\|_{L^2(W)}},$$

where $\|h\|^2_{L^2(X)} = \mathbb{E}[h(X)^2]$ with $\|\cdot\|_{L^2(W)}$ is defined similarly. The measure τ_J quantifies the degree of difficulty of inverting $\mathbb{E}[h_0(X)|W]$ to recover h_0. The model (1) is said to be *mildly ill-posed* (or in the *mild* regime) if $\tau_J \asymp J^{\varsigma/d}$ for some $\varsigma \geq 0$ and *severely ill-posed* (or in the *severe* regime) if $\tau_J \asymp e^{CJ^{\varsigma/d}}$ for some $C, \varsigma > 0$, where $d = \dim(X)$. Note that in nonparametric regression models $\tau_J = 1$ for all J, so the mildly ill-posed case includes nonparametric regression as a special case with $\varsigma = 0$.

In addition, let $\Pi_J : L^2(X) \to \Psi_J$ and $\Pi_K : L^2(W) \to B_K$ denote the least-squares projection operators onto Ψ_J and B_K, respectively, i.e.:

$$\Pi_J f = \arg\min_{g \in \Psi_J} \|f - g\|_{L^2(X)}, \qquad \Pi_K f = \arg\min_{g \in B_K} \|f - g\|_{L^2(W)}.$$

We also define the TSLS projection operator $Q_J : L^2(X) \to \Psi_J$ as

$$Q_J h_0(\cdot) = \arg\min_{h \in \Psi_J} \|\Pi_K T(h_0 - h)\|_{L^2(W)} = (\psi^J_{(\cdot)})'[S'_J G^{-1}_{b,J} S_J]^{-1} S'_J G^{-1}_{b,J} \mathbb{E}\big[b^{K(J)}_W h_0(X)\big],$$

where $G_{b,J} = \mathbb{E}[b^{K(J)}_W (b^{K(J)}_W)']$ and $S_J = \mathbb{E}[b^{K(J)}_W (\psi^J_X)']$.

Assumption 3 *(i)* $\sup_{h \in \Psi_J, \|h\|_{L^2(X)}=1} \tau_J \|\Pi_{K(J)} Th - Th\|_{L^2(W)} \leq v_J$ *where* $v_J < 1$ *for all* $J \in \mathcal{T}$ *and* $v_J \to 0$ *as* $J \to \infty$*;*
(ii) $\tau_J \|T(h_0 - \Pi_J h_0)\|_{L^2(W)} \leq C_T \|h_0 - \Pi_J h_0\|_{L^2(X)}$ *for all* $J \in \mathcal{T}$*;*
(iii) $\|Q_J(h_0 - \Pi_J h_0)\|_\infty \leq C_Q \|h_0 - \Pi_J h_0\|_\infty$ *for all* $J \in \mathcal{T}$.

By analogy with the definition of $\|\hat{\sigma}_{x,J}\|^2_{sd}$, the "population" sieve variance of $\hat{h}_J(x)$ is $\|\sigma_{x,J}\|^2_{sd} = L_{J,x} \Omega_J L'_{J,x}$ where $L_{J,x} = (\psi^J_x)'[S'_J G^{-1}_{b,J} S_J]^{-1} S'_J G^{-1}_{b,J}$ and $\Omega_J = \mathbb{E}[u^2 b^{K(J)}_W (b^{K(J)}_W)']$. We also let $\|\sigma_{x,J}\|^2 = [\psi^J_x]'[S'_J G^{-1}_{b,J} S_J]^{-1}[\psi^J_x]$ which, in view of Assumption 2(i)(ii), satisfies $\|\sigma_{x,J}\| \asymp \|\sigma_{x,J}\|_{sd}$ uniformly in x.

Assumption 4 *(i)* $\underline{c}\tau^2_J J \leq \inf_{x \in \mathcal{X}} \|\sigma_{x,J}\|^2 \leq \sup_{x \in \mathcal{X}} \|\sigma_{x,J}\|^2 \leq \overline{C}\tau^2_J J$ *for all* $J \in \mathcal{T}$*;*
(ii) $\limsup_{J \to \infty} \sup_{x \in \mathcal{X}, J_2 \in \mathcal{T} : J_2 > J} (\|\sigma_{x,J}\|^2_{sd} / \|\sigma_{x,J_2}\|^2_{sd}) < \gamma$.

Assumptions 1, 2, and 3 are the same as (or slightly modifications of) Assumptions 1(i)–(iii), 2(i)(iv), and 4 of Chen and Christensen (2018). To briefly summarize these conditions, Assumption 1(i)(ii) are standard while Assumption 1(iii) is generically satisfied in models with endogeneity (see Andrews (2017)) and is trivially satisfied for nonparametric regression because $W = X$ and T reduces to the identity operator. Assumption 3 is trivially satisfied for nonparametric regression models with $C_T, C_Q = 1$. Assumption 3(i) is imposed to ensure that the smallest singular value, denoted s_J, of $G_{b,J}^{-1/2} S_J G_{\psi,J}^{-1/2}$ is a suitable analog of τ_J^{-1}, in the sense that $s_J^{-1} \asymp \tau_J$. Our results apply without this assumption, provided τ_J is replaced by s_J^{-1} everywhere. We do however believe this is a mild condition on the approximation properties of the basis used for the instrument space, as it is trivially satisfied with $\|(\Pi_K T - T)h\|_{L^2(W)} = 0$ for all $h \in \Psi_J$ when the basis functions for B_K and Ψ_J form either a Riesz basis or eigenfunction basis for the conditional expectation operator. Assumption 3(ii) is the usual L^2 "stability condition" imposed in the NPIV literature. Assumption 3 (iii) is a L^∞ "stability condition" to control for the sup-norm bias. Assumption 4(i) is similar to a condition from Corollary 4.1 of Chen and Christensen (2018). Assumption 4(ii) says essentially that the sieve variance $\|\sigma_{x,J}\|_{sd}^2$ is increasing in $J \in \mathcal{T}$, uniformly in x. We view Assumption 4(ii) as a mild condition given that J is increasing exponentially over $\mathcal{T}$. Indeed, by Assumption 2 and 4(i) and the fact that $J = 2^{Ld}$ for some $L \in \mathbb{N}$ for CDV wavelet bases (B-splines are defined similarly, see Appendix A), for any $J, J_2 \in \mathcal{T}$ with $J_2 > J$ we have

$$\sup_{x \in \mathcal{X}} \frac{\|\sigma_{x,J}\|_{sd}^2}{\|\sigma_{x,J_2}\|_{sd}^2} \asymp \frac{\tau_J \sqrt{J}}{\tau_{J_2} \sqrt{J_2}} \leq \frac{\tau_{2^{Ld}}}{\tau_{2^{(L+1)d}}} 2^{-d/2} \leq 2^{-d/2} < 1\,.$$

4.2 Main Results: Adaptive Estimation

We begin by defining the parameter space over which our procedure is adaptive. Let $B_{\infty,\infty}^p(M)$ denote the p-Hölder ball of radius M (see Appendix A.3 for a formal definition). For given constants $C_T, C_Q, M > 0$ and $\overline{p} > \underline{p} > \frac{d}{2}$, let $\mathcal{H}^p = \mathcal{H}^p(M, C_T, C_Q)$ denote the subset of $B_{\infty,\infty}^p(M)$ that satisfies Assumption 3(ii)(iii) for any distribution of (X, W, u) satisfying Assumptions 1-4, and let $\mathcal{H} = \bigcup_{p \in [\underline{p}, \overline{p}]} \mathcal{H}^p$. For each $h_0 \in \mathcal{H}$, we let $\mathbb{P}_{h_0}$ denote the distribution of $(X_i, Y_i, W_i)_{i=1}^\infty$ where each observation is generated by IID draws of (X, W, u) from a distribution of (X, W, u) satisfying Assumptions 1-4 and setting $Y = h_0(X) + u$.

Our first main result on adaptive estimation shows that $\hat{h}_{\tilde{J}}$ converges at the optimal sup-norm rate in both the mildly and severely ill-posed cases:

Theorem 4.1 *Let Assumptions 1-4 hold.*

(i) Suppose the model is mildly ill-posed. Then: there is a universal constant $C_{4.1}$ for which

$$\sup_{p \in [\underline{p}, \overline{p}]} \sup_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}\left(\|\hat{h}_{\tilde{J}} - h_0\|_\infty > C_{4.1} \left(\frac{\log n}{n} \right)^{\frac{p}{2(p+\varsigma)+d}} \right) \to 0.$$

(ii) Suppose the model is severely ill-posed. Then: there is a universal constant $C_{4.1}$ for which

$$\sup_{p \in [\underline{p}, \overline{p}]} \sup_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}\left(\|\hat{h}_{\tilde{J}} - h_0\|_\infty > C_{4.1}(\log n)^{-p/\varsigma}\right) \to 0.$$

Remark 4.1 *The convergence rates in cases (i) and (ii) are the minimax convergence rates for estimating h_0 under sup-norm loss; see Chen and Christensen (2018).*

In fact, adaptivity carries over to estimation of derivatives of h_0 without having to modify the choice of sieve dimension $\tilde{J}$. Given a multi-index $a \in (\mathbb{N} \cup \{0\})^d$ with order $|a| = \sum_{i=1}^d a_i$ and any $f : \mathcal{X} \to \mathbb{R}$, let

$$\partial^a f(x) = \frac{\partial^{|a|} f(x)}{\partial^{a_1} x_1 \dots \partial^{a_d} x_d}.$$

A natural estimator of $\partial^a h_0$ is $\partial^a \hat{h}_{\tilde{J}}$ (i.e., just differentiate $\hat{h}_{\tilde{J}}$). Our second main result on adaptive estimation shows that $\partial^a \hat{h}_{\tilde{J}}$ converges at the optimal sup-norm rate in both the mildly and severely ill-posed cases:

Corollary 4.1 *Let Assumptions 1-4 hold and let $a \in (\mathbb{N} \cup \{0\})^d$ with $0 \leq |a| < \underline{p}$.*

(i) Suppose the model is mildly ill-posed. Then: there is a universal constant $C'_{4.1}$ for which

$$\sup_{p \in [\underline{p}, \overline{p}]} \sup_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}\left(\|\partial^a \hat{h}_{\tilde{J}} - \partial^a h_0\|_\infty > C'_{4.1}\left(\frac{\log n}{n}\right)^{\frac{p-|a|}{2(p+\varsigma)+d}}\right) \to 0.$$

(ii) Suppose the model is severely ill-posed. Then: there is a universal constant $C'_{4.1}$ for which

$$\sup_{p \in [\underline{p}, \overline{p}]} \sup_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}\left(\|\partial^a \hat{h}_{\tilde{J}} - \partial^a h_0\|_\infty > C'_{4.1}(\log n)^{-(p-|a|)/\varsigma}\right) \to 0.$$

Remark 4.2 *The convergence rates in cases (i) and (ii) are the minimax convergence rates for estimating $\partial^a h_0$ under sup-norm loss; see Chen and Christensen (2018).*

To the best of our knowledge, Theorem 4.1 and Corollary 4.1 represent the first results on adaptive estimation in sup-norm for NPIV models and, more generally, ill-posed inverse problems with unknown operator.

4.3 Main Results: Adaptive Uniform Confidence Bands

It has been known since Low (1997) that it is impossible to construct confidence bands that are simultaneously honest and adaptive over standard nonparametric classes (e.g. Hölder balls) of varying smoothness. As is standard following Picard and Tribouley (2000), Giné and

Nickl (2010), Bull (2012), Chernozhukov et al. (2014), and many others, we content ourselves with constructing adaptive confidence bands that deliver uniform coverage guarantees over a "generic" subclass $\mathcal{G}$ of $\mathcal{H}$.

To describe the class $\mathcal{G}$, first note by the discussion in Appendix A.3 that there exists a constant $\overline{B} < \infty$ for which $\sup_{h \in \mathcal{H}^p} \|h - \Pi_J h\|_\infty \leq \overline{B} J^{-\frac{p}{d}}$ holds for all $J \in \mathcal{T}$ and all $p \in [\underline{p}, \overline{p}]$. For any small fixed $\underline{B} \in (0, \overline{B})$ and any $\underline{J} \in \mathcal{T}$, we therefore define

$$\mathcal{G}^p = \left\{ h \in \mathcal{H}^p : \underline{B} J^{-\frac{p}{d}} \leq \|h - \Pi_J h\|_\infty \text{ for all } J \in \mathcal{T} \text{ with } J \geq \underline{J} \right\}, \quad \mathcal{G} = \bigcup_{p \in [\underline{p}, \overline{p}]} \mathcal{G}^p \,.$$

Note that neither $\underline{J}$ nor $\underline{B}$ need to be known to implement our procedure, which is valid for any $\underline{B}$ and $\underline{J}$. Giné and Nickl (2010) present several results establishing the genericity of the class $\mathcal{G}$ in unions of Hölder balls (see also Chapter 8.3 of Giné and Nickl (2016)), which says, in our notation, that for each p the set $\mathcal{H}^p \setminus (\cup_{\underline{B}>0, \underline{J} \in \mathcal{T}} \mathcal{G}^p)$ is nowhere dense in $\mathcal{H}^p$ under the norm topology of $\mathcal{H}^p$. Thus, the set of functions in $\mathcal{H}^p$ that do not belong to $\mathcal{G}^p$ for some $\underline{B}$ and $\underline{J}$ is topologically meagre. We refer the reader to Chapter 8.3 of Giné and Nickl (2016) for further details on importance of the class $\mathcal{G}$ for the existence of adaptive uniform confidence bands in Gaussian white noise models.

We say that a confidence band $\{C_n(x) : x \in \mathcal{X}\}$ is *honest* (asymptotically) over $\mathcal{G}$ with level α if

$$\liminf_{n \to \infty} \inf_{h_0 \in \mathcal{G}} \mathbb{P}_{h_0} \big(h_0(x) \in C_n(x) \quad \forall \; x \in \mathcal{X} \big) \geq 1 - \alpha \,, \tag{12}$$

and *adaptive* if for every $\epsilon > 0$ there exists a constant D for which

$$\liminf_{n \to \infty} \inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{G}^p} \mathbb{P}_{h_0} \left(\sup_{x \in \mathcal{X}} |C_n(x)| \leq D r_n(p) \right) \geq 1 - \epsilon \,,$$

where $| \cdot |$ is Lebesgue measure on $\mathbb{R}$ and $r_n(p)$ is the minimax sup-norm rate of estimation over $\mathcal{H}^p$. Let $C_n(x, A)$ denote the UCB from (8) replacing A_n with a fixed positive constant A. Our first main result is that $C_n(x, A)$ is honest and adaptive in the mildly ill-posed case:

Theorem 4.2 *Let Assumptions 1-4 hold and suppose the model is mildly ill-posed. Then: there is a constant $A^* > 0$ (independent of α) such that for all $A \geq A^*$, we have*

(i) $\displaystyle \liminf_{n \to \infty} \inf_{h_0 \in \mathcal{G}} \mathbb{P}_{h_0} \big(h_0(x) \in C_n(x, A) \quad \forall \; x \in \mathcal{X} \big) \geq 1 - \alpha \,;$

(ii) $\displaystyle \inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{G}^p} \mathbb{P}_{h_0} \left(\sup_{x \in \mathcal{X}} |C_n(x, A)| \leq C_{4.2} (1 + A) \left(\frac{\log n}{n} \right)^{\frac{p}{2(p+\varsigma)+d}} \right) \to 1 \,,$

where $C_{4.2} > 0$ is a universal constant.

It follows immediately from Theorem 4.2(i) that the data-driven UCBs $C_n(x,A)$ have uniform coverage over $\mathcal{G}$ for any sufficiently large A. The uniform coverage guarantees extend to one-sided UCBs also. In practice, the constant A can actually be quite small; see the additional simulation evidence in Section 5. Our recommended choice $A_n = 0.25 \log\log n$ from Section 3 ensures that the bands in (8) are asymptotically valid over $\mathcal{G}$ and contract at rate that is within a $\log\log n$ factor of the minimax sup-norm rate for mildly ill-posed NPIV models.

Remark 4.3 *As the mildly ill-posed case nests nonparametric regression as a special case with $\varsigma = 0$, Theorem 4.2 shows that our UCBs are honest and adaptive for general nonparametric regression models with non-Gaussian, heteroskedastic errors.*

Our UCBs are centered at the data-driven estimator $\hat{h}_{\tilde{J}}$ and have their width determined by a critical value that accounts for both sampling uncertainty and bias. In this respect, our approach is similar to the work Schennach (2020) for UCBs for a nonparametric regression function based on estimates of the bias of kernel estimators. Although her UCBs are valid pointwise, rather than uniformly, in h_0 (cf. display (12)), it is plausible that they could be made honest under additional conditions.

Let $C_n(x,A)$ denote the UCB from (9) replacing A_n with a fixed positive constant A. Our next main result is that $C_n(x,A)$ is honest in the severely ill-posed case, and contracts at the optimal sup-norm rate when the degree of smoothness $p = \underline{p}$:

Theorem 4.3 *Let Assumptions 1-4 hold and suppose the model is severely ill-posed. Then: there is a constant $A^* > 0$ (independent of α) such that for all $A \geq A^*$, we have*

(i) $$\liminf_{n\to\infty} \inf_{h_0 \in \mathcal{G}} \mathbb{P}_{h_0}\big(h_0(x) \in C_n(x,A) \quad \forall\ x \in \mathcal{X}\big) \geq 1-\alpha\,;$$

(ii) $$\inf_{p\in[\underline{p},\overline{p}]} \inf_{h_0 \in \mathcal{G}^p} \mathbb{P}_{h_0}\left(\sup_{x\in\mathcal{X}} |C_n(x,A)| \leq C_{4.3}(1+A)(\log n)^{-\underline{p}/\varsigma}\right) \to 1\,,$$

where $C_{4.3} > 0$ is a universal constant.

It follows immediately from Theorem 4.3(i) that the data-driven UCBs $C_n(x,A)$ have uniform coverage over $\mathcal{G}$ for any sufficiently large constant A. Again our recommended choice $A_n = 0.25 \log\log n$ from Section 3 ensures that the UCBs are asymptotically valid over $\mathcal{G}$ and contract at rate that is within a $\log\log n$ factor of the optimal sup-norm rate if the true smoothness is $p = \underline{p}$.

Remark 4.4 *The factor $\tilde{J}^{-\underline{p}/d}$ in the definition of the UCB from (9) represents a possibly conservative upper bound on the order of the bias component $\|\Pi_{\tilde{J}} h_0 - h_0\|_\infty$. If the true degree of smoothness is $p > \underline{p}$, then this term is conservative and the UCB does not contract at the* *minimax rate of estimation. This raises the question as to whether it is possible to construct UCBs that contract at the minimax rate of estimation in severely ill-posed settings. As stated in Chapter* 8.3 *of Giné and Nickl (2016), the existence of a rate-adaptive UCB implicitly requires the estimation of certain aspects of the unknown function, such as its smoothness, to be feasible. In mildly ill-posed settings, the condition* $h_0 \in \mathcal{G}^p$ *is sufficient to ensure that* $\tilde{J}$ *diverges at the oracle rate* $(n/\log n)^{d/(2(p+\varsigma)+d)}$*. As it turns out, this means* $\tilde{J}$ *is sufficiently informative about the unknown smoothness* p *to facilitate the construction of adaptive UCBs. In severely ill-posed models the oracle choice of* J *is* $J_{0,n} = (\alpha \log n)^{d/\varsigma}$ *for some* $0 < \alpha < (2C)^{-1}$*. Noticeably, here the oracle choice is independent of* p*. Therefore, the adaptivity of* $\tilde{J}$ *cannot be used to ascertain any information about* p*. We conjecture that this negative result is not specific to our choice of UCB construction: any UCB that is centered around an adaptive estimator that aims to mimic the oracle* $\hat{h}_{J_{0,n}}$ *over a sufficiently non-trivial class of functions (such as* $\mathcal{G}$*) will likely face the same "identifiability" problem of recovering information about* p *from* $(J_{0,n})_{n=1}^{\infty}$*.*

Remark 4.5 *It is helpful to compare the performance of our data-driven UCBs with undersmoothed UCBs in the severely ill-posed case in which* $\tau_J \asymp e^{CJ^{\varsigma/d}}$ *for some* $C, \varsigma > 0$*. Undersmoothing is possible in theory by choosing* $J = (\frac{1}{2C}\log n - \frac{d+2ap}{2C\varsigma}\log\log n)^{d/\varsigma}$ *with* $a \in (0,1)$*. Undersmoothed UCBs using this* J *sequence will contract at a slower rate than our data-driven UCBs if* $a < \underline{p}/p$*. Note that choosing any constant in the exponent different from* ς *or any constant multiplying* $\log n$ *different from* $1/(2C)$ *will result in either a failure undersmoothing (i.e., the bias term will dominate) or a potential failure of consistency of the undersmoothed estimator; a similar point is made in Remark 1 of Horowitz and Lee (2012).*

4.4 Adaptive Uniform Confidence Bands for Derivatives

Our data-driven UCBs for h_0 extend naturally to UCBs for derivatives $\partial^a h_0$ of h_0, as we now describe. UCBs for derivatives are implemented in exactly the same way as UCBs for h_0, except that we now use a bootstrap t-statistic process for $\partial^a h_0$, namely

$$\mathbb{Z}_n^{a*}(x,J) = \frac{1}{\|\hat{\sigma}_{x,J}^a\|_{sd}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \hat{L}_{J,x}^a b_{W_i}^{K(J)} \hat{u}_{i,J} \varpi_i \right),$$

where $\|\hat{\sigma}_{x,J}^a\|_{sd}^2 = \hat{L}_{J,x}^a \widehat{\Omega}_{J,J} (\hat{L}_{J,x}^a)'$ and $\hat{L}_{J,x}^a = (\partial^a \psi_x^J)' [\widehat{S}_J' \widehat{G}_{b,J}^{-1} \widehat{S}_J]^{-1} \widehat{S}_J' \widehat{G}_{b,J}^{-1}$ with $\partial^a \psi_x^J$ denoting the derivative applied element-wise: $\partial^a \psi_x^J = (\partial^a \psi_{J1}(x), \ldots, \partial^a \psi_{JJ}(x))'$. Let $z_{1-\alpha}^{a*}$ denote the $1-\alpha$ quantile of $\sup_{(x,J)\in\mathcal{X}\times\hat{\mathcal{J}}} |\mathbb{Z}_n^{a*}(x,J)|$. Our first UCB for $\partial^a h_0$ is analogous to the UCB for h_0 from (8), namely

$$C_n^a(x) = \left[\partial^a \hat{h}_{\tilde{J}}(x) - (z_{1-\alpha}^{a*} + A_n \theta^*(\hat{\alpha})) \frac{\|\hat{\sigma}_{x,\tilde{J}}^a\|_{sd}}{\sqrt{n}},\ \partial^a \hat{h}_{\tilde{J}}(x) + (z_{1-\alpha}^{a*} + A_n \theta^*(\hat{\alpha})) \frac{\|\hat{\sigma}_{x,\tilde{J}}^a\|_{sd}}{\sqrt{n}} \right]. \quad (13)$$

We recommend using this UCB in practice when $\tilde{J} = \hat{J}$ from Step 3 (see display (6)).

We require an additional condition to establish the theoretical properties of our UCBs for derivatives. Let $\|\sigma^a_{x,J}\|^2_{sd} = L^a_{J,x}\Omega_J(L^a_{J,x})'$ with $L^a_{J,x} = (\partial^a\psi^J_x)'[S'_J G^{-1}_{b,J} S_J]^{-1} S'_J G^{-1}_{b,J}$. Also let $\|\sigma^a_{x,J}\|^2 = [\partial^a\psi^J_x]'[S'_J G^{-1}_{b,J} S_J]^{-1}[\partial^a\psi^J_x]$.

Assumption 4 (continued) *(iii) There exist constants $\underline{c}, \overline{C} > 0$ for which $\underline{c}\tau^2_J J^{1+2|a|/d} \leq \inf_{x\in\mathcal{X}} \|\sigma^a_{x,J}\|^2 \leq \sup_{x\in\mathcal{X}} \|\sigma^a_{x,J}\|^2 \leq \overline{C}\tau^2_J J^{1+2|a|/d}$ for all $J \in \mathcal{T}$.*

Assumption 4(iii) is only required for the theoretical results for derivatives we establish in this subsection, and is similar to a condition from Corollary 4.1 of Chen and Christensen (2018).

Let $C^a_n(x, A)$ denote the UCB $C^a_n(x)$ from (13) when A_n is replaced by a fixed positive constant A. Analogously to Theorem 4.2, we have the following result which establishes honesty and adaptivity of $C^a_n(x, A)$ in mildly ill-posed models.

Theorem 4.4 *Let Assumptions 1-4 hold, $|a| < \underline{p}$, and suppose the model is mildly ill-posed. Then: there is a constant $A^* > 0$ (independent of α) such that for all $A \geq A^*$, we have*

(i) $$\liminf_{n\to\infty} \inf_{h_0\in\mathcal{G}} \mathbb{P}_{h_0}\left(\partial^a h_0(x) \in C^a_n(x, A) \quad \forall\ x \in \mathcal{X}\right) \geq 1 - \alpha\,;$$

(ii) $$\inf_{p\in[\underline{p},\overline{p}]} \inf_{h_0\in\mathcal{G}^p} \mathbb{P}_{h_0}\left(\sup_{x\in\mathcal{X}} |C^a_n(x, A)| \leq C_{4.4}(1 + A)\left(\frac{\log n}{n}\right)^{\frac{p-|a|}{2(p+\varsigma)+d}}\right) \to 1\,,$$

where $C_{4.4} > 0$ is a universal constant.

Remark 4.6 *As the mildly ill-posed case nests nonparametric regression as a special case with $\varsigma = 0$, Theorem 4.4 shows that our UCBs are honest and adaptive for derivatives of h_0 in general nonparametric regression models with non-Gaussian, heteroskedastic errors.*

In the severely ill-posed case we again require a larger bias-adjustment. To this end, let

$$C^a_n(x) = \left[\partial^a\hat{h}_{\tilde{J}}(x) - z^{a*}_{1-\alpha}\frac{\|\hat{\sigma}^a_{x,\tilde{J}}\|_{sd}}{\sqrt{n}} - \hat{A}^{a*}_n,\ \partial^a\hat{h}_{\tilde{J}}(x) + z^{a*}_{1-\alpha}\frac{\|\hat{\sigma}^a_{x,\tilde{J}}\|_{sd}}{\sqrt{n}} + \hat{A}^{a*}_n\right], \tag{14}$$

where

$$\hat{A}^{a*}_n = A_n \max\left\{\theta^*(\hat{\alpha})\frac{\|\hat{\sigma}^a_{x,\tilde{J}}\|_{sd}}{\sqrt{n}}\,,\ \tilde{J}^{(|a|-\underline{p})/d}\right\}.$$

We recommend using this band in practice when $\tilde{J} = \hat{J}_n$ from Step 3 (see display (6)). Let $C^a_n(x, A)$ denote the band (14) when A_n is replaced by a fixed positive constant A. Analogously to Theorem 4.3, the following result shows $C_n^a(x,A)$ is honest over $\mathcal{G}$ and shrinks at the optimal rate when h_0 is $\underline{p}$-smooth:

Theorem 4.5 *Let Assumptions 1-4 hold, $|a| < \underline{p}$, and suppose the model is severely ill-posed. Then: there is a constant $A^* > 0$ (independent of α) such that for all $A \geq A^*$, we have*

$$\text{(i)} \qquad \liminf_{n\to\infty} \inf_{h_0 \in \mathcal{G}} \mathbb{P}_{h_0}\big(\partial^a h_0(x) \in C_n^a(x,A) \quad \forall\ x \in \mathcal{X}\big) \geq 1-\alpha\,;$$

$$\text{(ii)} \qquad \inf_{p\in[\underline{p},\overline{p}]} \inf_{h_0\in\mathcal{G}^p} \mathbb{P}_{h_0}\left(\sup_{x\in\mathcal{X}} |C_n^a(x,A)| \leq C_{4.5}(1+A)(\log n)^{(|a|-\underline{p})/\varsigma}\right) \to 1\,,$$

where $C_{4.5} > 0$ is a universal constant.

We conclude this subsection by presenting one-sided versions of our UCBs for $\partial^a h_0$, which are useful for testing shape restrictions such as monotonicity, concavity, or convexity. Lower and upper UCBs for $\partial^a h_0$ in the mildly ill-posed case are

$$C_{L,n}^a(x) = \left[\partial^a \hat{h}_{\tilde{J}}(x) - (z_{L,1-\alpha}^{a*} + A_n\theta^*(\hat{\alpha}))\frac{\|\hat{\sigma}_{x,\tilde{J}}^a\|_{sd}}{\sqrt{n}},\ \infty\right),$$

$$C_{U,n}^a(x) = \left(-\infty,\ \partial^a \hat{h}_{\tilde{J}}(x) + (z_{U,1-\alpha}^{a*} + A_n\theta^*(\hat{\alpha}))\frac{\|\hat{\sigma}_{x,\tilde{J}}^a\|_{sd}}{\sqrt{n}}\right],$$

where the critical value $z_{L,1-\alpha}^{a*}$ is the $1-\alpha$ quantile of $\sup_{(x,J)\in\mathcal{X}\times\hat{\mathcal{J}}} \mathbb{Z}_n^{a*}(x,J)$ and $-z_{U,1-\alpha}^{a*}$ is the α quantile of $\inf_{(x,J)\in\mathcal{X}\times\hat{\mathcal{J}}} \mathbb{Z}_n^{a*}(x,J)$. One-sided UCBs in the severely ill-posed case are constructed similarly, replacing $A_n\theta^*(\hat{\alpha})\|\hat{\sigma}_{x,\tilde{J}}^a\|_{sd}/\sqrt{n}$ with $\hat{A}_n^{a*}$. The uniform coverage guarantees established in Theorem 4.4(i) and 4.5(i) extend to these one-sided UCBs in the mildly and severely ill-posed cases, respectively.

5 Additional Simulations

In this section we present additional simulation results to complement those presented in Section 2.1. We first present coverage properties of the data-driven band $C_n(x,A)$ from (8) for different A for the Engel curve design from Section 2.1. We then present a simulation design for a challenging nonlinear nonparametric regression. We finally present another simulation design for a NPIV model with a non-monotonic structure function h_0.

5.1 Additional Results for Section 2.1

The UCBs reported in Section 2.1 use our recommended default choice of $A_n = 0.25 \log\log n$, which is guaranteed to deliver asymptotically valid coverage irrespective of the design, and are therefore slightly conservative. To investigate how coverage depends on A, Table 2 shows the coverage of our 95% bands $C_n(x, A)$ for h_0 across simulations for different A. It appears that any $A \geq 0.5$ delivers valid coverage for the Engel curve design in Section 2.1.

NPIV	A					
n	0.00	0.01	0.05	0.10	0.50	1.00
1250	0.855	0.858	0.877	0.901	0.982	0.999
2500	0.855	0.865	0.886	0.907	0.985	1.000
5000	0.760	0.772	0.813	0.845	0.978	0.998
10000	0.467	0.488	0.571	0.644	0.951	0.999

Table 2: Coverage of data-driven 95% UCBs $C_n(x, A)$ for h_0 for different A (columns 2-7) across 1000 Monte Carlo simulations for the Engel curve design from Section 2.1, with 1000 bootstrap replications per simulation.

5.2 Nonparametric Regression

For this nonparametric regression (NPR) design, we simulate $X \sim U[0,1]$ and $U \sim N(0,1)$ independently, and then set

$$Y = \sin(15\pi X)\cos(X) + U\,. \tag{15}$$

The conditional mean function $h_0(x) = \sin(15\pi x)\cos(x)$ is very wiggly over $[0,1]$ and requires a high value of J to be selected in order to well approximate h_0 (see Figure 3). For each simulated data set, we compute our data-driven estimator $\hat{h}_{\tilde{J}}$ with $\tilde{J}$ chosen using the procedure described in Section 3.1 and data-driven UCBs for h_0 as described in display (8). As in Section 2.1, we compare our data-driven procedures to undersmoothing. We choose the undersmoothed sieve dimension J^u by setting its corresponding resolution level (see Appendix A) to be $L^u = \lfloor \tilde{L} + \log_2(\log(n)) \rfloor$, where $\tilde{L}$ is the resolution level corresponding to $\tilde{J}$ and $\lfloor a \rfloor$ denotes the largest integer less than or equal to a. This choice ensures that J^u diverges faster than $\tilde{J}$ by a factor of $\log n$. We compute the undersmoothed estimator $\hat{h}_{J^u}$ and UCBs that are centered at $\hat{h}_{J^u}$ and whose widths are determined using the bootstrap-based approach of Chen and Christensen (2018).[11] Results are presented in Table 3 for a B-spline basis of order 3.

While this simulation design is very different from that in Section 2.1, the findings presented in Table 3 are similar. In particular, the maximal estimation error of our data-driven estimator is several multiples smaller than that of the undersmoothed estimator. Moreover, undersmoothed UCBs are relatively wider than our data-driven UCBs for large n, despite the fact that our data-driven UCBs still provide (conservative) coverage guarantees. Here the conservativeness arises because of our default choice of $A_n = 0.25 \log\log n$, which is guaranteed to deliver asymptotically valid coverage irrespective of the design. For this NPR design, a much

[11] The UCB construction from Chen and Christensen (2018) implicitly assumes the chosen sieve dimension is non-random whereas our undersmoothed UCBs use a random choice $\tilde{J}^u$ and are therefore possibly too narrow.

NPR	$\|\hat{h}_{\tilde{J}} - h_0\|_\infty$		$\|\hat{h}_{J^u} - h_0\|_\infty$		Coverage		RMW	
n	mean	med	mean	med	90%	95%	mean	med
1250	0.590	0.534	1.648	1.349	0.999	0.999	1.499	1.195
2500	0.399	0.362	0.907	0.823	1.000	1.000	1.200	1.140
5000	0.290	0.262	0.955	0.868	1.000	1.000	1.707	1.610
10000	0.206	0.188	0.645	0.591	1.000	1.000	1.726	1.669

Table 3: Mean and median maximal estimation errors of $\hat{h}_{\tilde{J}}$ (data-driven) and $\hat{h}_{J^u}$ (undersmoothed) (columns 2-5), coverage of data-driven 90% and 95% UCBs for h_0 (columns 6-7), and mean and median relative maximal width (RMW) of undersmoothed 95% UCBs to data-driven 95% UCBs (columns 8-9) across 1000 Monte Carlo simulations for the NPR design (15), with 1000 bootstrap replications per simulation.

NPR	A					
n	0.00	0.01	0.05	0.10	0.50	1.00
1250	0.951	0.960	0.970	0.981	0.998	1.000
2500	0.984	0.984	0.990	0.998	1.000	1.000
5000	0.986	0.991	0.997	0.999	1.000	1.000
10000	0.991	0.993	0.998	1.000	1.000	1.000

Table 4: Coverage of data-driven 95% UCBs $C_n(x, A)$ for h_0 for different A (columns 2-7) across 1000 Monte Carlo simulations for the NPR design (15), with 1000 bootstrap replications per simulation.

smaller value of A suffices to deliver valid coverage. Table 4 shows the coverage of our 95% bands $C_n(x, A)$ for h_0 across simulations for different A, from which we see that even $A = 0$ suffices for valid coverage. The reason is that for this NPR design the set $\hat{\mathcal{J}}$ is large and the estimator $\hat{h}_J$ varies a lot across different J due to the wiggliness of h_0. Therefore, the critical value $z^*_{1-\alpha}$, which is the quantile of a sup-statistic over $\mathcal{X} \times \hat{\mathcal{J}}$, is relatively more conservative than for the NPIV designs in Section 2.1 and the next subsection. This extra conservativeness suffices to deliver valid coverage in this NPR design even with $A = 0$.

To illustrate the performance of our data-driven procedures, in Figure 3 we plot our data-driven estimator $\hat{h}_{\tilde{J}}$ and 95% UCBs for the conditional mean function for a sample of size 2500. In this sample, our data-driven choice of sieve dimension is $\tilde{J} = 34$. The data-driven estimator $\hat{h}_{\tilde{J}}$ well approximates the true conditional mean function h_0, which lies entirely within the 95% UCBs. We also plot two undersmoothed estimators using the first ("1-undersmoothed") and second ("2-undersmoothed") smallest values of J exceeding $\tilde{J}$, which are $J = 66$ and $J = 130$, respectively for this basis, along with 95% UCBs that are centered at the "undersmoothed" estimators and whose widths are determined using the procedure of Chen and Christensen (2018). The undersmoothed bands are of a similar width to our data-driven bands for this sample, even though they use a less conservative critical value which only accounts for sampling uncertainty conditional on the choice of sieve dimension. However, the undersmoothed estimator is much wigglier and does not approximate h_0 as well: the maximal estimation error

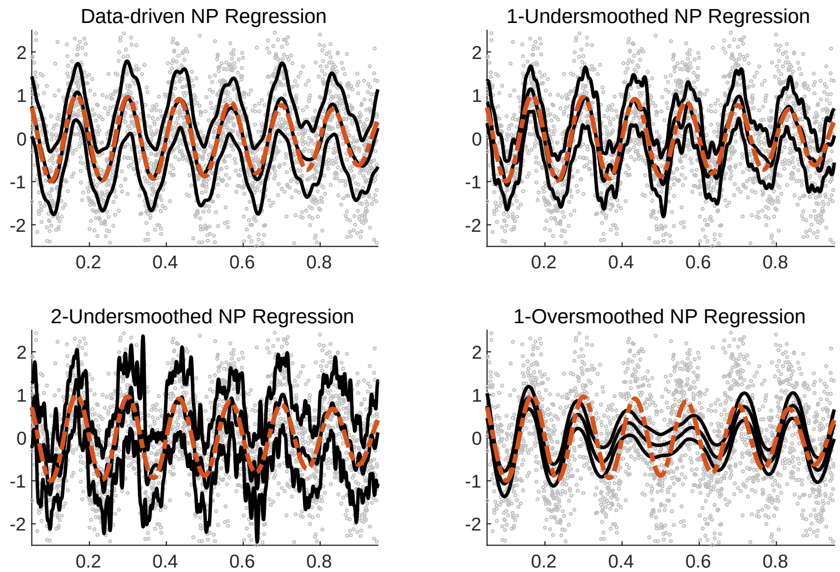

Figure 3: Estimates and 95% UCBs for h_0 (solid lines) using our data-driven procedure (top left panel) and undersmoothing (top right and bottom left panels) for a sample of size 2500 for the NPR design (15). The true conditional mean function h_0 is also shown (orange dashed lines). The bottom right panel shows estimates and 95% UCBs constructed using the largest sieve dimension less than $\tilde{J}$.

of our $\hat{h}_{\tilde{J}}$ is 0.289 while the 1-undersmoothed estimator has maximal estimation error 0.473 in this sample. Finally, in case the choice $\tilde{J} = 34$ seems unnecessarily large, we also plot an estimate of h_0 using the largest sieve dimension that is smaller than $\tilde{J}$, which is $J = 18$ for this basis ("1-oversmoothed"). It is clear that this value of J is too small, as the oversmoothed estimator fails to well approximate h_0 around the center of the support of X.

5.3 Nonparametric IV

For our second NPIV simulation design, we draw (U, V) from a bivariate normal distribution with mean zero, unit variances, and correlation 0.75, draw $Z \sim N(0, 1)$ independent of (U, V), and then set $W = \Phi(Z)$ where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, $X = \Phi(D(Z + V) + (1 - D)V)$ where D is an independent Bernoulli random variable taking the values 0 and 1 each with probability 0.5, and

$$Y = \sin(4X)\log(X) + U\,. \tag{16}$$

The structural function $h_0(x) = \sin(4x)\log(x)$ is plotted in Figure 4 and is non-monotonic and more nonlinear than the structural function for the Engel curve example in Section 2.1.

As in Section 2.1 and the previous subsection, for each simulated data set we compute

NPIV	$\|\hat{h}_{\tilde{J}} - h_0\|_\infty$		$\|\hat{h}_{J^u} - h_0\|_\infty$		Coverage		RMW	
n	mean	med	mean	med	90%	95%	mean	med
1250	0.520	0.463	1.076	1.006	1.000	1.000	1.111	1.071
2500	0.399	0.361	0.904	0.831	0.999	0.999	1.321	1.269
5000	0.308	0.278	1.218	1.164	1.000	1.000	2.441	2.389
10000	0.253	0.236	1.165	1.093	0.999	1.000	3.352	3.283

Table 5: Mean and median maximal estimation errors of $\hat{h}_{\tilde{J}}$ (data-driven) and $\hat{h}_{J^u}$ (undersmoothed) (columns 2-5), coverage of data-driven 90% and 95% UCBs for h_0 (columns 6-7), and mean and median relative maximal width (RMW) of undersmoothed 95% UCBs to data-driven 95% UCBs (columns 8-9) across 1000 Monte Carlo simulations for the NPIV design (16), with 1000 bootstrap replications per simulation.

NPIV	A					
n	0.00	0.01	0.05	0.10	0.50	1.00
1250	0.989	0.990	0.995	0.998	1.000	1.000
2500	0.974	0.978	0.982	0.988	1.000	1.000
5000	0.957	0.961	0.974	0.984	1.000	1.000
10000	0.861	0.869	0.898	0.928	1.000	1.000

Table 6: Coverage of data-driven 95% UCBs $C_n(x, A)$ for h_0 for different A (columns 2-7) across 1000 Monte Carlo simulations for the NPIV design (16), with 1000 bootstrap replications per simulation.

our data-driven estimator $\hat{h}_{\tilde{J}}$ and UCBs from (8), and compare these with an estimator using an undersmoothed sieve dimension J^u (computed as described in the previous subsection) and undersmoothed UCBs whose widths are determined using the bootstrap-based procedure of Chen and Christensen (2018). Results are presented in Table 5 for a B-spline basis which is constructed as described in Appendix A.1 with $r = 4$ and $q = 2$.

Here we again see that the maximal estimation errors of our data-driven estimator is several times smaller than that of the undersmoothed estimator, especially for large sample sizes. Our data-driven UCBs are around 10% narrower than undersmoothed UCBs for the smaller sample sizes, and around 70% narrower than undersmoothed UCBs for the larger sample sizes, even though our data-driven bands are conservative using our default choice $A_n = 0.25 \log\log n$. As seen in Table 6, a value $A > 0$ is required for correct coverage for this NPIV design, by contrast with the previous nonparametric regression design. The reason for this difference is that here the set $\hat{\mathcal{J}}$ is fairly small, so the critical value $z^*_{1-\alpha}$ is relatively less conservative. For this NPIV design, a value of $A > 0.1$, such as our default choice $A_n = 0.25 \log\log n$, suffices for correct coverage.

In Figure 4 we plot our data-driven estimator $\hat{h}_{\tilde{J}}$ and 95% UCBs for h_0 for a sample of size 2500, alongside corresponding estimates and UCBs based on undersmoothing. In this sample, our data-driven procedure chooses $\tilde{J} = 4$, the undersmoothed estimator uses $J^u = 7$. While both UCBs contain the true structural function h_0, the data-driven bands are narrower

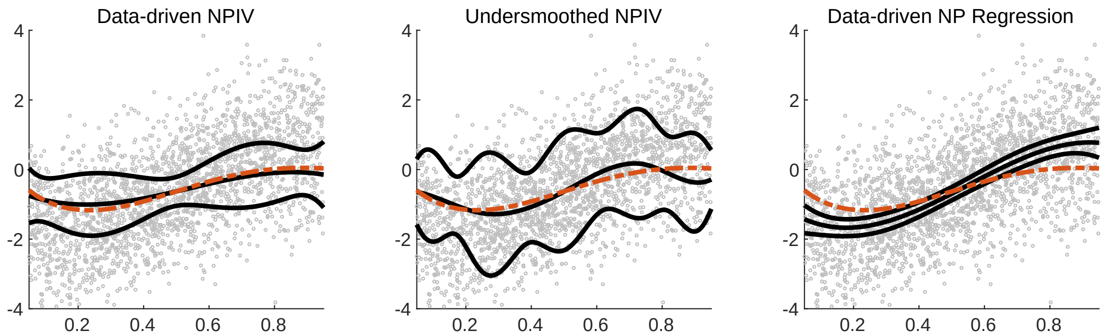

Figure 4: Estimated structural function h_0 and 95% UCBs for h_0 (solid lines) using our data-driven procedure (left panel) and undersmoothing (center panel) for a sample of size 2500 for the NPIV design (16). The true structural function h_0 is also shown (orange dashed lines). The right panel shows the estimated conditional mean of Y given X and 95% UCBs for the conditional mean function.

and more accurately convey the shape of h_0 than the undersmoothed bands, which are much more wiggly. For comparison, we also show data-driven estimates of the conditional mean function $\mathbb{E}[Y|X]$ and its UCBs. Although our data-driven procedure chooses $\tilde{J} = 4$ for both NPIV and nonparametric regression implementations, the maximal widths of the data-driven bands for NPIV is just under four times the maximal width of the bands for nonparametric regression. Note, however, that the true structural function h_0 falls outside the UCBs for the conditional mean function over almost all of the support of X, again illustrating the importance of estimating the structural function using instrumental variables for this design.

6 Conclusion

We introduce computationally simple, data-driven procedures for adaptive estimation and honest, adaptive UCBs for a structural function and its derivatives in nonparametric models using instrumental variables. Our first contribution is a data-driven choice of sieve dimension for sieve NPIV estimators. With this data-driven choice, estimators of the structural function and its derivatives converge at the minimax sup-norm rate, both for nonparametric regression models and NPIV models in both the mild and severe regimes. Our second contribution is a data-driven procedure for constructing UCBs for the structural function and its derivatives. The UCBs guarantee coverage uniformly over a generic class of data-generating processes and contract at the minimax sup-norm rate for nonparametric regression and NPIV models in the mild regime, and at near-optimal rates for NPIV models in the severe regime. We illustrate the usefulness of our procedures with an empirical application to international trade and several simulations, including an empirically relevant Engel curve design and a highly nonlinear nonparametric regression design. Although our methodology and theory are currently developed for model (1), they can be readily extend to other related models, such as partially linear NPIV and additive NPIV models.

A Basis Functions and Hölder Classes

A.1 B-splines

We first describe the construction of B-spline bases in the univariate and multivariate cases, then review some of their relevant properties.

We first consider the univariate case. The construction follows DeVore and Popov (1988) and Chapter 5.2 of DeVore and Lorentz (1993). The basis is characterized by a *resolution level* $l \in \mathbb{N} \cup \{0\}$ and *order* $r \in \mathbb{N}$ (or *degree* $r-1$). Let N_r denote the r-fold convolution of the indicator function of the unit interval, $N_r = \mathbb{1}_{[0,1]} * \cdots * \mathbb{1}_{[0,1]}$ (r-times). A (dyadic) B-spline basis on $[0,1]$ with resolution level l and order r is

$$\psi_{J_1 j}(x) = N_r(2^l x + r - j)\,, \quad j = 1, \ldots, 2^l + r - 1 =: J_1\,.$$

In the multivariate case we generate bases supported on $[0,1]^d$ with resolution level l and order r by taking tensor products of univariate bases. Here each basis function is of the form $\prod_{i=1}^d \psi_{j_i J_1}(x_i)$ with $\psi_{j_i} \in \{\psi_{J_1 1}, \ldots, \psi_{J_1 J_1}\}$. It follows that any B-spline basis of order r and resolution level l must have dimension $J = (2^l + r - 1)^d$. The set of possible sieve dimensions J as l varies over all resolution levels is therefore $\mathcal{T} = \{(2^l + r - 1)^d : l \in \mathbb{N} \cup \{0\}\}$.

As discussed further in Appendix A.3 below, the order r for the basis for the endogenous variable X should be chosen such that $r - 1 > \overline{p}$, where $\overline{p}$ is the maximal assumed degree of smoothness for h_0. Equivalently, our procedures deliver adaptivity over any smoothness range $[\overline{p}, \underline{p}]$ with $r - 1 > \overline{p} > \underline{p} > d/2$ when implemented with an order-r B-spline basis for X. Choosing r is therefore similar to choosing the order of a kernel in kernel-based nonparametric estimation.

We construct a B-spline basis $b_{K1}, \ldots, b_{KK}$ for the d_w-dimensional instrumental variable W similarly. Here we use a basis of order $r+1$ because the reduced form is smoother than h_0 (taking a conditional expectation is a smoothing operation similar to convolution). Given the resolution level l for the basis for X, the resolution level for the basis for W is $l_w = \lceil (l+q)d/d_w \rceil$ for some $q \in \mathbb{N} \cup \{0\}$, where $\lceil a \rceil$ denotes the smallest integer greater than or equal to a. Linking l_w to l in this manner defines a mapping $K(J)$ between the two bases that satisfies $\lim_{J \to \infty} K(J)/J = c \in (0, \infty)$, which is a condition Chen and Christensen (2018) use to establish that sieve NPIV estimators can attain their optimal sup-norm rates. By analogy with

NPIV	$q=0$		$q=1$		$q=2$		$q=3$		$q=4$	
n	mean	med	mean	med	mean	med	mean	med	mean	med
1250	0.158	0.124	0.151	0.118	0.139	0.112	0.146	0.123	0.201	0.183
2500	0.112	0.090	0.112	0.090	0.107	0.087	0.121	0.096	0.180	0.158
5000	0.097	0.078	0.095	0.077	0.090	0.075	0.112	0.083	0.162	0.144
10000	0.086	0.071	0.084	0.070	0.081	0.068	0.099	0.072	0.134	0.112

Table 7: Mean and median maximal estimation errors $\|\hat{h}_{\tilde{J}} - h_0\|_\infty$ of our data-driven estimator $\hat{h}_{\tilde{J}}$ across 1000 Monte Carlo simulations for the Engel curve design from Section 2.1, with 1000 bootstrap replications per simulation.

TSLS estimators for linear models, we clearly also need $K(J) \geq J$. In practice, we recommend taking q as the second- or third-smallest value for which $K(J) \geq J$ holds for all J (i.e., $q = 1$ or $q = 2$ if both X and W are of the same dimension). We advise against choosing q any larger, as the number of basis functions increases exponentially in the resolution level.

In Table 7 we report the average and median maximal estimation errors for our data-driven estimator $\hat{h}_{\tilde{J}}$ across simulations for the design from Section 2.1. Results shown are for a B-spline of order 3 for $\psi_{J1}, \ldots, \psi_{JJ}$ and order 4 for $b_{K1}, \ldots, b_{KK}$, where K and J are linked via $l_w = l + q$ for different values of q.

We now review some properties of B-spline bases that are used (sometimes implicitly) in the technical arguments below. The following Lemma summarizes Lemmas E.1 and E.2 of Chen and Christensen (2018). Let $\zeta_{\psi,J} = \sup_{x\in[0,1]^d} \|G_{\psi,J}^{-1/2}\psi_x^J\|_{\ell^2}$.

Lemma A.1 *Let Assumption 1(i) hold. Then: there are constants $C_\psi, a_\zeta > 0$ depending only on a_f for which*

(i) $\sup_{x\in[0,1]^d} \|\psi_x^J\|_{\ell^1} \leq C_\psi$;
(ii) $C_\psi^{-1} J^{-1} \leq \lambda_{\min}(G_{\psi,J}) \leq \lambda_{\max}(G_{\psi,J}) \leq C_\psi J^{-1}$;
(iii) $\sqrt{J} \leq \zeta_{\psi,J} \leq a_\zeta \sqrt{J}$.

Our choice of basis $b_{K1}, \ldots, b_{KK}$ for W satisfies essentially the same properties, in view of the fact that $J \leq K(J) \lesssim J$. Let $\zeta_{b,J} = \sup_{w\in[0,1]^{d_w}} \|G_{b,J}^{-1/2} b_w^{K(J)}\|_{\ell^2}$.

Corollary A.1 *Let Assumption 1(i) hold. Then: there are constants $C_b, a_\zeta > 0$ depending only on a_f for which*

(i) $\sup_{w\in[0,1]^{d_w}} \|b_w^{K(J)}\|_{\ell^1} \leq C_b$;
(ii) $C_b^{-1} J^{-1} \leq \lambda_{\min}(G_{b,J}) \leq \lambda_{\max}(G_{b,J}) \leq C_b J^{-1}$;
(iii) $\sqrt{J} \leq \zeta_{b,J} \leq a_\zeta \sqrt{J}$.

We also use some continuity properties of the basis in the proofs. Note that $N_r(\cdot)$ is $r-1$ times continuously differentiable on $(0, r)$. It therefore follows by Lemma A.1(ii) that the basis functions are Hölder continuous, in the sense that $\|G_{\psi,J}^{-1/2}([\psi_{x_1}^J]-[\psi_{x_2}^J])\|_{\ell^2} \le CJ^\omega \|x_1-x_2\|_{\ell^2}^{\omega'}$ holds for some positive constants C, ω, ω'. Finally, this basis also satisfies a Bernstein inequality (or inverse estimate): $\|\partial^a f\|_\infty \lesssim J^{|a|/d}\|f\|_\infty$ holds for any $f \in \Psi_J$ (the closed linear subspace of $L^2(X)$ spanned by $\psi_{J1}, \dots, \psi_{JJ}$) and multi-index a with $|a| < r-1$.

A.2 CDV Wavelets

We first describe the construction of CDV wavelet bases in the univariate and multivariate cases, then review some of their relevant properties.

We first consider the univariate case. The construction follows Cohen, Daubechies, and Vial (1993); see also chapter 4.3.5 of Giné and Nickl (2016). The basis is characterized by a *resolution level* $l \in \mathbb{N}$ and an *order* $N \in \mathbb{N}$. Let (φ, ψ) be a Daubechies pair consisting of a scaling function φ and wavelet ψ of order N. The function ψ has support contained in $[-N+1, N]$ and φ has support contained in $[0, 2N-1]$. We translate φ to have support $[-N+1, N]$ as well. Let L denote the smallest integer for which $2^L \ge 2N$ and define

$$\varphi_{L,j}(x) = 2^{L/2}\varphi(2^L x - j)\,, \quad \psi_{L,j} = 2^{L/2}\psi(2^L x - j)\,, \quad j \in \{N, \dots, 2^L - N - 1\}.^{12}$$

The functions $\varphi_{L,j}$ and $\psi_{L,j}$ with $N \le j \le 2^L - N - 1$ are supported on $[2^{-L}, 1-2^{-L}]$. We augment these with N left and right boundary corrected functions $\varphi_{L,j} = \varphi_{L,j}^{\text{left}}$ and $\varphi_{L,2^L-N+j} = \varphi_{L,j}^{\text{right}}$ for $j \in \{0, \dots, N-1\}$, with support contained in $[0, (2N-1)/2^L]$ and $[1-(2N-1)/2^L, 1]$, respectively. The boundary corrected functions are constructed as a finite linear combination of translates of φ (Giné and Nickl, 2016, p. 363-364). For each $l \ge L$ we similarly augment $\psi_{l,j}$, $j \in \{N, \dots, 2^l - N - 1\}$ with boundary corrected functions $\psi_{l,j} = \psi_{l,j}^{\text{left}}$ and $\psi_{l,2^L-N+j} = \psi_{l,j}^{\text{right}}$ for $j \in \{0, \dots, N-1\}$. This yields a total of $J_1 = 2^l$ basis functions, namely

$$\{\psi_{J_1 j} : j = 1, \dots, J_1\} := \{\varphi_{L,j}\}_{j=0}^{2^L-1} \cup \left(\cup_{k=L}^{l-1}\{\psi_{k,j}\}_{j=0}^{2^k-1}\right)$$

In the multivariate case we generate bases supported on $[0,1]^d$ by taking tensor products of univariate bases. The set of possible sieve dimensions J as l varies over all resolution levels is therefore $\mathcal{T} = \{2^{ld} : l = L+1, L+2, \dots\}$.

We say that the CDV wavelet sieve space is S-regular if the Daubechies functions φ and ψ are S times continuously differentiable on $\mathbb{R}$. A S-regular basis can always be chosen by choosing the order N such that $0.18(N-1) \ge S$ (Giné and Nickl, 2016, Theorem 4.2.10(e)). As discussed further in Appendix A.3 below, the regularity S of the basis for the endogenous variable X should be chosen such that $S > \overline{p}$, where $\overline{p}$ is the maximal assumed degree of smoothness for h_0. Equivalently, our procedures deliver adaptivity over any smoothness range

[12] We use this conventional notation without confusion with the ψ_{Jj} basis functions spanning Ψ_J.

$[\overline{p}, \underline{p}]$ with $S > \overline{p} > \underline{p} > d/2$ when implemented with a S-regular CDV wavelet basis for X. As with choosing the order r of B-splines, choosing S (equivalently, N) is therefore similar to choosing the order of a kernel in kernel-based nonparametric estimation.

A CDV wavelet basis $b_{K1}, \ldots, b_{KK}$ for the d_w-dimensional instrumental variable W is constructed similarly, using a basis of regularity $S+1$. Given the resolution level l for the basis for X, the resolution level for the basis for W is $l_w = \lceil (l+q)d/d_w \rceil$ for some $q \in \mathbb{N}$. Linking l_w to l in this manner again defines a mapping $K(J)$ between the two bases that satisfies $\lim_{J\to\infty} K(J)/J = c \in (0, \infty)$. As with B-splines, we recommend that q should be the second- or third-smallest value for which $K(J) \geq J$ holds for all J.

We now review some properties of B-spline bases that are used (sometimes implicitly) in the technical arguments below. The following Lemma summarizes Lemmas E.3 and E.4 of Chen and Christensen (2018). Let $\zeta_{\psi,J} = \sup_{x\in[0,1]^d} \|G_{\psi,J}^{-1/2} \psi_x^J\|_{\ell^2}$.

Lemma A.2 *Let Assumption 1(i) hold. Then: there are constants $C_\psi, a_\zeta > 0$ depending only on a_f for which*

(i) $\sup_{x\in[0,1]^d} \|\psi_x^J\|_{\ell^1} \leq C_\psi \sqrt{J}$;
(ii) $C_\psi^{-1} \leq \lambda_{\min}(G_{\psi,J}) \leq \lambda_{\max}(G_{\psi,J}) \leq C_\psi$;
(iii) $\sqrt{J} \leq \zeta_{\psi,J} \leq a_\zeta \sqrt{J}$.

Our choice of basis $b_{K1}, \ldots, b_{KK}$ for W satisfies essentially the same properties, in view of the fact that $J \leq K(J) \lesssim J$. Let $\zeta_{b,J} = \sup_{w\in[0,1]^{d_w}} \|G_{b,J}^{-1/2} b_w^{K(J)}\|_{\ell^2}$.

Corollary A.2 *Let Assumption 1(i) hold. Then: there are constants $C_b, a_\zeta > 0$ depending only on a_f for which*

(i) $\sup_{w\in[0,1]^{d_w}} \|b_w^{K(J)}\|_{\ell^1} \leq C_b \sqrt{J}$;
(ii) $C_b^{-1} \leq \lambda_{\min}(G_{b,J}) \leq \lambda_{\max}(G_{b,J}) \leq C_b$;
(iii) $\sqrt{J} \leq \zeta_{b,J} \leq a_\zeta \sqrt{J}$.

We also use some continuity properties of the basis in the proofs. As φ and ψ are S times continuously differentiable on their supports, it follows by Lemma A.2(ii) that the basis functions are Hölder continuous, in the sense that $\|G_{\psi,J}^{-1/2}\big([\psi_{x_1}^J] - [\psi_{x_2}^J]\big)\|_{\ell^2} \leq C J^\omega \|x_1 - x_2\|_{\ell^2}^{\omega'}$ holds for some positive constants C, ω, ω'. Finally, this basis also satisfies a Bernstein inequality (or inverse estimate): $\|\partial^a f\|_\infty \lesssim J^{|a|/d} \|f\|_\infty$ holds for any $f \in \Psi_J$ (the closed linear subspace of $L^2(X)$ spanned by $\psi_{J1}, \ldots, \psi_{JJ}$) and multi-index a with $|a| < S$.

A.3 Hölder Classes

Let $B_{\infty,\infty}^p = \{h \in L^\infty([0,1]^d) : \|h\|_{B_{\infty,\infty}^p} < \infty\}$ denote the Hölder space of smoothness p where $\|\cdot\|_{B_{\infty,\infty}^p}$ denotes the Hölder norm of smoothness $p > 0$ (see Giné and Nickl (2016), pp. 370-1), and let $B^p_{\infty,\infty}(M) = \{h \in B^p_{\infty,\infty} : \|h\|_{B^p_{\infty,\infty}} \leq M\}$ denote the Hölder ball of smoothness p and radius M.

The space $B^p_{\infty,\infty}$ may equivalently be defined by the error in approximating a function using the CDV wavelet or dyadic B-spline sieves (see Giné and Nickl (2016) for the wavelet construction and DeVore and Popov (1988) for the dyadic B-spline construction). To do so, let Ψ_J be a CDV wavelet sieve space of regularity $S > p$ or dyadic B-spline sieve space of order $r > p+1$ at resolution level L_J that generates J. In either case, let $d(h, \Psi_J) = \inf_{g \in \Psi_J} \|h - g\|_\infty$. We then have

$$h \in B^p_{\infty,\infty} \iff \|h\|_\infty + \sup_{J : J \in \mathcal{T}} J^{p/d} d(h, \Psi_J) < \infty\,.$$

Moreover, $\|h\|_\infty + \sup_{J:J\in\mathcal{T}} J^{p/d} d(h, \Psi_J)$ is equivalent to $\|h\|_{B^p_{\infty,\infty}}$. But note that

$$d(h, \Psi_J) \leq \|h - \Pi_J h\|_\infty \leq (1 + \|\Pi_J\|_\infty) d(h, \Psi_J)\,,$$

by Lebesgue's lemma (DeVore and Lorentz, 1993, p. 30), where $\|\Pi_J\|_\infty := \sup_{h : \|h\|_\infty \leq 1} \|\Pi_J h\|_\infty$ is the L^∞ norm of the $L^2(X)$ projection onto Ψ_J. Previously, Huang (2003) and Chen and Christensen (2015b) established that $\|\Pi_J\|_\infty \lesssim 1$ under Assumption 1(i) when Ψ_J is spanned by a (tensor product) B-spline or CDV wavelet basis, respectively. In this case,

$$h \in B^p_{\infty,\infty} \iff \|h\|_\infty + \sup_{J:J\in\mathcal{T}} J^{p/d} \|h - \Pi_J h\|_\infty < \infty\,,$$

and $\|h\|_\infty + \sup_{J:J\in\mathcal{T}} J^{p/d} \|h - \Pi_J h\|_\infty$ is equivalent to $\|\cdot\|_{B^p_{\infty,\infty}}$.

B Technical Results and Proofs of Main Results

In this Appendix we first introduce notation. We then present technical results that are used to establish the main results. We finally prove the main results in Subsections 4.2 and 4.3.

B.1 Notation

For any given sequence $(Z_i)_{i=1}^n$ of random vectors in $\mathbb{R}^m$ and any function $g : \mathbb{R}^m \to \mathbb{R}^k$, we denote $\mathbb{E}_n[g(Z)] = \frac{1}{n} \sum_{i=1}^n g(Z_i)$. We use the following notation for vectors and matrices formed from the basis functions

$$
\begin{aligned}
\psi_x^J &= (\psi_{J1}(x), \dots, \psi_{JJ}(x))' , & b_w^K &= (b_{K1}(w), \dots, b_{KK}(w))' , \\
\zeta_{\psi,J} &= \sup_{x\in[0,1]^d} \|G_{\psi,J}^{-1/2}\psi_x^J\|_{\ell^2} , & \zeta_{b,J} &= \sup_{w\in[0,1]^{d_w}} \|G_{b,J}^{-1/2} b_w^{K(J)}\|_{\ell^2} , \\
G_{\psi,J} &= \mathbb{E}\big[\psi_X^J(\psi_X^J)'\big] , & G_{b,J} &= \mathbb{E}\big[b_W^{K(J)}(b_W^{K(J)})'\big] , \\
S_J &= \mathbb{E}\big[b_W^{K(J)}(\psi_X^J)'\big] , & S_J^o &= G_{b,J}^{-1/2}\mathbb{E}\big[b_W^{K(J)}(\psi_X^J)'\big]G_{\psi,J}^{-1/2} .
\end{aligned}
$$

Estimators of the matrices above and their orthogonalized versions are

$$
\begin{aligned}
\widehat{G}_{\psi,J} &= \mathbb{E}_n\big[\psi_X^J(\psi_X^J)'\big] , & \widehat{G}_{b,J} &= \mathbb{E}_n\big[b_W^{K(J)}(b_W^{K(J)})'\big] , \\
\widehat{G}_{\psi,J}^o &= G_{\psi,J}^{-1/2}\mathbb{E}_n\big[\psi_X^J(\psi_X^J)'\big]G_{\psi,J}^{-1/2} , & \widehat{G}_{b,J}^o &= G_{b,J}^{-1/2}\mathbb{E}_n\big[b_W^{K(J)}(b_W^{K(J)})'\big]G_{b,J}^{-1/2} , \\
\widehat{S}_J &= \mathbb{E}_n\big[b_W^{K(J)}(\psi_X^J)'\big] , & \widehat{S}_J^o &= G_{b,J}^{-1/2}\mathbb{E}_n\big[b_W^{K(J)}(\psi_X^J)'\big]G_{\psi,J}^{-1/2} .
\end{aligned}
$$

Recall that Π_J is the $L^2(X)$ projection onto Ψ_J. We also define

$$
\Delta_J h_0 = h_0 - \Pi_J h_0 , \qquad \tilde{h}_J(x) = \psi_x^{J\prime}[\widehat{S}_J'\widehat{G}_{b,J}^{-1}\widehat{S}_J]^{1-}\widehat{S}_J'\widehat{G}_{b,J}^{-1}\mathbb{E}_n[b_W^{K(J)}h_0(X)] .
$$

Sieve variances and related terms are

$$
\begin{aligned}
\|\hat{\sigma}_{x,J,J_2}\|_{sd}^2 &= \|\hat{\sigma}_{x,J}\|_{sd}^2 + \|\hat{\sigma}_{x,J_2}\|_{sd}^2 - 2\hat{\sigma}_{x,J,J_2} , & \|\hat{\sigma}_{x,J}\|_{sd}^2 &= \hat{\sigma}_{x,J,J} , \\
\|\sigma_{x,J,J_2}\|_{sd}^2 &= \|\sigma_{x,J}\|_{sd}^2 + \|\sigma_{x,J_2}\|_{sd}^2 - 2\sigma_{x,J,J_2} , & \|\sigma_{x,J}\|_{sd}^2 &= \sigma_{x,J,J} ,
\end{aligned}
$$

where

$$
\begin{aligned}
\hat{\sigma}_{x,J,J_2} &= \hat{L}_{J,x}\widehat{\Omega}_{J,J_2}(\hat{L}_{J_2,x})' , & \hat{L}_{J,x} &= [\psi_x^J]'[\widehat{S}_J'\widehat{G}_{b,J}^{-1}\widehat{S}_J]^{-1}\widehat{S}_J'\widehat{G}_{b,J}^{-1} , \\
\sigma_{x,J,J_2} &= L_{J,x}\Omega_{J,J_2}(L_{J_2,x})' , & L_{J,x} &= [\psi_x^J]'[S_J'G_{b,J}^{-1}S_J]^{-1}S_J'G_{b,J}^{-1} ,
\end{aligned}
$$

with

$$
\begin{aligned}
\widehat{\Omega}_{J,J_2} &= \mathbb{E}_n\left[\hat{u}_J\hat{u}_{J_2}b_W^{K(J)}b_W^{K(J_2)}\right]' , \quad \hat{u}_{i,J} = Y_i - \hat{h}_J(X_i) , & \widehat{\Omega}_J &= \widehat{\Omega}_{J,J} , \\
\Omega_{J,J_2} &= \mathbb{E}\left[u^2 b_W^{K(J)}b_W^{K(J_2)}\right]' , & \Omega_J &= \Omega_{J,J} .
\end{aligned}
$$

For bootstrap and related processes, we use the notation

$$\mathbb{Z}_n^*(x,J) = \frac{1}{\|\hat\sigma_{x,J}\|_{sd}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat L_{J,x} b_{W_i}^{K(J)} \hat u_{i,J}\varpi_i\right), \tag{17}$$

$$\widehat{\mathbb{Z}}_n(x,J) = \frac{1}{\|\sigma_{x,J}\|_{sd}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n L_{J,x} b_{W_i}^{K(J)} u_i\varpi_i\right), \tag{18}$$

$$\mathbb{Z}_n(x,J) = \frac{1}{\|\sigma_{x,J}\|_{sd}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n L_{J,x} b_{W_i}^{K(J)} u_i\right). \tag{19}$$

The law of the processes $\mathbb{Z}_n^*(x,J)$ and $\widehat{\mathbb{Z}}_n(x,J)$ is determined from $(\varpi_i)_{i=1}^n$ conditional on the data $\mathcal{Z}^n := (X_i, Y_i, W_i)_{i=1}^n$. We let $\mathbb{P}^*$ denote their probability measure (i.e., with respect to the $(\varpi_i)_{i=1}^n$ conditional on the data) and $\mathbb{E}^*$ denote expectation under $\mathbb{P}^*$.

In view of the discussion in Appendix A, we have finite positive constants a_ζ and a_b for which

$$a_\zeta \geq \zeta_{\psi,J}/\sqrt{J} \geq 1\,, \qquad a_\zeta \geq \zeta_{b,J}/\sqrt{K(J)} \geq 1\,, \qquad a_b \geq K(J)/J\,.$$

By Lemma A.1 of Chen and Christensen (2018), under Assumptions 1(i) and 3(i) there is a finite positive constant a_τ for which

$$a_\tau^{-1} s_J^{-1} \leq \tau_J \leq s_J^{-1} \tag{20}$$

for all $J \in \mathcal{T}$.

Finally, we also shorten "with $\mathbb{P}_{h_0}$ probability approaching 1 (uniformly over $h_0 \in \mathcal{H}$)" to "wpa1 $\mathcal{H}$-uniformly". We also write $\mathcal{H}^p = \mathcal{H} \cap B^p_{\infty,\infty}(M)$ and $\mathcal{G}^p = \mathcal{G} \cap B^p_{\infty,\infty}(M)$.

B.2 Technical Results

In this Subsection we present several technical results that are used in the proofs of the main results in Subsections 4.2 and 4.3. The proofs of these technical results are provided in the Online Supplemental Appendix.

We first state two preliminary lemmas used in the proof of Theorem 4.1. The first relates to resolution levels in the mildly ill-posed case. For $D > 0$ and $p \in [\underline{p}, \overline{p}]$, define

$$\begin{aligned} J_0(p,D) &= \sup\left\{J \in \mathcal{T} : \tau_J \frac{\sqrt{J}\theta^*(\hat\alpha)}{\sqrt{n}} \leq D J^{-\frac{p}{d}}\right\}, \\ J_0^+(p,D) &= \inf\{J \in \mathcal{T} : J > J_0(p,D)\}\,. \end{aligned} \tag{21}$$

Lemma B.1 *Let Assumptions 1-4 hold and let* $\tau_J \asymp J^{\varsigma/d}$ *with* $\varsigma \geq 0$*. Then: with* $\bar{J}_{\max}(R)$ *as defined in (24) for any* $R > 0$ *and* $J_0^+(p, D)$ *as defined in (21) for any* $D > 0$*, we have*

$$\inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}(J_0^+(p, D) < \bar{J}_{\max}(R)) \to 1.$$

The second preliminary lemma relates to resolution levels in the severely ill-posed case. For $R > 0$ and $p \in [\underline{p}, \overline{p}]$, define

$$\bar{J}^*_{\max}(R) = \sup\left\{J \in \mathcal{T} : \tau_J J \sqrt{\log J} \leq R\sqrt{n}\right\}, \tag{22}$$

$$M_0(p, R) = \sup\{J \in \mathcal{T} : \tau_J J^{\frac{p}{d}+\frac{1}{2}} \sqrt{\log J} \leq R\sqrt{n}\}, \tag{23}$$

$$M_0^+(p, R) = \inf\{J \in \mathcal{T} : J > M_0(p, R)\}.$$

Note that $M_0(p, R)$ is (weakly) decreasing in p. In particular, as $\overline{p}/d + 1/2 \geq \underline{p}/d + 1/2 > 1$, we have $\bar{J}^*_{\max}(R) \geq M_0(\underline{p}, R) \geq M_0(p, R) \geq M_0(\overline{p}, R)$ for each R and each $p \in [\underline{p}, \overline{p}]$.

Lemma B.2 *Let* $\tau_J \asymp \exp(CJ^{\varsigma/d})$ *for some* $C, \varsigma > 0$*. Then for any* $R > 0$*, the inequality* $M_0^+(\overline{p}, R) \geq J^*_{\max}(R)$ *holds for all* n *sufficiently large.*

B.2.1 Uniform-in-J Convergence Rates for $\hat{h}_J$

For any positive constant R, define

$$\bar{J}_{\max}(R) = \sup\left\{J \in \mathcal{T} : J\sqrt{\log J}\left[(\log n)^4 \vee \tau_J\right] \leq R\sqrt{n}\right\}. \tag{24}$$

Recall that $\Delta_J h_0 = h_0 - \Pi_J h_0$. The main result we will prove in this subsection is the following:

Theorem B.1 *Let Assumptions 1, 2(i), and 3 hold, and for any positive constant* R *let* $\bar{J}_{\max} \equiv \bar{J}_{\max}(R)$*. Then: there exists a universal constant* $C_{B.1} > 0$ *such that*

$$(i) \quad \inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\|\tilde{h}_J - h_0\|_\infty \leq C_{B.1}\|\Delta_J h_0\|_\infty \quad \forall \;\; J \in \mathcal{T} \cap [1, \bar{J}_{\max}]\right) \to 1\,,$$

$$(ii) \quad \inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\|\hat{h}_J - \tilde{h}_J\|_\infty \leq C_{B.1}\tau_J \frac{\sqrt{J \log \bar{J}_{\max}}}{\sqrt{n}} \quad \forall \;\; J \in \mathcal{T} \cap [1, \bar{J}_{\max}]\right) \to 1\,.$$

B.2.2 Uniform-in-J Estimation of Sieve Variance Terms

Recall the definition of $\bar{J}_{\max}(R)$ from (24). In the remainder of this subsection, for any fixed $R > 0$, let $\bar{J}_{\max} \equiv \bar{J}_{\max}(R)$. Also let $J_{\min} \to \infty$ arbitrarily slowly. Given $\bar{J}_{\max}$ and $J_{\min}$, define

$$\delta_n = \tau_{\bar{J}_{\max}} \sqrt{\frac{\bar{J}_{\max} \log \bar{J}_{\max}}{n}} + \left(\frac{\bar{J}_{\max}^2 \log \bar{J}_{\max}}{n} \right)^{1/3} + J_{\min}^{-p/d}, \tag{25}$$

$\mathcal{J}_n = \{J \in \mathcal{T} : J_{\min} \le J \le \bar{J}_{\max}\}$, and $\mathcal{S}_n = \{(x, J, J_2) \in \mathcal{X} \times \mathcal{J}_n \times \mathcal{J}_n : J_2 > J\}$. The main result we will prove in this subsection is the following:

Lemma B.3 *Let Assumptions 1-4 hold. Then: there exists universal constants $C_{B.3} > 0$ and $N_{B.3} \in \mathbb{N}$ such that:*

(i) for every $x \in \mathcal{X}$ and $J, J_2 \in \mathcal{T}$ with $J_2 > J \ge N_{B.3}$, we have

$$C_{B.3}^{-1} \|\sigma_{x,J_2}\|_{sd} \le \|\sigma_{x,J,J_2}\|_{sd} \le C_{B.3} \|\sigma_{x,J_2}\|_{sd}\,;$$

(ii) we have

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(\sup_{(x,J,J_2) \in \mathcal{S}_n} \left| \frac{\|\hat{\sigma}_{x,J,J_2}\|_{sd}}{\|\sigma_{x,J,J_2}\|_{sd}} - 1 \right| \le C_{B.3} \delta_n \right) \to 1\,.$$

Lemma B.4 *Let Assumptions 1-3 hold. Then: there is a universal constant $C_{B.4} > 0$ such that*

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(\sup_{(x,J,J_2) \in \mathcal{S}_n} \frac{|\hat{\sigma}_{x,J,J_2} - \sigma_{x,J,J_2}|}{\|\sigma_{x,J}\|_{sd} \|\sigma_{x,J_2}\|_{sd}} \le C_{B.4} \delta_n \right) \to 1\,.$$

In particular,

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(\sup_{(x,J) \in \mathcal{X} \times \mathcal{J}_n} \left| \frac{\|\hat{\sigma}_{x,J}\|_{sd}^2}{\|\sigma_{x,J}\|_{sd}^2} - 1 \right| \le C_{B.4} \delta_n \right) \to 1\,.$$

B.2.3 Uniform Consistency of $\hat{J}_{\max}$

For the following lemma, recall $\hat{J}_{\max}$ from (4) and $\bar{J}_{\max}(R)$ from (24). The main result that we prove in this subsection is the following:

Lemma B.5 *Let Assumptions 1-3 hold. Then: replacing $10\sqrt{n}$ with $M\sqrt{n}$ for any $M > 0$ in the definition of $\hat{J}_{\max}$ from (4), there exists $R_1, R_2 > 0$ which satisfy*

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(\bar{J}_{\max}(R_1) \le \hat{J}_{\max} \le \bar{J}_{\max}(R_2) \right) \to 1\,.$$

Remark B.1 *For any $R_2 \ge R_1 > 0$ there exists a finite positive constant C for which*

$$\bar{J}_{\max}(R_1) \le \bar{J}_{\max}(R_2) \le C \bar{J}_{\max}(R_1)\,.$$

Lemma B.5 therefore provides an asymptotic rate of divergence for $\hat{J}_{\max}$.

B.2.4 Uniform-in-J Bounds for the Bootstrap

For the following Lemma, recall the process $\mathbb{Z}_n^*(x, J, J_2)$ from (5), and the set $\hat{\mathcal{S}}$ and critical value $\theta^*(\hat{\alpha})$ from Section 3.1. The first main result that we prove in this subsection is the following:

Lemma B.6 *Let Assumptions 1-4 hold. Then with $\bar{J}_{\max}(R)$ as defined in (24) for any $R > 0$, there exists constants $C_4, C_5 > 0$ for which*

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(C_4 \sqrt{\log \bar{J}_{\max}(R)} \leq \theta^*(\hat{\alpha}) \leq C_5 \sqrt{\log \bar{J}_{\max}(R)} \right) \to 1.$$

The second is a companion result concerning the critical value involved in the uniform confidence band construction:

Lemma B.7 *Let Assumptions 1-4 hold. For a given $\alpha \in (0,1)$, let $z^*_{1-\alpha}$ denote the $1-\alpha$ quantile of $\sup_{(x,J)\in\mathcal{X}\times\hat{\mathcal{J}}} |\mathbb{Z}_n^*(x, J)|$. Then: with $\bar{J}_{\max}(R)$ as defined in (24) for any $R > 0$,there exists a constant $C_{B.7} > 0$ for which*

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(z^*_{1-\alpha} \leq C_{B.7} \sqrt{\log \bar{J}_{\max}(R)} \right) \to 1\,.$$

B.2.5 Uniform Consistency for the Bootstrap

Recall $\bar{J}_{\max}(R)$ from (24). In this subsection, for any fixed $R > 0$, let $\bar{J}_{\max} \equiv \bar{J}_{\max}(R)$. Also let $J_{\min} \to \infty$ with $J_{\min} \leq \bar{J}_{\max}$. Define $\mathcal{J}_n = \{J \in \mathcal{T} : J_{\min} \leq J \leq \bar{J}_{\max}\}$ and $\mathcal{S}_n = \{(x, J, J_2) \in \mathcal{X} \times \mathcal{J}_n \times \mathcal{J}_n : J_2 > J\}$. The main result we prove in this subsection is:

Theorem B.2 *Let Assumptions 1-4 hold and let $J_{\min} \asymp (\log \bar{J}_{\max})^2$. Then: there exists a sequence $\gamma_n \downarrow 0$ for which the following inequalities hold wpa1 $\mathcal{H}$-uniformly:*

$$(i) \quad \sup_{s\in\mathbb{R}} \left| \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} \left| \sqrt{n}\frac{\hat{h}_J(x) - \tilde{h}_J(x)}{\|\hat{\sigma}_{x,J}\|_{sd}} \right| \leq s \right) - \mathbb{P}^*\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} |\mathbb{Z}_n^*(x, J)| \leq s \right) \right| \leq \gamma_n\,,$$

$$(ii) \quad \sup_{s\in\mathbb{R}} \left| \mathbb{P}_{h_0}\left(\sup_{(x,J,J_2)\in\mathcal{S}_n} \left| \sqrt{n}\frac{\hat{h}_J(x) - \hat{h}_{J_2}(x) - (\tilde{h}_J(x) - \tilde{h}_{J_2}(x))}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} \right| \leq s \right) - \mathbb{P}^*\left(\sup_{(x,J,J_2)\in\mathcal{S}_n} |\mathbb{Z}_n^*(x, J, J_2)| \leq s \right) \right| \leq \gamma_n\,.$$

B.3 Proofs of Main Results in Subsections 4.2 and 4.3

Proof of Theorem 4.1. We first list some constants that will be used throughout the proof. Fix $R_2 > 0$ in the definition of $\bar{J}_{\max}(R_2)$ from (24) sufficiently large so that by Lemma B.5 we have $\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}(\hat{J}_{\max} \leq \bar{J}_{\max}(R_2)) \to 1$. Let $\bar{J}_{\max} \equiv \bar{J}_{\max}(R_2)$ for the remainder of the proof. By Theorem B.1(i), there exists $C_{B.1} > 0$ which satisfies

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\|\tilde{h}_J - \Pi_J h_0\|_\infty \leq C_{B.1} \|\Pi_J h_0 - h_0\|_\infty \quad \forall\, J \in [1, \bar{J}_{\max}] \cap \mathcal{T} \right) \to 1. \tag{26}$$

For our choice of sieves, there exists $B_2 > 0$ which satisfies

$$\sup_{p \in [\underline{p}, \bar{p}]} \sup_{h_0 \in \mathcal{H}^p} J^{\frac{p}{d}} \|\Pi_J h_0 - h_0\|_\infty \leq B_2 \quad \forall\ J \in \mathcal{T}. \tag{27}$$

Lemmas B.3 and B.5, Assumption 4(i), and the fact that $\delta_n \downarrow 0$ (cf. (25)) imply that there exists $C_2, C_3 > 0$ which satisfy

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J,J_2) \in \hat{\mathcal{S}}} \frac{\tau_{J_2}\sqrt{J_2}}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} \leq C_3 \right) \to 1\,, \ \inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J,J_2) \in \hat{\mathcal{S}}} \frac{\|\hat{\sigma}_{x,J,J_2}\|_{sd}}{\tau_{J_2}\sqrt{J_2}} \leq C_2 \right) \to 1\,. \tag{28}$$

Additionally, by Lemma B.6 there exists constants $C_4, C_5 > 0$ which satisfy

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(C_4 \sqrt{\log \bar{J}_{\max}} \leq \theta^*(\hat{\alpha}) \leq C_5 \sqrt{\log \bar{J}_{\max}} \right) \to 1. \tag{29}$$

Part (i), step 1: We verify that $\hat{J}$ achieves the optimal rate under mild ill-posedness. Fix $\xi > 1$ (we take $\xi = 1.1$ in the main text) and let $D > 0$ be such that $2B_2(C_1+1)D^{-1}C_3 < (\xi - 1)$. Recall $J_0(p, D)$ and $J_0^+(p, D)$ from (21); we drop dependence of these quantities on (p, D) hereafter to simplify notation. By Lemma B.1, $\inf_{p \in [\underline{p}, \bar{p}]} \inf_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}(J_0^+ < \bar{J}_{\max}) \to 1$. It then follows from Lemmas B.1 and B.5 that $\inf_{p \in [\underline{p}, \bar{p}]} \inf_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}(J_0^+ < \hat{J}_{\max}) \to 1$. We therefore assume for the remainder of the proof of part (i) that $J_0^+ < \hat{J}_{\max}, \bar{J}_{\max}$.

Note by Lemma B.5 that $\hat{\mathcal{J}} \subseteq \mathcal{J}_n := \{J \in \mathcal{T} : 0.1(\log \bar{J}_{\max})^2 \leq J \leq \bar{J}_{\max}\}$ wpa1 $\mathcal{H}$-uniformly. Then for all $J \in \hat{\mathcal{J}}$ with $J > J_0^+$, by the triangle inequality, displays (26) and (27), and definition of J_0, we may deduce that

$$\begin{aligned}
&\left| \|\hat{h}_J - \hat{h}_{J_0^+}\|_\infty - \|\hat{h}_J - \hat{h}_{J_0^+} - (\tilde{h}_J - \tilde{h}_{J_0^+})\|_\infty \right| \\
&\leq \|\tilde{h}_J - \Pi_J h_0\|_\infty + \|\tilde{h}_{J_0^+} - \Pi_{J_0^+} h_0\|_\infty + \|\Pi_{J_0^+} h_0 - h_0\|_\infty + \|\Pi_J h_0 - h_0\|_\infty \\
&\leq 2B_2(1 + C_1)(J_0^+)^{-p/d} \\
&\leq 2B_2(1 + C_1) D^{-1} \theta^*(\hat{\alpha}) \tau_{J_0^+} \sqrt{J_0^+/n}
\end{aligned}$$

wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$. By (28), we also have that for all $J \in \hat{\mathcal{J}}$ with $J > J_0^+$

$$\tau_{J_0^+}\sqrt{J_0^+} \le \tau_J \sqrt{J} \le C_3 \|\hat{\sigma}_{x,J_0^+,J}\|_{sd} \qquad \forall \quad x \in \mathcal{X}$$

wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$. Combining the preceding two inequalities and using the definition of D, we obtain that for all $J \in \hat{\mathcal{J}}$ with $J > J_0^+$,

$$\sup_{x\in\mathcal{X}} \sqrt{n}\frac{|\hat{h}_J(x) - \hat{h}_{J_0^+}(x)|}{\|\hat{\sigma}_{x,J_0^+,J}\|_{sd}} \le \sup_{x\in\mathcal{X}} \sqrt{n}\frac{|\hat{h}_J(x) - \hat{h}_{J_0^+}(x) - (\tilde{h}_J(x) - \tilde{h}_{J_0^+}(x))|}{\|\hat{\sigma}_{x,J_0^+,J}\|_{sd}} + (\xi - 1)\theta^*(\hat{\alpha})$$

wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$. It follows by definition of $\hat{J}$ that

$$\begin{aligned}
&\sup_{p\in[\underline{p},\overline{p}]} \sup_{h_0\in\mathcal{H}^p} \mathbb{P}_{h_0}\big(\hat{J} > J_0^+\big) \\
&\le \sup_{p\in[\underline{p},\overline{p}]} \sup_{h_0\in\mathcal{H}^p} \mathbb{P}_{h_0}\bigg(\sup_{J\in\hat{\mathcal{J}}:J>J_0^+} \sup_{x\in\mathcal{X}} \frac{\sqrt{n}|\hat{h}_{J_0^+}(x) - \hat{h}_J(x)|}{\|\hat{\sigma}_{x,J_0^+,J}\|_{sd}} > \xi\theta^*(\hat{\alpha})\bigg) \\
&\le \sup_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\bigg(\sup_{(x,J,J_2)\in\hat{\mathcal{S}}} \frac{\sqrt{n}|\hat{h}_J(x) - \hat{h}_{J_2}(x) - (\tilde{h}_J(x) - \tilde{h}_{J_2}(x))|}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} > \theta^*(\hat{\alpha})\bigg) + o(1)\,. \qquad (30)
\end{aligned}$$

To control the r.h.s. probability in (30), let $\hat{\mathcal{J}}(\tilde{J}) = \{J \in \mathcal{T} : 0.1(\log\tilde{J})^2 \le J \le \tilde{J}\}$, $\hat{\mathcal{S}}(\tilde{J}) = \{(x, J, J_2) \in \mathcal{X} \times \hat{\mathcal{J}}(\tilde{J}) \times \hat{\mathcal{J}}(\tilde{J}) : J_2 > J\}$, and let $\theta^*(\hat{\alpha}; \tilde{J})$ denote the $(1 - 0.5 \wedge \tilde{J}^{-1})$ quantile of $\sup_{(x,J,J_2)\in\hat{\mathcal{S}}(\tilde{J})} |\mathbb{Z}_n^*(x, J, J_2)|$. Then by Lemma B.5, the union bound, and Theorem B.2(ii), we have

$$\begin{aligned}
&\sup_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\bigg(\sup_{(x,J,J_2)\in\hat{\mathcal{S}}} \frac{\sqrt{n}|\hat{h}_J(x) - \hat{h}_{J_2}(x) - (\tilde{h}_J(x) - \tilde{h}_{J_2}(x))|}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} > \theta^*(\hat{\alpha})\bigg) \\
&\le \sup_{h_0\in\mathcal{H}} \sum_{\tilde{J}\in\mathcal{T}:\tilde{J}=\bar{J}_{\max}(R_1)}^{\bar{J}_{\max}(R_2)} \mathbb{P}_{h_0}\bigg(\sup_{(x,J,J_2)\in\hat{\mathcal{S}}(\tilde{J})} \frac{\sqrt{n}|\hat{h}_J(x) - \hat{h}_{J_2}(x) - (\tilde{h}_J(x) - \tilde{h}_{J_2}(x))|}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} > \theta^*(\hat{\alpha};\tilde{J})\bigg) \\
&\le \sum_{\tilde{J}\in\mathcal{T}:\tilde{J}=\bar{J}_{\max}(R_1)}^{\bar{J}_{\max}(R_2)} \Big(\tilde{J}^{-1} + \gamma_n + o(1)\Big) \to 0\,, \qquad (31)
\end{aligned}$$

because $\gamma_n \downarrow 0$ and, by our choice of sieve and Remark B.1, for some constant $C > 0$ we have

$$\begin{aligned}
\#\{J \in \mathcal{T} : \bar{J}_{\max}(R_1) \le J \le \bar{J}_{\max}(R_2)\} &\le \#\{J \in \mathcal{T} : \bar{J}_{\max}(R_1) \le J \le C\bar{J}_{\max}(R_1)\} \\
&\le \#\{l \in \mathbb{N} : \bar{J}_{\max}(R_1) \le 2^{ld} \le C\bar{J}_{\max}(R_1)\} \le C\,.
\end{aligned}$$

In view of (30), this proves $\hat{J} \le J_0^+$ wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$.

Whenever $\hat{J} \le J_0^+ < \hat{J}_{\max}, \bar{J}_{\max}$, it follows by definition of $\hat{J}$ and display (28) that wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$, we have

$$\begin{aligned}\|\hat{h}_{\hat{J}} - h_0\|_\infty &\leq \|\hat{h}_{\hat{J}} - \hat{h}_{J_0^+}\|_\infty + \|\hat{h}_{J_0^+} - h_0\|_\infty \\ &\leq C_2 \xi \theta^*(\hat{\alpha}) \tau_{J_0^+} \sqrt{J_0^+/n} + \|\hat{h}_{J_0^+} - \tilde{h}_{J_0^+}\|_\infty + \|\tilde{h}_{J_0^+} - h_0\|_\infty.\end{aligned}$$

Then by Theorem B.1, definition of J_0^+, and the lower bound on $\theta^*(\hat{\alpha})$ in display (29), we may deduce that there exists a constant $C > 0$ for which

$$\inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0} \left(\|\hat{h}_{\hat{J}} - h_0\|_\infty \leq C \theta^*(\hat{\alpha}) \tau_{J_0^+} \sqrt{J_0^+/n} \right) \to 1.$$

As the model is mildly ill-posed, there exists a constant $C' > 0$ for which $\tau_{J_0^+} \sqrt{J_0^+} \leq C' \tau_{J_0} \sqrt{J_0}$. It then follows by definition of J_0 that

$$\inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0} \left(\|\hat{h}_{\hat{J}} - h_0\|_\infty \leq C C' D J_0^{-p/d} \right) \to 1. \tag{32}$$

By the upper bound on $\theta^*(\hat{\alpha})$ in display (29) and the fact that $\sqrt{\log \bar{J}_{\max}} \asymp \sqrt{\log n}$ (because the model is mildly ill-posed), there exists a constant $E > 0$ such that by defining

$$J_n^*(p, E) = \sup \left\{ J \in \mathcal{T} : \tau_J \sqrt{(J \log n)/n} \leq E J^{-p/d} \right\}$$

we have $\inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{H}^p} \left(J_n^*(p, E) \leq J_0(p, D) \right) \to 1$. It follows from $\tau_J \asymp J^{\varsigma/d}$ that $J_n^*(p, E) \asymp (n/\log n)^{d/(2(p+\varsigma)+d)}$. The desired result now follows from (32).

Part (i), step 2: We verify that $\tilde{J}$ achieves the optimal rate under mild ill-posedness. By step 1, we have $\inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}(\hat{J} \leq J_0^+) \to 1$. As such, if we can show that $\hat{J}_n > J_0^+$ wpa1 $\mathcal{H}$-uniformly then $\tilde{J} = \hat{J}$ wpa1 $\mathcal{H}$-uniformly and the result follows by step 1.

By the lower bound on $\theta^*(\hat{\alpha})$ in display (29) and the fact that $\sqrt{\log \bar{J}_{\max}} \asymp \sqrt{\log n}$ (because the model is mildly ill-posed), we may deduce that there exists a constant $E' > 0$ such that $\inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{H}^p} \left(J_n^\dagger(p, E') \geq J_0^+(p, D) \right) \to 1$ where

$$J_n^\dagger(p, E') = \inf \left\{ J \in \mathcal{T} : \tau_J \sqrt{(J \log n)/n} > E' J^{-p/d} \right\}.$$

But note that $\max_{p \in [\underline{p}, \overline{p}]} J_0^\dagger(p, E') = J_0^\dagger(\underline{p}, E')$. The result now follows by Lemma B.5, noting that $\bar{J}_{\max}(R_1)/J_0^\dagger(\underline{p}, E') \to \infty$ when the model is mildly ill-posed because $\underline{p} > d/2$.

Part (ii), step 1: We verify that $\hat{J}_n$ achieves the optimal rate under severe ill-posedness. We do so assuming a CDV wavelet basis, though we note a similar argument applies (albeit with more complicated notation) for B-spline bases. Note that when the model is severely ill-posed, for any $R > 0$ we have $n^\beta \lesssim \tau_{\bar{J}_{\max}(R)}$ for some $\beta > 0$ and so $\tau_{\bar{J}_{\max}(R)} > (\log n)^4$ for all sufficiently large n. Therefore $\bar{J}_{\max}(R) = \bar{J}^*_{\max}(R)$ for all n sufficiently large, where $\bar{J}^*_{\max}(R)$ is defined in (22). By Theorem B.1, Lemma B.5, and Remark B.1, we may deduce that there exist constants $D, D' > 0$ for which

$$\begin{aligned}
\|\hat{h}_{\hat{J}_n} - h_0\|_\infty &\le \|\hat{h}_{\hat{J}_n} - \tilde{h}_{\hat{J}_n}\|_\infty + \|\tilde{h}_{\hat{J}_n} - h_0\|_\infty \\
&\le D\biggl((2^{-d}\bar{J}^*_{\max}(R_1))^{-\frac{p}{d}} + \tau_{2^{-d}\bar{J}^*_{\max}(R_2)}\sqrt{2^{-d}\bar{J}^*_{\max}(R_2)\log(2^{-d}\bar{J}^*_{\max}(R_2))/n}\biggr) \\
&\le D'\biggl((2^{-d}\bar{J}^*_{\max}(R_2))^{-\frac{p}{d}} + \tau_{2^{-d}\bar{J}^*_{\max}(R_2)}\sqrt{2^{-d}\bar{J}^*_{\max}(R_2)\log(2^{-d}\bar{J}^*_{\max}(R_2))/n}\biggr)
\end{aligned}$$

wpa1 uniformly over $\mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$.

Recall the definition of $M_0(p, R_2)$ from (23). By Lemma B.2, for all $p \in [\underline{p}, \overline{p}]$ we have that $M_0(p, R_2) \ge M_0(\overline{p}, R_2) \ge 2^{-d}J^*_{\max}(R_2)$ holds for all n sufficiently large, in which case by definition of $M_0(p, R_2)$ we must have

$$\tau_{2^{-d}\bar{J}^*_{\max}(R_2)}\sqrt{2^{-d}\bar{J}^*_{\max}(R_2)\log(2^{-d}\bar{J}^*_{\max}(R_2))/n} \le R_2(2^{-d}\bar{J}^*_{\max}(R_2))^{-\frac{p}{d}}\,.$$

Combining the preceding two inequalities then yields

$$\|\hat{h}_{\hat{J}_n} - h_0\|_\infty \le D'(1 + R_2)2^p(\bar{J}^*_{\max}(R_2))^{-\frac{p}{d}}$$

wpa1 uniformly over $\mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$.

It remains to show $(\log n)^{d/\varsigma} \lesssim \bar{J}^*_{\max}(R_2)$ when $\tau_J \asymp \exp(CJ^{\varsigma/d})$ for $C, \varsigma > 0$. Suppose $\liminf_{n\to\infty} \bar{J}^*_{\max}(R_2)/(\log n)^{d/\varsigma} = 0$. Then along a subsequence $\{n_k\}_{k\ge 1}$ we have $\bar{J}^*_{\max}(R_2) = (2^{-\varsigma}C^{-1}u_{n_k}\log n_k)^{d/\varsigma}$ for some sequence $u_{n_k} \downarrow 0$. Then $2^d\bar{J}^*_{\max}(R_2) \in \mathcal{T}$ satisfies

$$\tau_{2^d\bar{J}^*_{\max}(R_2)}2^d\bar{J}^*_{\max}(R_2)\sqrt{\log(2^d\bar{J}^*_{\max}(R_2))/n_k} \lesssim n_k^{u_{n_k}-\frac{1}{2}}(\log n_k)^{d/\varsigma}\sqrt{\log\log n_k} \xrightarrow[k\to\infty]{} 0\,,$$

thereby contradicting the definition of $\bar{J}^*_{\max}(R_2)$ from (22) for all sufficiently large k.

Part (ii), step 2: We verify that $\tilde{J}$ achieves the optimal rate under severe ill-posedness. For any constant $D > 0$, by definition of $\tilde{J}$ we have

$$\begin{aligned}
&\sup_{p\in[\underline{p},\overline{p}]}\sup_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}\bigl(\|\hat{h}_{\tilde{J}} - h_0\|_\infty > D(\log n)^{-p/\varsigma}\bigr) \\
&\le \sup_{p\in[\underline{p},\overline{p}]}\sup_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}\bigl(\|\hat{h}_{\hat{J}} - h_0\|_\infty > D(\log n)^{-p/\varsigma} \text{ and } \hat{J} < \hat{J}_n\bigr) \\
&\quad + \sup_{p\in[\underline{p},\overline{p}]}\sup_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}\bigl(\|\hat{h}_{\hat{J}_n} - h_0\|_\infty > D(\log n)^{-p/\varsigma}\bigr).
\end{aligned}$$

By part (ii), step 1, the constant D can be chosen sufficiently large so that the second term on the r.h.s. is $o(1)$. For the first term, note that $\|\hat{h}_{\hat{J}} - h_0\|_\infty \leq \|\hat{h}_{\hat{J}} - \hat{h}_{\hat{J}_n}\|_\infty + \|\hat{h}_{\hat{J}_n} - h_0\|_\infty$, so it suffices to show that there exists a constant $D > 0$ for which

$$\sup_{p \in [\underline{p}, \overline{p}]} \sup_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}\big(\|\hat{h}_{\hat{J}} - \hat{h}_{\hat{J}_n}\|_\infty > D(\log n)^{-p/\varsigma} \text{ and } \hat{J} < \hat{J}_n\big) \to 0\,.$$

But by definition of $\hat{J}$ and displays (28) and (29), we have

$$\begin{aligned}
&\sup_{p \in [\underline{p}, \overline{p}]} \sup_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}\big(\|\hat{h}_{\hat{J}} - \hat{h}_{\hat{J}_n}\|_\infty > D(\log n)^{-p/\varsigma} \text{ and } \hat{J} < \hat{J}_n\big) \\
&\leq \sup_{p \in [\underline{p}, \overline{p}]} \sup_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}\left(\xi C_2 \theta^*(\hat{\alpha}) \tau_{\hat{J}_n} \sqrt{\hat{J}_n/n} > D(\log n)^{-p/\varsigma}\right) + o(1) \\
&\leq \sup_{p \in [\underline{p}, \overline{p}]} \mathbb{1}\left[\xi C_2 C_5 \tau_{2^{-d}\bar{J}^*_{\max}(R_2)} \sqrt{2^{-d}\bar{J}^*_{\max}(R_2) \log(2^{-d}\bar{J}^*_{\max}(R_2))/n} > D(\log n)^{-p/\varsigma}\right] + o(1)\,.
\end{aligned}$$

By step 1, we have $\tau_{2^{-d}\bar{J}^*_{\max}(R_2)} \sqrt{2^{-d}\bar{J}^*_{\max}(R_2) \log(2^{-d}\bar{J}^*_{\max}(R_2))/n} \lesssim (\log n)^{-p/\varsigma}$ uniformly for $p \in [\underline{p}, \overline{p}]$, so the constant D can be chosen sufficiently large that the indicator function on the r.h.s. is zero uniformly for $p \in [\underline{p}, \overline{p}]$ for all n sufficiently large. ■

Proof of Corollary 4.1. Part (i): Recall $J_0^+ \equiv J_0(p, D)^+$ from (21). We have

$$\|\partial^a \hat{h}_{\hat{J}} - \partial^a h_0\|_\infty \leq \|\partial^a \hat{h}_{\hat{J}} - \partial^a \hat{h}_{J_0^+}\|_\infty + \|\partial^a \hat{h}_{J_0^+} - \partial^a \tilde{h}_{J_0^+}\|_\infty + \|\partial^a \tilde{h}_{J_0^+} - \partial^a h_0\|_\infty\,.$$

As $\hat{J} \leq J_0^+ < \hat{J}_{\max}, \bar{J}_{\max}$ holds wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$, by part (i), step 1 of the proof of Theorem 4.1, we may appeal to a Bernstein inequality (or inverse estimate) for our choice of basis to write

$$\|\partial^a \hat{h}_{\hat{J}} - \partial^a h_0\|_\infty \lesssim (J_0^+)^{|a|/d}\left(\|\hat{h}_{\hat{J}} - \hat{h}_{J_0^+}\|_\infty + \|\hat{h}_{J_0^+} - \tilde{h}_{J_0^+}\|_\infty\right) + \|\partial^a \tilde{h}_{J_0^+} - \partial^a h_0\|_\infty\,.$$

By similar arguments to the proof of Corollary 3.1 of Chen and Christensen (2018), we may also deduce $\|\partial^a \tilde{h}_{J_0^+} - \partial^a h_0\|_\infty \lesssim (J_0^+)^{(|a|-p)/d}$ and so

$$\|\partial^a \hat{h}_{\hat{J}} - \partial^a h_0\|_\infty \lesssim (J_0^+)^{|a|/d}\left(\|\hat{h}_{\hat{J}} - \hat{h}_{J_0^+}\|_\infty + \|\hat{h}_{J_0^+} - \tilde{h}_{J_0^+}\|_\infty + (J_0^+)^{-p/d}\right)\,.$$

It now follows by similar arguments to part (i), step 1 of the proof of Theorem 4.1 and definition of J_0 that there exists a constant $C > 0$ for which

$$\inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}\left(\|\partial^a \hat{h}_{\hat{J}} - \partial^a h_0\|_\infty \leq C J_0^{(|a|-p)/d}\right) \to 1.$$

The result follows from noting, as in the proof of part (i), step 1 of the proof of Theorem 4.1, that $\inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{H}^p}\big(J_n^*(p, E) \leq J_0(p, D)\big) \to 1$ where $J_n^*(p, E) \asymp (n/\log n)^{d/(2(p+\varsigma)+d)}$, and by part (i), step 2 of the proof of Theorem 4.1 (which shows that $\tilde{J} = \hat{J}$ wpa1 $\mathcal{H}$-uniformly).

Part (ii): Recall $\bar{J}^*_{\max}(R)$ from (22) and $\hat{J}_n$ from the definition of $\tilde{J}$. By similar arguments to part (ii), step 1 of the proof of Theorem 4.1, and the proof of Corollary 3.1 of Chen and Christensen (2018), we may deduce

$$\begin{aligned}&\|\partial^a \hat{h}_{\hat{J}_n} - \partial^a h_0\|_\infty \\ &\lesssim (\bar{J}^*_{\max}(R_2))^{\frac{|a|}{d}} \left((2^{-d}\bar{J}^*_{\max}(R_2))^{-\frac{p}{d}} + \tau_{2^{-d}\bar{J}^*_{\max}(R_2)} \sqrt{2^{-d}\bar{J}^*_{\max}(R_2) \log(2^{-d}\bar{J}^*_{\max}(R_2))/n} \right)\end{aligned}$$

wpa1 uniformly over $\mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$. It follows by part (ii), step 1 of the proof of Theorem 4.1 that

$$\|\partial^a \hat{h}_{\hat{J}_n} - \partial^a h_0\|_\infty \lesssim (\log n)^{(|a|-p)/d}$$

wpa1 uniformly over $\mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$.

By similar arguments to part (ii), step 2 of the proof of Theorem 4.1, it suffices to show that there exists a constant $C > 0$ for which

$$\sup_{p \in [\underline{p}, \overline{p}]} \sup_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}\big(\|\partial^a \hat{h}_{\hat{J}} - \partial^a \hat{h}_{\hat{J}_n}\|_\infty > C(\log n)^{(|a|-p)/\varsigma} \text{ and } \hat{J} < \hat{J}_n\big) \to 0\,.$$

But for any $\hat{J} \leq \hat{J}_n$ by a Bernstein inequality (or inverse estimate) for our choice of basis, we have

$$\|\partial^a \hat{h}_{\hat{J}} - \partial^a \hat{h}_{\hat{J}_n}\|_\infty \lesssim (\hat{J}_n)^{|a|/d} \|\hat{h}_{\hat{J}} - \hat{h}_{\hat{J}_n}\|_\infty \lesssim (\bar{J}^*_{\max}(R_2))^{|a|/d} \|\hat{h}_{\hat{J}} - \hat{h}_{\hat{J}_n}\|_\infty$$

wpa1 uniformly over $\mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$, where the second inequality is because $\hat{J}_n \leq \hat{J}_{\max} \leq \bar{J}_{\max}(R_2)$ wpa1 $\mathcal{H}$-uniformly by Lemma B.5 and because $\bar{J}_{\max}(R_2) = \bar{J}^*_{\max}(R_2)$ for all n sufficiently large. But note by severe ill-posedness and definition of $\bar{J}^*_{\max}(R_2)$, we have that $C(\bar{J}^*_{\max}(R_2))^{\varsigma/d} \asymp \log \tau_{\bar{J}^*_{\max}(R_2)} \leq \log(R_2\sqrt{n}) \asymp \log n$, and so $\bar{J}^*_{\max}(R_2) \lesssim (\log n)^{d/\varsigma}$. The result now follows by part (ii), step 2 of the proof of Theorem 4.1. ■

Proof of Theorem 4.2. In some of what follows, we use the fact that the sieve dimensions for CDV wavelet bases are linked via $J^+ = 2^d J$ for $J \in \mathcal{T}$. We do so for notational convenience; a similar argument (but with more complicated notation) applies for B-spline bases.

Part (i), step 1: By part (i), step 2 of the proof of Theorem 4.1, we have $\hat{J} = \tilde{J}$ wpa1 $\mathcal{H}$-uniformly. It therefore suffices to prove the claim with $\hat{J}$ in place of $\tilde{J}$. Fix $R_2 > 0$ in the definition of $\bar{J}_{\max}(R_2)$ from (24) sufficiently large so that by Lemma B.5 we have $\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}(\hat{J}_{\max} \leq \bar{J}_{\max}(R_2)) \to 1$. Let $\bar{J}_{\max} \equiv \bar{J}_{\max}(R_2)$ for the remainder of the proof. Recall the constants $C_{B.1}$ from (26), $\underline{B}$ and $\overline{B}$ from the discussion preceding the statement of this theorem, and C_4 and C_5 from (29). Also note that by Lemmas B.3 and B.5, Assumption

4(i), and the fact that $\delta_n \downarrow 0$ (cf. (25)) imply that there exists $C_2, C_3 > 0$ which satisfy

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\hat{\mathcal{J}}} \frac{\tau_J\sqrt{J}}{\|\hat{\sigma}_{x,J}\|_{sd}} \leq C_3 \right) \to 1\,, \inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\hat{\mathcal{J}}} \frac{\|\hat{\sigma}_{x,J}\|_{sd}}{\tau_J\sqrt{J}} \leq C_2 \right) \to 1\,. \quad (33)$$

Let $v = \inf_{J\in\mathcal{T}}(1+\|\Pi_J\|_\infty)^{-1} > 0$, where $\|\Pi_J\|_\infty \lesssim 1$ is the Lebesgue constant for Ψ_J (see Appendix A.3). Choose $\beta \in (0,1)$ and $E > 0$ such that $(v\underline{B}\beta^{-\underline{p}/d} - (C_{B.1}+1)\overline{B}) > 0$ and $E^{-1}(v\underline{B}\beta^{-\underline{p}/d} - (C_{B.1}+1)\overline{B}) > C_2(\xi+1)$, where $\xi > 1$ (we take $\xi = 1.1$ in the main text).

Define $J_0(p,E)$ as in (21). It is established in part (i), step 1 of the proof of Theorem 4.1 that $J_0(p,E) \gtrsim (n/\log n)^{d/(2(p+\varsigma)+d)}$. By Lemma B.5 and mild ill-posedness, for any constant $C > 0$ we have $J_0(p,E)/(\log \hat{J}_{\max})^2 \geq C$ wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p},\overline{p}]$. Therefore, $\inf\{J \in \mathcal{T} : J \geq \beta J_0(p,E)\} > \log \hat{J}_{\max}$ wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p},\overline{p}]$.

Fix any $J \in \hat{\mathcal{J}}$ with $J < \beta J_0(p,E)$ (this is justified wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p},\overline{p}]$ by the preceding paragraph) and note (dropping dependence of J_0 on (p,E))

$$\begin{aligned}
\|\hat{h}_J - \hat{h}_{J_0}\|_\infty &= \|\hat{h}_J - \hat{h}_{J_0} - \tilde{h}_J + \tilde{h}_J - \tilde{h}_{J_0} + \tilde{h}_{J_0} - h_0 + h_0\|_\infty \\
&\geq \|\tilde{h}_J - h_0\|_\infty - \|\tilde{h}_{J_0} - h_0\|_\infty - \|\hat{h}_J - \tilde{h}_J - (\hat{h}_{J_0} - \tilde{h}_{J_0})\|_\infty.
\end{aligned}$$

For a given $h_0 \in \mathcal{G}^p$, let $h_{0,J} \in \arg\min_{h\in\Psi_J} \|h - h_0\|_\infty$. Recall $\underline{J}$ from the definition of $\mathcal{G}^p$ and note that $\inf\{J : J \in \hat{\mathcal{J}}\} \geq \underline{J}$ holds wpa1 $\mathcal{H}$-uniformly by Lemma B.5. Recalling the Lebesgue constant $\|\Pi_J\|_\infty$ from Appendix A.3, we may then deduce

$$\|\tilde{h}_J - h_0\|_\infty \geq \|h_{0,J} - h_0\|_\infty \geq (1+\|\Pi_J\|_\infty)^{-1}\|h_0 - \Pi_J h_0\|_\infty \geq v\underline{B}J^{-p/d}\,,$$

for all $J \in \hat{\mathcal{J}}$ wpa1, uniformly for all $h_0 \in \mathcal{G}^p$ and all $p \in [\underline{p},\overline{p}]$. It follows by (26) and the discussion preceding the statement of this theorem that

$$\begin{aligned}
\|\hat{h}_J - \hat{h}_{J_0}\|_\infty &\geq v\underline{B}J^{-p/d} - (C_{B.1}+1)\overline{B}J_0^{-p/d} - \|\hat{h}_J - \tilde{h}_J - (\hat{h}_{J_0} - \tilde{h}_{J_0})\|_\infty \\
&\geq (v\underline{B}\beta^{-\underline{p}/d} - (C_{B.1}+1)\overline{B})J_0^{-p/d} - \|\hat{h}_J - \tilde{h}_J - (\hat{h}_{J_0} - \tilde{h}_{J_0})\|_\infty \\
&> C_2(\xi+1)\tau_{J_0}\frac{\sqrt{J_0}\theta^*(\hat{\alpha})}{\sqrt{n}} - \|\hat{h}_J - \tilde{h}_J - (\hat{h}_{J_0} - \tilde{h}_{J_0})\|_\infty\,,
\end{aligned}$$

where the second line uses $J < \beta J_0$ and the third uses definition of E and $J_0(p,E)$. It now follows by the preceding display and (33) that

$$
\begin{aligned}
&\sup_{p\in[\underline{p},\overline{p}]}\sup_{h_0\in\mathcal{G}^p}\mathbb{P}_{h_0}\big(\hat{J}<\beta J_0(p,E)\big)\\
&\le \sup_{p\in[\underline{p},\overline{p}]}\sup_{h_0\in\mathcal{G}^p}\mathbb{P}_{h_0}\left(\inf_{J\in\hat{\mathcal{J}}:J<\beta J_0}\sup_{x\in\mathcal{X}}\frac{\sqrt{n}|\hat{h}_J(x)-\hat{h}_{J_0}(x)|}{\|\hat{\sigma}_{x,J,J_0}\|_{sd}}\le \xi\theta^*(\hat{\alpha})\right)\\
&\le \sup_{p\in[\underline{p},\overline{p}]}\sup_{h_0\in\mathcal{G}^p}\mathbb{P}_{h_0}\left(\sup_{(x,J,J_2)\in\hat{\mathcal{S}}}\frac{\sqrt{n}|\hat{h}_J(x)-\hat{h}_{J_2}(x)-(\tilde{h}_J(x)-\tilde{h}_{J_2}(x))|}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}}>\theta^*(\hat{\alpha})\right)+o(1)\\
&\le \sup_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\sup_{(x,J,J_2)\in\hat{\mathcal{S}}}\frac{\sqrt{n}|\hat{h}_J(x)-\hat{h}_{J_2}(x)-(\tilde{h}_J(x)-\tilde{h}_{J_2}(x))|}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}}>\theta^*(\hat{\alpha})\right)+o(1)\to 0\,,
\end{aligned}
$$

where the final line is by (31).

Part (i), step 2: Recall $J_0^+(p,D)$ from part (i), step 1 of the proof of Theorem 4.1. By the previous step of this proof and part (i), step 1 of the proof of Theorem 4.1, we have

$$
\inf_{p\in[\underline{p},\overline{p}]}\inf_{h_0\in\mathcal{G}^p}\mathbb{P}_{h_0}\big(\beta J_0(p,E)\le\hat{J}\le J_0^+(p,D)\big)\to 1\,. \tag{34}
$$

Therefore, by (26), (33), (34), and definition of $\overline{B}$, for every $h_0\in\mathcal{G}^p$ and $x\in\mathcal{X}$ we have

$$
\frac{|\tilde{h}_{\hat{J}}(x)-h_0(x)|}{\|\hat{\sigma}_{x,\hat{J}}\|_{sd}}\le (C_{B.1}+1)C_3\overline{B}\frac{\hat{J}^{-p/d}}{\tau_{\hat{J}}\sqrt{\hat{J}}}\le (C_{B.1}+1)C_3\overline{B}\beta^{-\overline{p}/d}2^p\frac{(2^dJ_0(p,E))^{-p/d}}{\tau_{\lceil\beta J_0(p,E)\rceil}\sqrt{\beta J_0(p,E)}}\,,
$$

wpa1 uniformly for $h_0\in\mathcal{G}^p$ and $p\in[\underline{p},\overline{p}]$ and $x\in\mathcal{X}$, where $\tau_{\lceil\beta J_0(p,E)\rceil}$ denotes the ill-posedness at resolution level $\inf\{J\in\mathcal{T}:J\ge\beta J_0(p,E)\}$. It now follows from definition of $2^dJ_0(p,E)\equiv J_0^+(p,E)$ from (21) that whenever the preceding inequality holds, we have

$$
\sup_{x\in\mathcal{X}}\sqrt{n}\frac{|\tilde{h}_{\hat{J}}(x)-h_0(x)|}{\|\hat{\sigma}_{x,\hat{J}}\|_{sd}}\le C_3(C_{B.1}+1)\overline{B}\beta^{-\overline{p}/d-1/2}2^{\overline{p}+d/2}E^{-1}\frac{\tau_{2^dJ_0(p,E)}}{\tau_{\lceil\beta J_0(p,E)\rceil}}\theta^*(\hat{\alpha})<A_0\theta^*(\hat{\alpha})\,,
$$

where the final inequality holds uniformly for $h_0\in\mathcal{G}^p$ and $p\in[\underline{p},\overline{p}]$ for some constant $A_0>0$ because $\sup_{J\in\mathcal{T}}\tau_{2^dJ}/\tau_{\lceil\beta J\rceil}<\infty$ by virtue of mild ill-posedness. It follows from the preceding display that for any $A \geq A_0$,

$$\begin{aligned}
&\inf_{h_0 \in \mathcal{G}} \mathbb{P}_{h_0}\left(h_0(x) \in C_n(x,A) \ \ \forall \ \ x \in \mathcal{X}\right) \\
&\geq \inf_{p \in [\underline{p},\overline{p}]} \inf_{h_0 \in \mathcal{G}^p} \mathbb{P}_{h_0}\left(\sup_{x \in \mathcal{X}} \sqrt{n} \frac{|\hat{h}_{\hat{J}}(x) - h_0(x)|}{\|\hat{\sigma}_{x,\hat{J}}\|_{sd}} \leq z^*_{1-\alpha} + A\theta^*(\hat{\alpha}) \right) + o(1) \\
&\geq \inf_{p \in [\underline{p},\overline{p}]} \inf_{h_0 \in \mathcal{G}^p} \mathbb{P}_{h_0}\left(\sup_{x \in \mathcal{X}} \sqrt{n} \frac{|\hat{h}_{\hat{J}}(x) - \tilde{h}_{\hat{J}}(x)|}{\|\hat{\sigma}_{x,\hat{J}}\|_{sd}} \leq z^*_{1-\alpha} \right) + o(1) \\
&\geq \inf_{p \in [\underline{p},\overline{p}]} \inf_{h_0 \in \mathcal{G}^p} \mathbb{P}_{h_0}\left(\sup_{(x,J) \in \mathcal{X} \times \underline{\mathcal{J}}_n} \sqrt{n} \frac{|\hat{h}_J(x) - \tilde{h}_J(x)|}{\|\hat{\sigma}_{x,J}\|_{sd}} \leq z^*_{1-\alpha} \right) + o(1)\,,
\end{aligned}$$

where the final line is because $\hat{J} \in \underline{\mathcal{J}}_n := \{J \in \mathcal{T} : 0.1(\log \bar{J}_{\max}(R_2))^2 \leq J \leq \bar{J}_{\max}(R_1)\}$ and $\hat{\mathcal{J}} \supseteq \underline{\mathcal{J}}_n$ both hold wpa1 uniformly for $h_0 \in \mathcal{G}^p$ and $p \in [\underline{p},\overline{p}]$; the former holds by (34) and Lemma B.1 and the latter holds by Lemma B.5. Let $\underline{z}^*_{1-\alpha}$ denote the $1-\alpha$ quantile of $\sup_{(x,J) \in \mathcal{X} \times \underline{\mathcal{J}}_n} |\mathbb{Z}^*_n(x,J)|$. As $\underline{z}^*_{1-\alpha} \leq z^*_{1-\alpha}$ must hold whenever $\hat{\mathcal{J}} \supseteq \underline{\mathcal{J}}_n$, we therefore have

$$\begin{aligned}
&\inf_{h_0 \in \mathcal{G}} \mathbb{P}_{h_0}\left(h_0(x) \in C_n(x,A) \ \ \forall \ \ x \in \mathcal{X}\right) \\
&\geq \inf_{p \in [\underline{p},\overline{p}]} \inf_{h_0 \in \mathcal{G}^p} \mathbb{P}_{h_0}\left(\sup_{(x,J) \in \mathcal{X} \times \underline{\mathcal{J}}_n} \sqrt{n} \frac{|\hat{h}_J(x) - \tilde{h}_J(x)|}{\|\hat{\sigma}_{x,J}\|_{sd}} \leq \underline{z}^*_{1-\alpha} \right) + o(1) = (1-\alpha) + o(1)\,,
\end{aligned}$$

where the last equality follows from Theorem B.2(i) and the definition of $\underline{z}^*_{1-\alpha}$.

Part (ii): By Lemmas B.4, B.6, and B.7 and Assumption 4(i), we have

$$\sup_{x \in \mathcal{X}} |C_n(x,A)| \lesssim (1+A)\tau_{\hat{J}} \sqrt{(\hat{J} \log \bar{J}_{\max})/n}$$

wpa1 $\mathcal{H}$-uniformly. Then by (34) with $J_0 = J_0(p,D)$ and $\bar{A} = 1 + A$, we have that

$$\sup_{x \in \mathcal{X}} |C_n(x,A)| \lesssim \bar{A}\tau_{J_0^+} \sqrt{(J_0^+ \log \bar{J}_{\max})/n} \lesssim \bar{A}\tau_{J_0} \sqrt{(J_0 \log \bar{J}_{\max})/n} \lesssim \bar{A} \frac{\sqrt{\log \bar{J}_{\max}}}{\theta^*(\hat{\alpha})} J_0^{-p/d}$$

holds wpa1 uniformly for $h_0 \in \mathcal{G}^p$ and $p \in [\underline{p},\overline{p}]$ and for all $A > 0$, where the second inequality follows from the fact that the model is mildly ill-posed and the third is by definition (21). It follows by Lemma B.6 that there is a constant $C > 0$ (independent of A) for which

$$\inf_{p \in [\underline{p},\overline{p}]} \inf_{h_0 \in \mathcal{G}^p} \mathbb{P}_{h_0}\left(\sup_{x \in \mathcal{X}} |C_n(x,A)| \leq C(1+A)(J_0(p,D))^{-p/d} \right) \to 1\,.$$

The result now follows from part (i), step 2 of the proof of Theorem 4.1, which shows that $\inf_{p \in [\underline{p},\overline{p}]} \inf_{h_0 \in \mathcal{H}^p} \left(J^*_n(p,E) \leq J_0(p,D)\right) \to 1$ with $J^*_n(p,E) \asymp (n/\log n)^{d/(2(p+\varsigma)+d)}$. ■

Proof of Theorem 4.3. In some of what follows, we use the fact that the sieve dimensions for CDV wavelet bases are linked via $J^+ = 2^d J$ for $J \in \mathcal{T}$. We do so for notational convenience; a similar argument (but with more complicated notation) applies for B-spline bases.

Part (ii): First note by Lemma B.5 and the fact that $\bar{J}_{\max}(R) = \bar{J}^*_{\max}(R)$ (see (22)) holds for any $R > 0$ for all n sufficiently large (see part (ii), step 1 of the proof of Theorem 4.1), we have that $J^*_{\max}(R_1) \leq \hat{J}_{\max} \leq J^*_{\max}(R_2)$ wpa1 $\mathcal{H}$-uniformly.

Recall $M_0(p, R_2)$ from (23). By Lemma B.2, for all $p \in [\underline{p}, \overline{p}]$ we have that $M_0(p, R_2) \geq M_0(\overline{p}, R_2) \geq 2^{-d} J^*_{\max}(R_2)$ holds for all n sufficiently large. Then by Lemmas B.4, B.6, and B.7 and Assumption 4(i), there exist constants $C, C' > 0$ for which

$$\sup_{x \in \mathcal{X}} |C_n(x, A)| \leq C(1+A)\tau_{\tilde{J}}\sqrt{(\tilde{J}\log(\bar{J}^*_{\max}(R_2)))/n} + A\tilde{J}^{-\underline{p}/d} \leq C'(1+A)(J^*_{\max}(R_2))^{-p/d} + A\tilde{J}^{-\underline{p}/d}$$

holds wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$, where the second inequality is by definition of $M_0(p, R_2)$. The proofs of Theorem 4.1 and Corollary 4.1 show that $\bar{J}^*_{\max}(R_2) \asymp (\log n)^{d/\varsigma}$ in the severely ill-posed case. Therefore, it suffices to show that there is a constant $c > 0$ for which $\hat{J} \geq c(\log n)^{d/\varsigma}$ holds wpa1 uniformly for $h_0 \in \mathcal{G}^p$ and $p \in [\underline{p}, \overline{p}]$.

Recall β and E from the proof of Theorem 4.2 and $J_0(p, E)$ from (21). By similar arguments to Lemma B.2, we may deduce that $\inf\{J \in \mathcal{T} : J \geq \beta J_0(p, E)\} > \log \hat{J}_{\max}$ wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$. It then follows by the same argument as part (i), step 1 of the proof of Theorem 4.2 that $\hat{J} \geq \beta J_0(p, E)$ holds wpa1 uniformly for $h_0 \in \mathcal{G}^p$ and $p \in [\underline{p}, \overline{p}]$. But by Lemma B.6 and the fact that $\log \bar{J}^*_{\max}(R_2) \asymp \log\log n$ for severely ill-posed models, it follows that there is a constant $C'' > 0$ for which, by defining

$$J^*(p, C'') = \sup\left\{J \in \mathcal{T} : \tau_J \sqrt{(J \log\log n)/n} \leq C'' J^{-p/d}\right\},$$

we have $\inf_{p \in [\underline{p}, \overline{p}]} \inf_{h_0 \in \mathcal{H}^p} \mathbb{P}_{h_0}(J_0(p, E) \geq J^*(p, C'')) \to 1$. Finally, we may deduce by a similar argument to part (ii), step 1 of the proof of Theorem 4.1 that $J^*(p, C'') \gtrsim (\log n)^{d/\varsigma}$ for all $p \in [\underline{p}, \overline{p}]$, which establishes the desired behavior of $\hat{J}$.

Part (i): By Theorem B.1 and Lemma B.5, there exists a constant $A_0 > 0$ for which

$$|\hat{h}_{\tilde{J}}(x) - h_0(x)| \leq |\hat{h}_{\tilde{J}}(x) - \tilde{h}_{\tilde{J}}(x)| + A_0 \tilde{J}^{-\underline{p}/d}$$

holds for all $x \in \mathcal{X}$ wpa1 $\mathcal{H}$-uniformly. Then for any $A \geq A_0$, we have

$$\inf_{h_0 \in \mathcal{G}} \mathbb{P}_{h_0}\big(h_0(x) \in C_n(x, A) \ \ \forall\, x \in \mathcal{X}\big) \geq \inf_{h_0 \in \mathcal{G}} \mathbb{P}_{h_0}\left(\sup_{x \in \mathcal{X}} \left|\frac{\hat{h}_{\tilde{J}}(x) - \tilde{h}_{\tilde{J}}(x)}{\|\hat{\sigma}_{x,\tilde{J}}\|_{sd}}\right| \leq z^*_{1-\alpha}\right) + o(1)\,.$$

Suppose that $J^*_{\max}(R_2) \geq 2^{2d} J^*_{\max}(R_1) \in \mathcal{T}$. Then by definition of $\bar{J}^*_{\max}(R)$ and Remark

B.1, we have

$$\frac{\tau_{J^*_{\max}(R_2)}}{\tau_{2^{2d}J^*_{\max}(R_1)}} \asymp \frac{\tau_{J^*_{\max}(R_2)}}{\tau_{2^{2d}J^*_{\max}(R_1)}} \frac{J^*_{\max}(R_2)\sqrt{\log J^*_{\max}(R_2)}}{2^{2d}J^*_{\max}(R_1)\sqrt{\log J^*_{\max}(R_1)}} \leq \frac{R_2}{R_1}\,. \tag{35}$$

But note that if $J^*_{\max}(R_2) \geq 2^{2d}J^*_{\max}(R_1)$ then by severe ill-posedness we have

$$\frac{\tau_{J^*_{\max}(R_2)}}{\tau_{2^d J^*_{\max}(R_1)}} \geq \frac{\tau_{2^{2d}J^*_{\max}(R_1)}}{\tau_{2^d J^*_{\max}(R_1)}} \asymp e^{C((2^{2d}J^*_{\max}(R_1))^{\varsigma/d}-(2^d J^*_{\max}(R_1))^{\varsigma/d})} = e^{C2^\varsigma(2^\varsigma-1)(J^*_{\max}(R_1))^{\varsigma/d}} \to +\infty\,,$$

which contradicts (35). Therefore, $\bar{J}^*_{\max}(R_1) \in \{2^{-d}\bar{J}^*_{\max}(R_2), \bar{J}^*_{\max}(R_2)\}$ holds for all n sufficiently large, from which it follows by Lemma B.5 that $\hat{J}_{\max} \in \{2^{-d}\bar{J}^*_{\max}(R_2), \bar{J}^*_{\max}(R_2)\}$ wpa1 $\mathcal{H}$-uniformly. Therefore, $\tilde{J} \leq 2^{-d}\bar{J}^*_{\max}(R_2)$ holds wpa1 $\mathcal{H}$-uniformly. But by part (ii) we also have that $\tilde{J} \geq c\bar{J}^*_{\max}(R_2)$ holds for a sufficiently small $c > 0$ wpa1 uniformly $h_0 \in \mathcal{G}^p$ and $p \in [\underline{p}, \overline{p}]$. Therefore, $\tilde{J} \in \underline{\mathcal{J}}_n := \{J \in \mathcal{T} : c\bar{J}^*_{\max}(R_2) \leq J \leq 2^{-d}\bar{J}_{\max}(R_2)\}$ and $\hat{\mathcal{J}} \supseteq \underline{\mathcal{J}}_n$ both hold wpa1 uniformly for $h_0 \in \mathcal{G}^p$ and $p \in [\underline{p}, \overline{p}]$.

Let $\underline{z}^*_{1-\alpha}$ denote the $1-\alpha$ quantile of $\sup_{(x,J)\in\mathcal{X}\times\underline{\mathcal{J}}_n} |\mathbb{Z}^*_n(x,J)|$. As $\underline{z}^*_{1-\alpha} \leq z^*_{1-\alpha}$ must hold whenever $\hat{\mathcal{J}} \supseteq \underline{\mathcal{J}}_n$, we therefore have

$$\begin{aligned}
&\inf_{h_0\in\mathcal{G}} \mathbb{P}_{h_0}\left(h_0(x) \in C_n(x,A) \;\; \forall \;\; x \in \mathcal{X}\right) \\
&\geq \inf_{p\in[\underline{p},\overline{p}]} \inf_{h_0\in\mathcal{G}^p} \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\underline{\mathcal{J}}_n} \sqrt{n}\frac{|\hat{h}_J(x) - \tilde{h}_J(x)|}{\|\hat{\sigma}_{x,J}\|_{sd}} \leq \underline{z}^*_{1-\alpha}\right) + o(1) = (1-\alpha) + o(1)\,,
\end{aligned}$$

where the last equality follows from Theorem B.2(i) and the definition of $\underline{z}^*_{1-\alpha}$. ■

References

Adão, R., C. Arkolakis, and S. Ganapati (2020). Aggregate implications of firm heterogeneity: A nonparametric analysis of monopolistic competition trade models. NBER Working Paper, w28081.

Adao, R., A. Costinot, and D. Donaldson (2017). Nonparametric counterfactual predictions in neoclassical models of international trade. *American Economic Review 107*(3), 633–689.

Andrews, D. W. (2017). Examples of L2-complete and boundedly-complete distributions. *Journal of Econometrics 199*(2), 213–220.

Babii, A. (2020). Honest confidence sets in nonparametric IV regression and other ill-posed models. *Econometric Theory 36*(4), 658–706.

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. *The Review of Economics and Statistics 99*(2), 291–304.

Breunig, C. and X. Chen (2020). Adaptive, rate-optimal testing in instrumental variables models. arxiv:2006.09587 [econ.em].

Breunig, C. and J. Johannes (2016). Adaptive estimation of functionals in nonparametric instrumental regression. *Econometric Theory 32*(3), 612–654.

Bull, A. D. (2012). Honest adaptive confidence bands and self-similar functions. *Electronic Journal of Statistics 6*, 1490–1516.

Chaney, T. (2008). Distorted gravity: The intensive and extensive margins of international trade. *American Economic Review 98*(4), 1707–1721.

Chen, X. and T. Christensen (2013). Optimal uniform convergence rates for sieve nonparametric instrumental variables regression. arxiv:1311.0412 [math.st].

Chen, X. and T. Christensen (2015a). Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variables estimation. arxiv:1508.03365v1 [stat.me].

Chen, X. and T. M. Christensen (2015b). Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions. *Journal of Econometrics 188*(2), 447–465.

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 M. Reiss (2011). On rate optimality for ill-posed inverse problems in econometrics. *Econometric Theory*, 497–521.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics 42*(5), 1787–1818.

Chernozhukov, V., W. K. Newey, and A. Santos (2015). Constrained conditional moment restriction models. arxiv:1509.06311 [math.st].

Chetverikov, D. and D. Wilhelm (2017). Nonparametric instrumental variable estimation under monotonicity. *Econometrica 85*(4), 1303–1320.

Cohen, A., I. Daubechies, and P. Vial (1993). Wavelets on the interval and fast wavelet transforms.

Compiani, G. (2020). Market counterfactuals and the specification of multi-product demand: A nonparametric approach. Working paper.

Darolles, S., Y. Fan, J.-P. Florens, and E. Renault (2011). Nonparametric instrumental regression. *Econometrica 79*(5), 1541–1565.

DeVore, R. A. and G. G. Lorentz (1993). *Constructive approximation.* Springer.

DeVore, R. A. and V. A. Popov (1988). Interpolation of Besov spaces. *Transactions of the American Mathematical Society 305*(1), 397–414.

Eaton, J., S. Kortum, and F. Kramarz (2011). An anatomy of international trade: Evidence from french firms. *Econometrica 79*(5), 1453–1498.

Fang, Z. and J. Seo (2021). A projection framework for testing shape restrictions that form convex cones. *Econometrica (forthcoming).*

Freyberger, J. and B. Reeves (2019). Inference under shape restrictions. Working paper.

Giné, E. and R. Nickl (2010). Confidence bands in density estimation. *The Annals of Statistics 38*(2), 1122–1170.

Giné, E. and R. Nickl (2016). *Mathematical foundations of infinite-dimensional statistical models.* Cambridge University Press.

Hall, P. and J. L. Horowitz (2005). Nonparametric methods for inference in the presence of instrumental variables. *The Annals of Statistics 33*(6), 2904–2929.

Head, K., T. Mayer, and M. Thoenig (2014). Welfare and trade without Pareto. *American Economic Review 104*(5), 310–316.

Horowitz, J. L. (2011). Applied nonparametric instrumental variables estimation. *Econometrica 79*(2), 347–394.

Horowitz, J. L. (2014). Adaptive nonparametric instrumental variables estimation: Empirical choice of the regularization parameter. *Journal of Econometrics 180*(2), 158–173.

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

Huang, J. Z. (2003). Local asymptotics for polynomial spline regression. *The Annals of Statistics 31*(5), 1600–1635.

Jansson, M. and D. Pouzo (2020). Towards a general large sample theory for regularized estimators. arxiv:1712.07248 [math.st].

Liu, C.-A. and J. Tao (2014). Model selection and model averaging in nonparametric instrumental variables models. Working paper, National University of Singapore and University of Wisconsin-Madison.

Low, M. G. (1997). On nonparametric confidence intervals. *The Annals of Statistics 25*(6), 2547–2554.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics 21*(1), 255–285.

Melitz, M. J. and S. J. Redding (2015). New trade models, new welfare implications. *American Economic Review 105*(3), 1105–1146.

Newey, W. K. and J. L. Powell (2003). Instrumental variable estimation of nonparametric models. *Econometrica 71*(5), 1565–1578.

Picard, D. and K. Tribouley (2000). Adaptive confidence interval for pointwise curve estimation. *The Annals of Statistics 28*(1), 298–335.

Schennach, S. M. (2020). A bias bound approach to non-parametric inference. *Review of Economic Studies 87*(5), 2439–2472.

Spokoiny, V. and N. Willrich (2019). Bootstrap tuning in Gaussian ordered model selection. *The Annals of Statistics 47*(3), 1351–1380.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics 10*(4), 1040–1053.

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

Online Appendix to "Adaptive Uniform Confidence Bands for Nonparametric IV"

Xiaohong Chen   Timothy Christensen   Sid Kankanala

C Supplemental Results and Proofs

In this Supplemental Appendix we first present additional technical lemmas and proofs of all the technical results in Appendix B. We then present lemmas and proofs of theorems in Subsection 4.4 on UCBs for derivatives of NPIV functions.

C.1 Supplemental Results: Proofs of Technical Results in Appendix B

Proof of Lemma B.1. In what follows, we drop dependence of $\bar{J}_{\max}$ on R to simplify notation. As $\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\big(\theta^*(\hat{\alpha}) \geq C_4\sqrt{\log \bar{J}_{\max}}\big) \to 1$ for some constant $C_4 > 0$ (cf. Lemma B.6), we have $J_0(p, D) \leq \tilde{J}_0$ wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$, where

$$\tilde{J}_0 := \sup\left\{J \in \mathcal{T} : \tau_J J\sqrt{(\log \bar{J}_{\max})/n} \leq \frac{D}{C_4} J^{1/2-\underline{p}/d}\right\}.$$

Note that $\tilde{J}_0$ is deterministic and independent of p. It therefore suffices to show that the inequality $\tilde{J}_0^+ < \bar{J}_{\max}$ holds for all n sufficiently large, where $\tilde{J}_0^+ := \inf\{J \in \mathcal{T} : J > \tilde{J}_0\}$. To do so we argue by contradiction. Suppose that $\tilde{J}_0^+ \geq \bar{J}_{\max}$. Then by definition of $\tilde{J}_0$, we have

$$\tau_{\bar{J}_{\max}^-} \bar{J}_{\max}^- \sqrt{(\log \bar{J}_{\max})/n} \leq \frac{D}{C_4}(\bar{J}_{\max}^-)^{1/2-\underline{p}/d},$$

where $\bar{J}_{\max}^- := \sup\{J \in \mathcal{T} : J < \bar{J}_{\max}\}$. It follows from the fact that $2^{-L}J^+ \asymp J \asymp 2^L J^-$ for some $L \in \mathbb{N}$ and $\tau_J \asymp J^{\varsigma/d}$ that whenever $\tilde{J}_0^+ \geq \bar{J}_{\max}$, we must have

$$\tau_{\bar{J}_{\max}} \bar{J}_{\max} \sqrt{(\log \bar{J}_{\max})/n} \lesssim \bar{J}_{\max}^{1/2-\underline{p}/d}. \tag{36}$$

On the other hand, by definition of $\bar{J}_{\max}$ from (24) we may similarly deduce

$$((\log n)^4 \vee \tau_{\bar{J}_{\max}})\bar{J}_{\max}\sqrt{(\log \bar{J}_{\max})/n} \asymp 1 \tag{37}$$

The result now follows by noting that for $\varsigma > 0$ we have that $\tau_{\bar{J}_{\max}}$ grows like some power of n, so displays (36) and (37) yield a contradiction for all n sufficiently large. Similarly, with

$\varsigma = 0$ displays (36) and (37) lead to $(\log n)^{-4} \lesssim \bar{J}_{\max}^{-1/2-p/d}$, which yields a contradiction for all n sufficiently large because $\bar{J}_{\max}$ grows like some power of n. ■

Proof of Lemma B.2. As every $J \in \mathcal{T}$ can be expressed as $J = 2^{dL}$ for some $L \in \mathbb{N}$, we can define $\bar{J}^*_{\max}(R)$ and $M_0(\overline{p}, R)$ through the resolution levels. We do so here for the case of CDV wavelet sieves, but our results apply equally (but with more complicated notation) for B-spline bases. Let

$$L^*_{\max}(R) = \sup\left\{L \in \mathbb{N} : \tau_{2^{Ld}} 2^{Ld}\sqrt{\log(2^{Ld})} \le R\sqrt{n}\right\},$$
$$L_0(R) = \sup\left\{L \in \mathbb{N} : \tau_{2^{Ld}} 2^{\kappa Ld}\sqrt{\log(2^{Ld})} \le R\sqrt{n}\right\},$$

where $\kappa = \overline{p}/d + 1/2$. Define the sequences

$$x_l = \left(\tau_{2^{ld}} 2^{ld}\sqrt{\log(2^{ld})}\right)^2, \quad y_l = \left(\tau_{2^{ld}} 2^{l\kappa d}\sqrt{\log(2^{ld})}\right)^2.$$

Note that $y_l/x_l = 2^{2l(\kappa-1)d} \to \infty$ as $l \to \infty$ because $\kappa > 1$. As $\tau_J \asymp \exp(CJ^{\varsigma/d})$ for some $C, \varsigma > 0$, we have

$$\frac{y_l}{x_{l+1}} \asymp \frac{e^{2C2^{l\varsigma}} 2^{2l\kappa d}\log(2^{ld})}{e^{2C2^{(l+1)\varsigma}} 2^{2(l+1)d}\log(2^{(l+1)d})} \to 0$$

as $l \to \infty$. Let L be sufficiently large that $x_l < y_l < x_{l+1}$ for all $l \ge L$. Then for any $n \in \mathbb{N}$ for which $R^2 n \ge x_L$, we have that

$$R^2 n \in [x_l, y_l) \implies L^*_{\max}(R) = l\,, \quad L_0(R) = l - 1\,,$$
$$R^2 n \in [y_l, x_{l+1}) \implies L^*_{\max}(R) = L_0(R) = l\,,$$

in which case $L^*_{\max}(R) - 1 \le L_0(R) \le L^*_{\max}(R)$. Therefore, we have shown that the inequality $M_0^+(r, R) \ge \bar{J}^*_{\max}(R)$ holds for all sufficiently large n. ■

C.1.1 Supplemental Results: Uniform-in-J Convergence Rates for $\hat{h}_J$

We first state and prove some preliminary lemmas before proving Theorem B.1.

Lemma C.1 *Let Assumptions 1(i)(ii) hold. Then: for any $\bar{J}_{\max} \to \infty$ with $(\log\log \bar{J}_{\max})^2/n \to 0$, we have*

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\left\| G_{b,J}^{-1/2}(\mathbb{E}_n - \mathbb{E})\left[b_W^{K(J)}\Delta_J h_0(X)\right]\right\|_{\ell^2} \le (1 + a_b)\|\Delta_J h_0\|_\infty \sqrt{\bar{J}_{\max}/n} \quad \forall\ J \in [1, \bar{J}_{\max}] \cap \mathcal{T}\right) \to 1.$$

Proof of Lemma C.1. If there exists $J \in \mathcal{T}$ for which $\|h_0 - \Pi_J h_0\|_\infty = 0$, then the inequality is trivially true at that J. Fix any $J \in \mathcal{T} \cap [1, \bar{J}_{\max}]$ for which $\|h_0 - \Pi_J h_0\|_\infty > 0$ and note that

$$G_{b,J}^{-1/2}(\mathbb{E}_n - \mathbb{E})\big[b_W^{K(J)} \Delta_J h_0(X)\big] = \sum_{i=1}^n e_{i,J}$$

where

$$e_{i,J} := n^{-1} G_{b,J}^{-1/2}\left(b_{W_i}^{K(J)} \Delta_J h_0(X_i) - \mathbb{E}[b_W^{K(J)} \Delta_J h_0(X)]\right).$$

The $e_{1,J}, \dots, e_{n,J}$ are IID centered random vectors of dimension $K(J)$ and $\|\sum_{i=1}^n e_{i,J}\|_{\ell^2} = \sup_{v \in \mathbb{Q}^{K(J)} : \|v\|_{\ell^2}=1} \sum_{i=1}^n v' e_{i,J}$. For any fixed $v \in \mathbb{Q}^{K(J)}$ with $\|v\|_{\ell^2} = 1$, we have that

$$\begin{aligned}
\mathbb{E}\big[(v'e_{i,J})^2\big] &\le \frac{\|\Delta_J h_0\|_\infty^2}{n^2}, \\
\big|v'e_{i,J}\big| &\le \frac{2\|\Delta_J h_0\|_\infty \zeta_{b,J}}{n}, \\
\mathbb{E}\left[\left\|\sum_{i=1}^n e_{i,J}\right\|_{\ell^2}\right] &\le \left(\mathbb{E}\left[\left\|\sum_{i=1}^n e_{i,J}\right\|_{\ell^2}^2\right]\right)^{1/2} \le \|\Delta_J h_0\|_\infty \sqrt{\frac{K(J)}{n}},
\end{aligned}$$

where the final inequality is because $\mathbb{E}[\|e_{i,J}\|^2] \le n^{-1}\|\Delta_J h_0\|_\infty \mathbb{E}[\|G_{b,J}^{-1/2} b_W^{K(J)}\|^2]$ and $L^2(W)$-orthonormality of the elements of $G_{b,J}^{-1/2} b_w^{K(J)}$. By Talagrand's Inequality (Giné and Nickl, 2016, Theorem 3.3.9), we may deduce

$$\begin{aligned}
&\mathbb{P}_{h_0}\left(\left\|\sum_{i=1}^n e_{i,J}\right\|_{\ell^2} \ge \|\Delta_J h_0\|_\infty \sqrt{\frac{K(J)}{n}} + \|\Delta_J h_0\|_\infty \sqrt{\frac{\bar{J}_{\max}}{n}}\right) \\
&\qquad \le \exp\left(-\frac{\bar{J}_{\max}}{2 + 4\zeta_{b,J}(2\sqrt{K(J)} + \sqrt{\bar{J}_{\max}}/3)/\sqrt{n}}\right).
\end{aligned}$$

As the right-hand side does not depend on h_0 and $\#\{J \in \mathcal{T} : J \le \bar{J}_{\max}\} \le \log_2(\bar{J}_{\max})/d$, a union bound yields

$$\begin{aligned}
&\sup_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\bigcup_{J \in \mathcal{T} : J \le \bar{J}_{\max}} \left\{\left\|\sum_{i=1}^n e_{i,J}\right\|_{\ell^2} \ge \|\Delta_J h_0\|_\infty \left(\frac{\sqrt{K(J)} + \sqrt{\bar{J}_{\max}}}{\sqrt{n}}\right)\right\}\right) \\
&\le \frac{1}{d\log 2} \exp\left(\log\log \bar{J}_{\max} - \frac{\bar{J}_{\max}}{2 + 4(\max_{J \in \mathcal{T} : J \le \bar{J}_{\max}} \zeta_{b,J})(2\sqrt{K(\bar{J}_{\max})} + \sqrt{\bar{J}_{\max}}/3)/\sqrt{n}}\right) \\
&\le \frac{1}{d\log 2} \exp\left(\log\log \bar{J}_{\max} - \frac{\bar{J}_{\max}}{2 + C\bar{J}_{\max}/\sqrt{n}}\right),
\end{aligned}$$

where final inequality holds for some universal constant $C > 0$ because $\zeta_{b,J} \asymp \sqrt{J}$ (cf. Lemmas A.1 and A.2) and $K(J) = a_b J$. If $\limsup \bar{J}_{\max}^2/n < \infty$ then this final term in the preceding display $\to 0$ because $\bar{J}_{\max} \to \infty$. Alternatively, if $\liminf \bar{J}_{\max}^2/n = \infty$ then for n sufficiently large the final term in the preceding display is bounded by $\frac{1}{d\log 2}\exp(\log\log \bar{J}_{\max} - \sqrt{n}/(2C))$ which vanishes provided $(\log\log \bar{J}_{\max})^2/n \to 0$. ■

Lemma C.2 *Let Assumptions 1(i)(ii) and 2(i) hold and let $\bar{J}_{\max} \to \infty$ with $(\log\log \bar{J}_{\max})^3/n \to 0$. Then: there is a universal constant $C_{C.2} > 0$ such that*

$$\inf_{h\in\mathcal{H}_0} \mathbb{P}_{h_0}\left(\sup_{J:J\le \bar{J}_{\max}, J\in\mathcal{T}} \|G_{b,J}^{-1/2}\mathbb{E}_n[b_W^{K(J)}u]\|_{\ell^2} \le C_{C.2}\sqrt{\bar{J}_{\max}/n}\right) \to 1.$$

Proof of Lemma C.2. Define $u_i^+ = u_i\mathbb{1}\{|u_i| \le n^{1/6}\}$ and $u_i^- = u_i\mathbb{1}\{|u_i| > n^{1/6}\}$. We may then write

$$G_{b,J}^{-1/2}\mathbb{E}_n[b_W^{K(J)}u] = \sum_{i=1}^n \Xi_{+,i}^J + \sum_{i=1}^n \Xi_{-,i}^J =: T_{1,J} + T_{2,J}$$

where $\Xi_{\pm,i}^J = n^{-1}G_{b,J}^{-1/2}\big(b_{W_i}^{K(J)}u_i^{\pm} - \mathbb{E}\big[b_W^{K(J)}u^{\pm}\big]\big)$.

Control of $T_{1,J}$: Note $\|\sum_{i=1}^n \Xi_{+,i}^J\|_{\ell^2} = \sup_{v\in\mathbb{Q}^{K(J)}:\|v\|_{\ell^2}=1}\sum_{i=1}^n v'\Xi_{+,i}^J$. By Assumption 2(i), let $\sigma^2 > 0$ be such that $\mathbb{E}[u^2|W] \le \sigma^2$ (almost surely). For any fixed $v \in \mathbb{Q}^{K(J)}$ with $\|v\|_{\ell^2} = 1$, we have that

$$\begin{aligned}
\mathbb{E}\big[\big(v'\Xi_{+,i}\big)^2\big] &\le n^{-1}\mathbb{E}\left[v'G_{b,J}^{-1/2}b_{W_i}^K(b_{W_i}^K)'G_{b,J}^{-1/2}v|u_{1,i}|^2\right] \le \frac{\sigma^2}{n^2}\,,\\
\big|v'\Xi_{+,i}\big| &\le 2n^{-5/6}\zeta_{b,\bar{J}_{\max}}\,,\\
\mathbb{E}\left[\left\|\sum_{i=1}^n \Xi_{+,i}^J\right\|_{\ell^2}\right] &\le \left(\mathbb{E}\left[\left\|\sum_{i=1}^n \Xi_{+,i}^J\right\|_{\ell^2}\right]\right)^{1/2} \le \sigma\sqrt{\frac{K(\bar{J}_{\max})}{n}}.
\end{aligned}$$

By Talagrand's Inequality (Giné and Nickl, 2016, Theorem 3.3.9) and $\#\{J \in \mathcal{T} : J \le \bar{J}_{\max}\} \le \log_2(\bar{J}_{\max})/d$, we may deduce by a union bound argument that for any $C > 0$,

$$\begin{aligned}
&\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\bigcup_{J\in\mathcal{T}:J\le\bar{J}_{\max}} \left\{\|T_{1,J}\|_{\ell^2} \ge \sigma\sqrt{\frac{K(\bar{J}_{\max})}{n}} + C\sqrt{\frac{\bar{J}_{\max}}{n}}\right\}\right)\\
&\le \sum_{J\in\mathcal{T}:J\le\bar{J}_{\max}} \exp\left(-\frac{C^2\bar{J}_{\max}}{2\sigma^2 + 4n^{-1/3}\zeta_{b,\bar{J}_{\max}}(2\sigma\sqrt{K(\bar{J}_{\max})} + C\sqrt{\bar{J}_{\max}}/3)}\right)\\
&\le \frac{1}{d\log 2}\exp\left(\log\log \bar{J}_{\max} - \frac{C^2\bar{J}_{\max}}{2\sigma^2 + 4C'n^{-1/3}\bar{J}_{\max}}\right)
\end{aligned}$$

holds for n sufficiently large, where the final line holds for some $C' > 0$ because $\zeta_{b,J} \asymp \sqrt{J}$ (cf. Lemmas A.1 and A.2) and $K(\bar{J}_{\max}) \asymp \bar{J}_{\max}$. If $\limsup \bar{J}_{\max}^3/n < \infty$ then the final term in the preceding display $\to 0$ because $\bar{J}_{\max} \to \infty$. Alternatively, if $\liminf \bar{J}_{\max}^3/n = \infty$ then for n sufficiently large the final term in the preceding display is bounded by $\frac{1}{d\log 2}\exp(\log\log \bar{J}_{\max} -$

$n^{1/3}C'')$ for some $C'' > 0$ and therefore vanishes because $(\log\log \bar{J}_{\max})^3/n \to 0$.

Control of $T_{2,J}$: As $\zeta_{b,J} \asymp \sqrt{J}$ for all $J \in \mathcal{T}$ (cf. Lemmas A.1 and A.2), we may deduce that $\sup_{J\in\mathcal{T}:J\le\bar{J}_{\max}} J^{-1/2}\|\Xi^J_{-,i}\|_{\ell^2} \lesssim n^{-1}|u_i|\mathbb{1}\{|u_i| > n^{1/6}\}$. As such, for any $C''' > 0$, we have

$$\sup_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J\in\mathcal{T},J\le\bar{J}_{\max}} J^{-1/2}\|T_{2,J}\|_{\ell^2} > \frac{C'''}{\sqrt{n}}\right) \lesssim (C''')^{-1}\mathbb{E}\big[|u|^4\mathbb{1}\{|u| > n^{1/6}\}\big] \to 0$$

by Assumption 2(i). ■

Lemma C.3 *Let Assumption 1(i)(ii) hold and let $\bar{J}_{\max} \to \infty$ with $(\bar{J}_{\max}\log\bar{J}_{\max})/n \to 0$. Then: there is a universal constant $C_{C.3} > 0$ such that*

$$\sup_{J\in\mathcal{T}:J\le\bar{J}_{\max}} \max\left\{\|\widehat{G}^o_{b,J} - I_{K(J)}\|_{\ell^2}, \|\widehat{G}^o_{\psi,J} - I_J\|_{\ell^2}, \|\widehat{S}^o_J - S^o_J\|_{\ell^2}\right\} \le C_{C.3}\sqrt{(\bar{J}_{\max}\log\bar{J}_{\max})/n}$$

wpa1 $\mathcal{H}$-uniformly. Moreover, there exists universal constants $0 < \underline{c} < \bar{c} < \infty$ which satisfy:

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\big(\underline{c}\lambda_{\min}(G_{\psi,J}) \le \lambda_{\min}(\widehat{G}_{\psi,J}) \le \bar{c}\lambda_{\min}(G_{\psi,J}) \quad \forall\ J\ \in [1,\bar{J}_{\max}]\ \cap\mathcal{T}\big) \to 1,$$
$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\big(\underline{c}\lambda_{\max}(G_{\psi,J}) \le \lambda_{\max}(\widehat{G}_{\psi,J}) \le \bar{c}\lambda_{\max}(G_{\psi,J}) \quad \forall\ J\ \in [1,\bar{J}_{\max}]\ \cap\mathcal{T}\big) \to 1,$$
$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\big(\underline{c}\lambda_{\min}(G_{b,J}) \le \lambda_{\min}(\widehat{G}_{b,J}) \le \bar{c}\lambda_{\min}(G_{b,J}) \quad \forall\ J\ \in [1,\bar{J}_{\max}]\ \cap\mathcal{T}\big) \to 1,$$
$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\big(\underline{c}\lambda_{\max}(G_{b,J}) \le \lambda_{\max}(\widehat{G}_{b,J}) \le \bar{c}\lambda_{\max}(G_{b,J}) \quad \forall\ J\ \in [1,\bar{J}_{\max}]\ \cap\mathcal{T}\big) \to 1.$$

Proof of Lemma C.3. The first result on the convergence rate of the three matrix estimators may be deduced from the exponential inequalities in Lemma F.7 of Chen and Christensen (2018) using a union bound argument, Lemmas A.1 and A.2, and $K(J) = a_bJ$ for all $J \in \mathcal{T}$.

For the final claim, by Weyl's inequality we have

$$\begin{aligned}\left|\lambda_{\min}(\widehat{G}_{\psi,J}) - \lambda_{\min}(G_{\psi,J})\right|, \left|\lambda_{\max}(\widehat{G}_{\psi,J}) - \lambda_{\max}(G_{\psi,J})\right| &\le \|\widehat{G}_{\psi,J} - G_{\psi,J}\|_{\ell^2}\\ &\le \|G_{\psi,J}\|_{\ell^2}\|\widehat{G}^o_{b,J} - I_{K(J)}\|_{\ell^2}\\ &= \lambda_{\max}(G_{\psi,J})\|\widehat{G}^o_{b,J} - I_{K(J)}\|_{\ell^2}\end{aligned}$$

and so the desired result for $\lambda_{\max}(\widehat{G}_{\psi,J})$ from the first claim because $\sqrt{(\bar{J}_{\max}\log\bar{J}_{\max})/n} \to 0$. The result for $\lambda_{\min}(\widehat{G}_{\psi,J})$ follows similarly, using $\lambda_{\max}(G_{\psi,J})/\lambda_{\min}(G_{\psi,J}) \lesssim 1$ (cf. Lemmas A.1 and A.2). The result for $G_{b,J}$ follows similarly, using $K(J) = a_bJ$ for all $J \in \mathcal{T}$. ■

Lemma C.4 *Let Assumptions 1 and 3(i) hold and let $\bar{J}_{\max} \to \infty$ with $\tau_{\bar{J}_{\max}}\sqrt{(\bar{J}_{\max}\log\bar{J}_{\max})/n} \to$*

0. *Then: there is a universal constant $C_{C.4} > 0$ such that:*

$$(i)\ \inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\bigg(\sup_{J\in\mathcal{T}\cap[1,\bar{J}_{\max}]} \tau_J^{-2} \big\| ((\widehat{G}^o_{b,J})^{-1/2}\widehat{S}^o_J)^-_l (\widehat{G}^o_{b,J})^{-1/2} - (S^o_J)^-_l \|_{\ell^2} \leq C_{C.4}\sqrt{(\bar{J}_{\max}\log \bar{J}_{\max})/n}\bigg) \to 1\,;$$

$$(ii)\ \inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\bigg(\sup_{J\in\mathcal{T}\cap[1,\bar{J}_{\max}]} \tau_J^{-1} \big\| S^o_J\{ ((\widehat{G}^o_{b,J})^{-1/2}\widehat{S}^o_J)^-_l (\widehat{G}^o_{b,J})^{-1/2} - (S^o_J)^-_l \}\big\|_{\ell^2} \leq C_{C.4}\sqrt{(\bar{J}_{\max}\log \bar{J}_{\max})/n}\bigg) \to 1\,.$$

Proof of Lemma C.4. Part (i): First note that

$$\begin{aligned}&\|((\widehat{G}^o_{b,J})^{-1/2}\widehat{S}^o_J)^-_l (\widehat{G}^o_{b,J})^{-1/2} - (\widehat{S}^o_J)^-_l\|_{\ell^2}\\ &\leq \|((\widehat{G}^o_{b,J})^{-1/2}\widehat{S}^o_J)^-_l - (S^o_J)^-_l\|_{\ell^2}\|(\widehat{G}^o_{b,J})^{-1/2}\|_{\ell^2} + \|(\widehat{G}^o_{b,J})^{-1/2} - I\|_{\ell^2}\|(S^o_J)^-_l\|_{\ell^2} =: T_{1,J} + T_{2,J}.\end{aligned}$$

Control of $T_{2,J}$: By Lemma C.3, we have $\max_{J\in\mathcal{T}:J\leq\bar{J}_{\max}} \|(\widehat{G}^o_{b,J})^{-1/2}\|_{\ell^2} \leq 2$ wpa1 $\mathcal{H}$-uniformly. It then follows by Lemmas F.2 and F.3 of Chen and Christensen (2018) that

$$\|(\widehat{G}^o_{b,J})^{-1/2} - I\|_{\ell^2} \leq \frac{\|(\widehat{G}^o_{b,J})^{-1/2}\|_{\ell^2}}{\sqrt{\lambda_{\min}(\widehat{G}^o_{b,J})}+1}\|\widehat{G}^o_{b,J} - I\|_{\ell^2} \leq \frac{2}{3}\|\widehat{G}^o_{b,J} - I\|_{\ell^2} \quad \forall\ J\in[1,\bar{J}_{\max}]\cap\mathcal{T} \tag{38}$$

wpa1 $\mathcal{H}$-uniformly. Finally, as $\|(S^o_J)^-_l\|_{\ell^2} = s_J^{-1} \leq a_\tau \tau_J$ for all $J\in\mathcal{T}$ (cf. (20)), it follows by Lemma C.3 that

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\bigg(\sup_{J\in\mathcal{T}\cap[1,\bar{J}_{\max}]} \tau_J^{-1}\|(\widehat{G}^o_{b,J})^{-1/2} - I\|_{\ell^2}\|(S^o_J)^-_l\|_{\ell^2} \leq \frac{2}{3}a_\tau C_{C.3}\sqrt{(\bar{J}_{\max}\log\bar{J}_{\max})/n}\bigg) \to 1.$$

Control of $T_{1,J}$: It remains to bound $\|((\widehat{G}^o_{b,J})^{-1/2}\widehat{S}^o_J)^-_l - (S^o_J)^-_l\|_{\ell^2}$. Note

$$\|((\widehat{G}^o_{b,J})^{-1/2}\widehat{S}^o_J - S^o_J\|_{\ell^2} \leq \|(\widehat{G}^o_{b,J})^{-1/2} - I\|_{\ell^2}\|\widehat{S}^o_J\|_{\ell^2} + \|\widehat{S}^o_J - S^o_J\|_{\ell^2}\,,$$

and $\|S^o_J\|_{\ell^2} \leq 1$ for all $J\in\mathcal{T}$. It follows by Lemma C.3, that $\max_{J\in\mathcal{T}:J\leq\bar{J}_{\max}}\|\widehat{S}^o_J\|_{\ell^2} \leq 2$ wpa1 $\mathcal{H}$-uniformly. Using (38) and applying Lemma C.3, we may deduce

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\bigg(\sup_{J\in\mathcal{T}:J\leq\bar{J}_{\max}} \|((\widehat{G}^o_{b,J})^{-1/2}\widehat{S}^o_J - S^o_J\|_{\ell^2} \leq \frac{7}{3}C_{C.3}\sqrt{\frac{\bar{J}_{\max}\log\bar{J}_{\max}}{n}}\bigg) \to 1. \tag{39}$$

As $\tau_{\bar{J}_{\max}}\sqrt{(\bar{J}_{\max}\log\bar{J}_{\max})/n} \to 0$, $s_J^{-1} \leq a_\tau\tau_J$ holds for all $J\in\mathcal{T}$, and $J\mapsto\tau_J$ is monotone, we have

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J \in \mathcal{T} \cap [1, \bar{J}_{\max}]} s_J^{-1} \|((\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J - S^o_J\|_{\ell^2} \leq \frac{1}{2} \right) \to 1.$$

From this observation and Lemma F.4 of Chen and Christensen (2018), we can deduce that

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J \in \mathcal{T} \cap [1, \bar{J}_{\max}]} \tau_J^{-2} \|((\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J)^-_l - (S^o_J)^-_l\|_{\ell^2} \leq C'' \sqrt{(\bar{J}_{\max} \log \bar{J}_{\max})/n} \right) \to 1 \quad (40)$$

with $C'' = \frac{14(1+\sqrt{5})}{3} a_\tau^2 C_{C.3}$.

Part (ii): We start by observing that

$$\begin{aligned}
&\left\| S^o_J \{ ((\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J)^-_l (\widehat{G}^o_{b,J})^{-1/2} - (S^o_J)^-_l \} \right\|_{\ell^2} \\
&= \left\| S^o_J [\widehat{S}^{o\prime}_J (\widehat{G}^o_{b,J})^{-1} \widehat{S}^o_J]^{-1} \widehat{S}^{o\prime}_J (\widehat{G}^o_{b,J})^{-1} - S^o_J [S^{o\prime}_J S^o_J]^{-1} S^{o\prime}_J \right\|_{\ell^2} \\
&\leq \left\| S^o_J [\widehat{S}^{o\prime}_J (\widehat{G}^o_{b,J})^{-1} \widehat{S}^o_J]^{-1} \widehat{S}^{o\prime}_J (\widehat{G}^o_{b,J})^{-1/2} ((\widehat{G}^o_{b,J})^{-1/2} - I) \right\|_{\ell^2} \\
&\quad + \left\| (S^o_J - (\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J) [\widehat{S}^{o\prime}_J (\widehat{G}^o_{b,J})^{-1} \widehat{S}^o_J]^{-1} \widehat{S}^{o\prime}_J (\widehat{G}^o_{b,J})^{-1/2} \right\|_{\ell^2} \\
&\quad + \left\| (\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J [\widehat{S}^{o\prime}_J (\widehat{G}^o_{b,J})^{-1} \widehat{S}^o_J]^{-1} \widehat{S}^{o\prime}_J (\widehat{G}^o_{b,J})^{-1/2} - S^o_J [S^{o\prime}_J S^o_J]^{-1} S^{o\prime}_J \right\|_{\ell^2} =: T_{3,J} + T_{4,J} + T_{5,J}.
\end{aligned}$$

Control of $T_{3,J}$: By the proof of part (i), we have

$$\begin{aligned}
T_{3,J} &\leq \left\| S^o_J \right\|_{\ell^2} \left\| ((\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J)^-_l \right\|_{\ell^2} \left\| (\widehat{G}^o_{b,J})^{-1/2} - I \right\|_{\ell^2} \\
&\leq \|((\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J)^-_l\|_{\ell^2} \times \frac{2}{3} C_{C.3} \sqrt{(\bar{J}_{\max} \log \bar{J}_{\max})/n} \quad \forall \ J \in \mathcal{T} \cap [1, \bar{J}_{\max}]
\end{aligned}$$

wpa1 $\mathcal{H}$-uniformly. For the remaining term, by the reverse triangle inequality we have

$$\|((\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J)^-_l\|_{\ell^2} \leq s_J^{-1} + \|((\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J)^-_l - (S^o_J)^-_l\|_{\ell^2}.$$

It follows by similar arguments to those used to deduce (40) that

$$\|((\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J)^-_l\|_{\ell^2} \leq a_\tau \tau_J \left(1 + C''' \tau_{\bar{J}_{\max}} \sqrt{(\bar{J}_{\max} \log \bar{J}_{\max})/n} \right) \quad \forall \ J \in \mathcal{T} \cap [1, \bar{J}_{\max}]$$

wpa1 $\mathcal{H}$-uniformly with $C''' = \frac{14(1+\sqrt{5})}{3} a_\tau C_{C.3}$. The conditions on $\bar{J}_{\max}$ then imply that

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J \in \mathcal{T} \cap [1, \bar{J}_{\max}]} \tau_J^{-1} \|((\widehat{G}^o_{b,J})^{-1/2} \widehat{S}^o_J)^-_l\|_{\ell^2} \leq 2 a_\tau \right) \to 1 \quad (41)$$

and hence

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J \in \mathcal{T} \cap [1, \bar{J}_{\max}]} \tau_J^{-1} T_{3,J} \leq \frac{4}{3} a_\tau C_{C.3} \sqrt{(\bar{J}_{\max} \log \bar{J}_{\max})/n} \right) \to 1.$$

Control of $T_{4,J}$: Use $T_{4,J} \le \|S_J^o - (\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o\|_{\ell^2}\|((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-\|_{\ell^2}$ with (39) and (41).

Control of $T_{5,J}$: The norm is the difference between two projection matrices corresponding to the columns of $(\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o$ and S_J^o, respectively. The desired result may be deduced from Lemma F.6 of Chen and Christensen (2018) using (39) and (41). ∎

Proof of Theorem B.1. Part (i): As $\|\tilde h_J - h_0\|_\infty \le \|\tilde h_J - \Pi_J h_0\|_\infty + \|\Delta_J h_0\|_\infty$, it suffices to show

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\Big(\|\tilde h_J - \Pi_J h_0\|_\infty \le D\|\Delta_J h_0\|_\infty \quad \forall \;\; J \in \mathcal{T}\cap[1,\bar J_{\max}]\Big) \to 1$$

where $D > 0$ is a universal constant. We write the expression as

$$\begin{aligned}
\tilde h_J(x) - \Pi_J h_0(x) &= Q_J(\Delta_J h_0)(x) \\
&\quad + (\psi_x^J)'(G_{b,J}^{-1/2}S_J)_l^- G_{b,J}^{-1/2}(\mathbb{E}_n - \mathbb{E})\big[b_W^{K(J)}\Delta_J h_0(X)\big] \\
&\quad + (\psi_x^J)'\big\{(\widehat{G}_{b,J}^{-1/2}\widehat{S}_J)_l^- \widehat{G}_{b,J}^{-1/2}G_{b,J}^{1/2} - (G_{b,J}^{-1/2}S_J)_l^-\big\}G_{b,J}^{-1/2}\mathbb{E}_n\big[b_W^{K(J)}\Delta_J h_0(X)\big] \\
&=: T_{1,J}(x) + T_{2,J}(x) + T_{3,J}(x)\,.
\end{aligned}$$

Control of $\|T_{1,J}\|_\infty$: Follows immediately from Assumption 3(iii).

Control of $\|T_{2,J}\|_\infty$: By Lemma C.1, we have

$$\|G_{b,J}^{-1/2}(\mathbb{E}_n - \mathbb{E})\big[b_W^{K(J)}\Delta_J h_0(X)\big]\|_{\ell^2} \le (1+a_b)\|\Delta_J h_0\|_\infty\sqrt{\bar J_{\max}/n} \quad \forall\; J \in [1,\bar J_{\max}] \,\cap\, \mathcal{T}$$

wpa1 $\mathcal{H}$-uniformly. For the remaining term, we note that

$$\|(\psi_x^J)'(G_{b,J}^{-1/2}S_J)_l^-\|_{\ell^2} \le \|G_{\psi,J}^{-1/2}\psi_x^J\|_{\ell^2}s_J^{-1} \le \zeta_{\psi,J}s_J^{-1} \quad \forall\; J\in\mathcal{T}\cap[1,\bar J_{\max}]\,.$$

Combining the preceding two displays and noting that $\max_{J\in\mathcal{T}\cap[1,\bar J_{\max}]}\zeta_{\psi,J} \asymp \sqrt{\bar J_{\max}}$ (cf. Lemmas A.1 and A.2), we obtain for some finite positive constant C'

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\Big(\|T_{2,J}\|_\infty \le C' s_J^{-1}\frac{\bar J_{\max}}{\sqrt{n}}\|\Delta_J h_0\|_\infty \quad \forall \;\; J\in\mathcal{T}\cap[1,\bar J_{\max}]\Big) \to 1.$$

As $s_J^{-1} \le a_\tau \tau_J$ (cf. 20) and $J \mapsto \tau_J$ is monotone, the last display above holds with $a_\tau C'$ in place of C' and $\tau_{\bar J_{\max}}$ in place of s_J^{-1}. It then follows from $\tau_{\bar J_{\max}}\bar J_{\max}/\sqrt{n} \lesssim (\log \bar J_{\max})^{-1/2} \to 0$ that

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\Big(\|T_{2,J}\|_\infty \le \|\Delta_J h_0\|_\infty \quad \forall \;\; J\in\mathcal{T}\cap[1,\bar J_{\max}]\Big) \to 1.$$

Control of $\|T_{3,J}\|_\infty$: Note that

$$\|T_{3,J}\|_\infty \le \zeta_{\psi,J}\big\|((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J)_l^-(\widehat{G}_{b,J}^o)^{-1/2} - (S_J^o)_l^-\big\|_{\ell^2}\big\|G_{b,J}^{-1/2}\mathbb{E}_n\big[b_W^{K(J)}\Delta_J h_0(X)\big]\big\|_{\ell^2}.$$

As the functions that make up the vector $G_{b,J}^{-1/2} b_w^{K(J)}$ are $L^2(W)$-orthonormal, the Bessel inequality implies

$$\|G_{b,J}^{-1/2}\mathbb{E}[b_W^{K(J)}\Delta_J h_0(X)]\|_{\ell^2} = \|G_{b,J}^{-1/2}\mathbb{E}[b_W^{K(J)}\mathbb{E}[\Delta_J h_0(X)|W]]\|_{\ell^2} \leq \|T\Delta_J h_0\|_{L^2(W)}.$$

Using the reverse triangle inequality, we may deduce from the preceding bound, Lemma C.1 and Assumption 3(ii) that

$$\left\|G_{b,J}^{-1/2}\mathbb{E}_n\left[b_W^{K(J)}\Delta_J h_0(X)\right]\right\|_{\ell^2} \leq \left((1+a_b)\sqrt{\bar{J}_{\max}/n} + \tau_J^{-1}\right)\|\Delta_J h_0\|_\infty \quad \forall\ J \in \mathcal{T}\cap[1,\bar{J}_{\max}]$$

holds wpa1 $\mathcal{H}$-uniformly. Moreover, by Lemma C.4(i),

$$\left\|((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J)_l^-(\widehat{G}_{b,J}^o)^{-1/2} - (S_J^o)_l^-\right\|_{\ell^2} \leq C_{C.4}\tau_J^2\sqrt{(\bar{J}_{\max}\log \bar{J}_{\max})/n} \quad \forall\ J \in \mathcal{T}\cap[1,\bar{J}_{\max}]$$

wpa1 $\mathcal{H}$-uniformly. Combining these bounds and noting that $\max_{J\in\mathcal{T},J\leq\bar{J}_{\max}}\zeta_{\psi,J} \asymp \sqrt{\bar{J}_{\max}}$ (cf. Lemmas A.1 and A.2), there is a finite positive constant C such that

$$\|T_{3,J}\|_\infty \leq C\tau_J^2\left((1+a_b)\sqrt{\bar{J}_{\max}/n} + \tau_J^{-1}\right)\bar{J}_{\max}\sqrt{(\log\bar{J}_{\max})/n}\|\Delta_J h_0\|_\infty \quad \forall\ J \in \mathcal{T}\cap[1,\bar{J}_{\max}]$$

wpa1 $\mathcal{H}$-uniformly. As $\tau_{\bar{J}_{\max}}\bar{J}_{\max}\sqrt{(\log\bar{J}_{\max})/n} \leq R$, we obtain

$$\|T_{3,J}\|_\infty \leq C(1+R)\|\Delta_J h_0\|_\infty \qquad \forall\ J \in \mathcal{T}\cap[1,\bar{J}_{\max}]$$

wpa1 $\mathcal{H}$-uniformly.

Part (ii): First note that

$$\|\hat{h}_J - \tilde{h}_J\|_\infty = \sup_{x\in\mathcal{X}}|(\psi_x^J)'(\hat{c}_J - \tilde{c}_J)| \leq \sup_{x\in\mathcal{X}}\|\psi_x^J\|_{\ell^1}\|(\hat{c}_J - \tilde{c}_J)\|_{\ell^\infty}.$$

As $\zeta_{\psi,J} = \sup_{x\in\mathcal{X}}\|\psi_x^J\|_{\ell^1}\|G_{\psi,J}^{-1/2}\|_{\ell^2} \lesssim \sqrt{J}$ (cf. Lemmas A.1 and A.2) and $\tau_J \leq s_J^{-1} \leq a_\tau\tau_J$ for all $J \in \mathcal{T}$ (cf. (20)), it is enough to show that there exists a finite constant $C > 0$ which satisfies

$$\inf_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\|\hat{c}_J - \tilde{c}_J\|_\infty \leq C\|G_{\psi,J}^{-1/2}\|_{\ell^2}s_J^{-1}\sqrt{(\log\bar{J}_{\max})/n}\ \ \forall\ J \in \mathcal{T}\cap[1,\bar{J}_{\max}]\right) \to 1.$$

Given a sequence $M_n \to \infty$, let $u_i^+ = u_i\mathbb{1}\{|u_i| \leq M_n\} - \mathbb{E}[u_i\mathbb{1}\{|u_i| \leq M_n\}]$ and $u_i^- = u_i - u_i^+$.

We may then write

$$
\begin{aligned}
\|\hat{c}_J - \tilde{c}_J\|_\infty \leq\ & \big\|(G_{b,J}^{-1/2}S_J)_l^- G_{b,J}^{-1/2}\mathbb{E}_n[b_W^{K(J)}u^+]\big\|_{\ell^\infty} \\
& + \big\|(G_{b,J}^{-1/2}S_J)_l^- G_{b,J}^{-1/2}\mathbb{E}_n[b_W^{K(J)}u^-]\big\|_{\ell^\infty} \\
& + \big\|\{(\widehat{G}_{b,J}^{-1/2}\widehat{S}_J)_l^- \widehat{G}_{b,J}^{-1/2} G_{b,J}^{1/2} - (G_{b,J}^{-1/2}S_J)_l^-\} G_{b,J}^{-1/2}\mathbb{E}_n[b_W^{K(J)}u]\big\|_{\ell^\infty} \\
=:\ & T_{4,J} + T_{5,J} + T_{6,J}\,.
\end{aligned}
$$

Control of $T_{4,J}$: We have

$$
T_{4,J} = \max_{1\leq m\leq J}\left|\frac{1}{n}\sum_{i=1}^n q_{m,J}(W_i)u_i^+\right|, \qquad q_{m,J}(W_i) = ((G_{b,J}^{-1/2}S_J)_l^- G_{b,J}^{-1/2} b_{W_i}^{K(J)})_m\,,
$$

where $(v)_m$ denotes the mth element of a vector v. By Lemma F.5 of Chen and Christensen (2018) and the Cauchy–Schwarz inequality, we obtain

$$
\frac{1}{n}|q_{m,J}(W_i)u_i^+| \leq \frac{2M_n}{n}\|G_{\psi,J}^{-1/2}\|_{\ell^2}\|(S_J^o)_l^-\|_{\ell^2}\|G_{b,J}^{-1/2}b_{W_i}^{K(J)}\|_{\ell^2} \leq \frac{2M_n\|G_{\psi,J}^{-1/2}\|_{\ell^2}\zeta_{b,J}}{ns_J}. \tag{42}
$$

Let $(A)_{mm}$ denote the m^{th} diagonal element of a square matrix A. By virtue of Assumption 2, there exists $\sigma^2 > 0$ which satisfies $\mathbb{E}[(u_i^+)^2|W_i] \leq \sigma^2$ (almost everywhere). Therefore,

$$
\begin{aligned}
\mathbb{E}\left[\left(\frac{1}{n}q_{m,J}(W_i)u_{1,i}\right)^2\right] &\leq \frac{\sigma^2}{n^2}\big((G_{b,J}^{-1/2}S_J)_l^-((G_{b,J}^{-1/2}S_J)_l^-)'\big)_{mm} \\
&\leq \frac{\sigma^2}{n^2}\|(G_{b,J}^{-1/2}S_J)_l^-\|_{\ell^2}^2 = \frac{\sigma^2}{n^2}\|G_{\psi,J}^{-1/2}(S_J^o)_l^-\|_{\ell^2}^2 \leq \frac{\sigma^2}{n^2}\|G_{\psi,J}^{-1/2}\|_{\ell^2}^2 s_J^{-2}.
\end{aligned}
$$

By Bernstein's inequality (Giné and Nickl, 2016, Theorem 3.1.7) and the union bound, for any constant $C > 2\sigma$ we have that

$$
\begin{aligned}
&\inf_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\bigcup_{J\in\mathcal{T}\cap[1,\bar{J}_{\max}]}\left\{T_{4,J} > C\|G_{\psi,J}^{-1/2}\|_{\ell^2}s_J^{-1}\sqrt{(\log\bar{J}_{\max})/n}\right\}\right) \\
&\qquad\leq \sum_{J\in\mathcal{T}\cap[1,\bar{J}_{\max}]} 2J\exp\left(-\frac{C^2\log\bar{J}_{\max}}{2\sigma^2 + 8CM_n\zeta_{b,J}\sqrt{(\log\bar{J}_{\max})/(9n)}}\right) \to 0
\end{aligned}
$$

provided

$$
\sup_{J\in\mathcal{T}:J\leq\bar{J}_{\max}} M_n\zeta_{b,J}\sqrt{(\log\bar{J}_{\max})/n} \to 0\,.
$$

In view of Lemmas A.1 and A.2 and the fact that $K(J) = a_bJ$ for all $J \in \mathcal{T}$, any M_n that satisfies $M_n\sqrt{\bar{J}_{\max}(\log\bar{J}_{\max})/n} \to 0$ suffices. In the control of $T_{5,J}$ below, we shall provide an explicit form for M_n that satisfies this latter sufficient condition.

Control of $T_{5,J}$: Similar to (42), we have

$$|T_{5,J}| \leq \frac{1}{n}\sum_{i=1}^n |q_{m,J}(W_i)u_i^-| \leq \|G_{\psi,J}^{-1/2}\|_{\ell^2} s_J^{-1} \zeta_{b,J} \frac{1}{n}\sum_{i=1}^n |u_i^-|.$$

As $\sup_{J \in \mathcal{T}: J \leq \bar{J}_{\max}} \zeta_{b,K(J)} \asymp \sqrt{\bar{J}_{\max}}$ (cf. Lemmas A.1 and A.2), it follows that

$$\sup_{J \in \mathcal{T} \cap [1,\bar{J}_{\max}]} \frac{|T_{5,J}|}{\|G_{\psi,J}^{-1/2}\|_{\ell^2} s_J^{-1} \sqrt{(\log \bar{J}_{\max})/n}} \lesssim \left(\frac{1}{n}\sum_{i=1}^n |u_i^-|\right) \times \sqrt{\frac{n\bar{J}_{\max}}{\log \bar{J}_{\max}}},$$

and so, by virtue of the fact that $u_i^- = u_i \mathbb{1}\{|u_i| > M_n\} - \mathbb{E}[u_i \mathbb{1}\{|u_i| > M_n\}]$, we obtain

$$\sup_{h_0 \in \mathcal{H}} \mathbb{E}_{h_0}\left[\sup_{J \in \mathcal{T} \cap [1,\bar{J}_{\max}]} \frac{|T_{5,J}|}{\|G_{\psi,J}^{-1/2}\|_{\ell^2} s_J^{-1} \sqrt{(\log \bar{J}_{\max})/n}}\right] \lesssim \frac{\mathbb{E}[|u|^4 \mathbb{1}\{|u| > M_n\}]}{M_n^3} \sqrt{\frac{n\bar{J}_{\max}}{\log \bar{J}_{\max}}}.$$

As $\mathbb{E}[|u|^4] < \infty$ by Assumption 2(i)), setting $M_n^3 = \sqrt{n\bar{J}_{\max}/(\log \bar{J}_{\max})}$ ensures the r.h.s. of the preceding display goes to 0 asymptotically. It follows by Markov's inequality that

$$|T_{5,J}| \leq \|G_{\psi,J}^{-1/2}\|_{\ell^2} s_J^{-1} \sqrt{(\log \bar{J}_{\max})/n} \quad \forall \ J \in \mathcal{T} \cap [1, \bar{J}_{\max}]$$

wpa1 $\mathcal{H}$-uniformly. Also note that $M_n \sqrt{\bar{J}_{\max}(\log \bar{J}_{\max})/n} = (\bar{J}_{\max}^2 (\log \bar{J}_{\max})/n)^{1/3} \to 0$, as required for control of $T_{4,J}$.

Control of $T_{6,J}$: As $\|\cdot\|_\infty \leq \|\cdot\|_{\ell^2}$ on $\mathbb{R}^n$, we have

$$\begin{aligned}
&\sup_{J \in \mathcal{T} \cap [1,\bar{J}_{\max}]} \frac{T_{6,J}}{\|G_{\psi,J}^{-1/2}\|_{\ell^2} s_J^{-1} \sqrt{(\log \bar{J}_{\max})/n}} \\
&\leq \sup_{J \in \mathcal{T} \cap [1,\bar{J}_{\max}]} \frac{\|((\widehat{G}_{b,J}^o)^{-1/2} \widehat{S}_J^o)_l^- (\widehat{G}_{b,J}^o)^{-1/2} - (S_J^o)_l^-\|_{\ell^2}}{s_J^{-1} \sqrt{(\log \bar{J}_{\max})/n}} \times \sup_{J \in \mathcal{T}: J \leq \bar{J}_{\max}} \|G_{b,J}^{-1/2} \mathbb{E}_n[b_W^{K(J)} u]\|_{\ell^2}.
\end{aligned}$$

The first term in the r.h.s. product is bounded by $C_{C.4} \tau_{\bar{J}_{\max}} \sqrt{\bar{J}_{\max}}$ wpa1 $\mathcal{H}$-uniformly by Lemma C.4(i) and the fact that $s_J \leq \tau_J^{-1}$. The second term in the product is bounded $C_{C.2} \sqrt{\bar{J}_{\max}/n}$ wpa1 $\mathcal{H}$-uniformly by Lemma C.2. The result now follows because $\bar{J}_{\max}$ by definition satisfies $\tau_{\bar{J}_{\max}} \bar{J}_{\max} \sqrt{(\log \bar{J}_{\max})/n} \leq R$. ■

C.1.2 Supplemental Results: Uniform-in-J Estimation of Sieve Variance Terms

We first state and prove some preliminary lemmas before proving Lemma B.3.

Lemma C.5 *Let Assumptions 1, 2(i) and 3 hold. Then: there is a universal constant $C_{C.5}$ for* *which*

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(J,J_2)\in\mathcal{J}_n} \|\widehat{\Omega}^o_{J,J_2} - \Omega^o_{J,J_2}\|_{\ell^2} \le C_{C.5}\delta_n \right) \to 1\,.$$

Proof of Lemma C.5. First write

$$\hat{u}_{i,J}\hat{u}_{i,J_2} = u_i^2 + (\hat{u}_{i,J} - u_i)(\hat{u}_{i,J_2} - u_i) + u_i(\hat{u}_{i,J_2} - u_i) + u_i(\hat{u}_{i,J} - u_i)\,.$$

As such, we have the bound

$$\begin{aligned}
\|\widehat{\Omega}^o_{J,J_2} - \Omega^o_{J,J_2}\|_{\ell^2} \le & \left\| G_{b,J}^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n u_i^2 b_{W_i}^{K(J)}(b_{W_i}^{K(J_2)})'\right) G_{b,J_2}^{-1/2} - \Omega^o_{J,J_2}\right\|_{\ell^2} \\
&+ \left\| G_{b,J}^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n u_i(\hat{u}_{i,J} - u_i) b_{W_i}^{K(J)}(b_{W_i}^{K(J_2)})'\right) G_{b,J_2}^{-1/2}\right\|_{\ell^2} \\
&+ \left\| G_{b,J}^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n u_i(\hat{u}_{i,J_2} - u_i) b_{W_i}^{K(J)}(b_{W_i}^{K(J_2)})'\right) G_{b,J_2}^{-1/2}\right\|_{\ell^2} \\
&+ \left\| G_{b,J}^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n (\hat{u}_{i,J} - u_i)(\hat{u}_{i,J_2} - u_i) b_{W_i}^{K(J)}(b_{W_i}^{K(J_2)})'\right) G_{b,J_2}^{-1/2}\right\|_{\ell^2} \\
=:& \, T_{1,J,J_2} + T_{2,J,J_2} + T_{3,J,J_2} + T_{4,J,J_2}.
\end{aligned}$$

Control of T_{1,J,J_2} : Given a sequence $M_n \to \infty$, let $u_i = u_i^+ + u_i^-$ where

$$u_i^+ = u_i \mathbb{1}\{u_i^2 \max\left(\|G_{b,J}^{-1/2} b_{W_i}^{K(J)}\|_{\ell^2}^2, \|G_{b,J_2}^{-1/2} b_{W_i}^{K(J_2)}\|_{\ell^2}^2\right) \le M_n^2\}$$

and set $\Xi^\pm_{i,J,J_2} = (u_i^\pm)^2 G_{b,J}^{-1/2} b_{W_i}^{K(J)}(b_{W_i}^{K(J_2)})' G_{b,J_2}^{-1/2} - \mathbb{E}\big[(u_i^\pm)^2 G_{b,J}^{-1/2} b_{W_i}^{K(J)}(b_{W_i}^{K(J_2)})' G_{b,J_2}^{-1/2}\big]$. We may then write

$$G_{b,K(J)}^{-1/2}\left(\frac{1}{n}\sum_{i=1}^n u_i^2 b_{W_i}^{K(J)}(b_{W_i}^{K(J_2)})'\right) G_{b,K(J_2)}^{-1/2} - \Omega^o_{J,J_2} = \frac{1}{n}\sum_{i=1}^n \Xi^+_{i,J,J_2} + \frac{1}{n}\sum_{i=1}^n \Xi^-_{i,J,J_2}\,.$$

By definition of u_i^+ and Jensen's inequality, we have $\|\Xi^+_{i,J,J_2}\|_{\ell^2} \le 2M_n^2$ for all i. By Assumption 2(i) there is $\sigma^2 > 0$ such that $\mathbb{E}[u^2|W] \le \sigma^2$ (almost everywhere). For any $v \in \mathbb{R}^{K(J)}$ satisfying $\|v\|_{\ell^2} = 1$, we have that

$$\begin{aligned}
v'\mathbb{E}[\Xi^+_{i,J,J_2}(\Xi^+_{i,J,J_2})']v &= \mathbb{E}\left[(u_i^+)^4 \|G_{b,J_2}^{-1/2} b_{W_i}^{K(J_2)}\|_{\ell^2}^2 v' G_{b,J}^{-1/2} b_{W_i}^{K(J)}(b_{W_i}^{K(J)})' G_{b,J}^{-1/2} v\right] \\
&\le M_n^2\sigma^2 \mathbb{E}\left[v' G_{b,J}^{-1/2} b_{W_i}^{K(J)}(b_{W_i}^{K(J)})' G_{b,J}^{-1/2} v\right] = M_n^2\sigma^2
\end{aligned}$$

and so $\|\mathbb{E}[\Xi^+_{i,J,J_2}(\Xi^+_{i,J,J_2})']\|_{\ell^2} \le M_n^2\sigma^2$. Similarly, $\|\mathbb{E}[(\Xi^+_{i,J,J_2})'\Xi^+_{i,J,J_2}]\|_{\ell^2} \le M_n^2\sigma^2$. By Bernstein's inequality for random matrices (Tropp, 2012, Theorem 1.6), for any $C > 0$ we have

$$\mathbb{P}_{h_0}\left(\frac{1}{n}\sum_{i=1}^{n} \Xi^{+}_{i,J,J_2} > CM_n\sqrt{(\log \bar{J}_{\max})/n}\right) \leq 2K(J_2)\exp\left(-\frac{C^2\log \bar{J}_{\max}}{2\sigma^2 + 4CM_n\sqrt{(\log \bar{J}_{\max})/(9n)}}\right).$$

As $K(J) \leq a_b J$, it follow by a union bound argument that

$$\sup_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J,J_2\in\mathcal{J}_n}\frac{1}{n}\sum_{i=1}^{n} \Xi^{+}_{i,J,J_2} > CM_n\sqrt{(\log \bar{J}_{\max})/n}\right)$$
$$\leq 2\exp\left(3\log \bar{J}_{\max} - \frac{C^2\log \bar{J}_{\max}}{2\sigma^2 + 4CM_n\sqrt{(\log \bar{J}_{\max})/(9n)}}\right) \to 0$$

for any $C > \sigma\sqrt{6}$ provided $M_n\sqrt{(\log \bar{J}_{\max})/n} \to 0$ which holds for M_n defined below.

For $\frac{1}{n}\sum_{i=1}^{n}\Xi^{-}_{i,J,J_2}$, by definition of u_i^-, the Cauchy–Schwarz inequality, and the fact that $\sqrt{J} \leq \zeta_{b,J} \leq C'\sqrt{J}$ for all $J \in \mathcal{T}$ for some $C' > 0$ implies that

$$\sup_{J,J_2\in\mathcal{J}_n}\left\|(u_i^-)^2 G_{b,J}^{-1/2} b_{W_i}^{K(J)} (b_{W_i}^{K(J_2)})' G_{b,J_2}^{-1/2}\right\|_{\ell^2} \leq (C')^2 \bar{J}_{\max}|u_i|^2 \mathbb{1}\{C'\sqrt{\bar{J}_{\max}}|u_i| > M_n\}.$$

And so by Markov's inequality, for any $C > 0$ we have

$$\sup_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J,J_2\in\mathcal{J}_n}\frac{1}{n}\sum_{i=1}^{n} \Xi^{-}_{i,J,J_2} > CM_n\sqrt{(\log \bar{J}_{\max})/n}\right)$$
$$\leq \frac{(C')^4 \bar{J}_{\max}^2\sqrt{n}}{CM_n^3\sqrt{\log \bar{J}_{\max}}}\mathbb{E}\Big[|u|^4 \mathbb{1}\{|u| > M_n/(C'\sqrt{\bar{J}_{\max}})\}\Big] \to 0$$

with $M_n^3 = \bar{J}_{\max}^2\sqrt{n/(\log \bar{J}_{\max})}$ because $\mathbb{E}[|u|^4] < \infty$ under Assumption 2(i) and this choice of M_n satisfies $M_n/\sqrt{\bar{J}_{\max}} \to \infty$. This choice of M_n also satisfies the requirement imposed above, namely $M_n\sqrt{(\log \bar{J}_{\max})/n} = (\bar{J}_{\max}^2(\log \bar{J}_{\max})/n)^{\frac{1}{3}} \to 0$ because $\bar{J}_{\max}^2(\log \bar{J}_{\max}/n) \to 0$.

We have therefore shown that for any $C > \sigma\sqrt{6}$,

$$\sup_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J,J_2\in\mathcal{J}_n} T_{1,J,J_2} > C\left(\frac{\bar{J}_{\max}^2\log \bar{J}_{\max}}{n}\right)^{\frac{1}{3}}\right) \to 0\,.$$

Control of T_{2,J,J_2}: First note $|\hat{u}_{i,J} - u_i| \leq \|\hat{h}_J - h_0\|_\infty$. Using the Cauchy–Schwarz inequality, for any $v_1 \in \mathbb{R}^{K(J)}$ and $v_2 \in \mathbb{R}^{K(J_2)}$ satisfying $\|v_1\|_{\ell^2} = \|v_2\|_{\ell^2} = 1$ we may deduce

$$\begin{aligned} & v_1' \left(\frac{1}{n} \sum_{i=1}^n u_i (\hat{u}_{i,J} - u_i) G_{b,J}^{-1/2} b_{W_i}^{K(J)} (b_{W_i}^{K(J_2)})' G_{b,J_2}^{-1/2} \right) v_2 \\ & \leq \|\hat{h}_J - h_0\|_\infty^{1/2} \sqrt{ v_1' \left(\frac{1}{n} \sum_{i=1}^n (1 + u_i^2) G_{b,J}^{-1/2} b_{W_i}^{K(J)}) b_{W_i}^{K(J)})' G_{b,J}^{-1/2} \right) v_1 } \\ & \quad \times \|\hat{h}_{J_2} - h_0\|_\infty^{1/2} \sqrt{ v_2' \left(\frac{1}{n} \sum_{i=1}^n (1 + u_i^2) G_{b,J_2}^{-1/2} b_{W_i}^{K(J_2)} (b_{W_i}^{K(J_2)})' G_{b,J_2}^{-1/2} \right) v_2 } \,. \end{aligned}$$

It follows by the variational characterization of singular values that

$$\sup_{J,J_2 \in \mathcal{J}_n} T_{2,J,J_2} \leq \sup_{J \in \mathcal{J}_n} \|\hat{h}_J - h_0\|_\infty \times \sup_{J \in \mathcal{J}_n} \left\| \frac{1}{n} \sum_{i=1}^n (1 + u_i^2) G_{b,J}^{-1/2} b_{W_i}^{K(J)} (b_{W_i}^{K(J)})' G_{b,J}^{-1/2} \right\|_{\ell^2} .$$

Note that

$$\frac{1}{n} \sum_{i=1}^n (1 + u_i^2) G_{b,J}^{-1/2} b_{W_i}^{K(J)} (b_{W_i}^{K(J)})' G_{b,J}^{-1/2} = \widehat{G}_{b,J}^o + G_{b,J}^{-1/2} \left(\frac{1}{n} \sum_{i=1}^n u_i^2 b_{W_i}^{K(J)} (b_{W_i}^{K(J)})' \right) G_{b,J}^{-1/2}$$

Using Lemma C.3, control of T_{1,J,J_2} above, and the fact that $\Omega_{J,J}^o \leq \sigma^2 I_J$ for some $\sigma^2 > 0$ (by Assumption 2(i)), we may deduce that for any $c > 0$,

$$\sup_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(\sup_{J \in \mathcal{J}_n} \left\| \frac{1}{n} \sum_{i=1}^n (1 + u_i^2) G_{b,J}^{-1/2} b_{W_i}^{K(J)} (b_{W_i}^{K(J)})' G_{b,J}^{-1/2} \right\|_{\ell^2} > 1 + \sigma^2 + c \right) \to 0 \,.$$

Theorem B.1 and Appendix A.3 now imply that

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(\sup_{J,J_2 \in \mathcal{J}_n} T_{2,J,J_2} \leq C' \left(J_{\min}^{-\frac{p}{d}} + \tau_{\bar{J}_{\max}} \sqrt{ \frac{\bar{J}_{\max} \log \bar{J}_{\max}}{n} } \right) \right) \to 1$$

with $C' = (1 + \sigma^2 + c) C_{B.1} \max\{(1 + C_\Pi) B, 1\}$.

Control of T_{3,J,J_2}: Identical to that of T_{2,J,J_2}.

Control of T_{4,J,J_2}: Similar logic to that for T_{2,J,J_2} implies that there exists a constant $C'' > 0$ which satisfies

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(\sup_{J,J_2 \in \mathcal{J}_n} T_{4,J,J_2} \leq C'' \left(J_{\min}^{-\frac{p}{d}} + \sqrt{ \frac{\bar{J}_{\max} \log \bar{J}_{\max}}{n} } \right)^2 \right) \to 1.$$

The result follows becasuse $(J_{\min}^{-p/d} + \sqrt{(\bar{J}_{\max} \log \bar{J}_{\max})/n})^2 \leq J_{\min}^{-p/d} + \sqrt{(\bar{J}_{\max} \log \bar{J}_{\max})/n}$ for all sufficiently large n. ■

Proof of Lemma B.4. Define

$$\hat{\gamma}_{x,J} = (\widehat{G}^o_{b,J})^{-1}\widehat{S}^o_J[(\widehat{S}^o_J)'(\widehat{G}^o_{b,J})^{-1}\widehat{S}^o_J]^{-1}G_{\psi,J}^{-1/2}(\psi_x^J)\,, \qquad \gamma_{x,J} = S^o_J[(S^o_J)'S^o_J]^{-1}G_{\psi,J}^{-1/2}(\psi_x^J)$$

and so

$$\begin{aligned}\hat{\sigma}_{x,J,J_2} - \sigma_{x,J,J_2} &= \hat{\gamma}'_{x,J}\widehat{\Omega}^o_{J,J_2}\hat{\gamma}_{x,J_2} - \gamma'_{x,J}\Omega^o_{J,J_2}\gamma_{x,J_2} \\ &= (\hat{\gamma}_{x,J} - \gamma_{x,J})'\Omega^o_{J,J_2}\hat{\gamma}_{x,J_2} + \gamma'_{x,J}\Omega^o_{J,J_2}(\hat{\gamma}_{x,J_2} - \gamma_{x,J_2}) + \hat{\gamma}'_{x,J}(\widehat{\Omega}^o_{J,J_2} - \Omega^o_{J,J_2})\hat{\gamma}'_{x,J_2} \\ &=: T_{1,x,J,J_2} + T_{2,x,J,J_2} + T_{3,x,J,J_2}\,.\end{aligned}$$

First note $\|\sigma_{x,J}\| = \|(\psi_x^J)'G_{\psi,J}^{-1/2}(S^o_J)^-_l\|_{\ell^2} = \|\gamma_{x,J}\|_{\ell^2}$. By Assumption 2, there is a universal constant $C > 0$ such that $C^{-1}\|\sigma_{x,J}\| \le \|\sigma_{x,J}\|_{sd} \le C\|\sigma_{x,J}\|$ for all $(x,J) \in \mathcal{X} \times \mathcal{T}$. Therefore,

$$\begin{aligned}\frac{\|\hat{\gamma}_{x,J} - \gamma_{x,J}\|_{\ell^2}}{\|\sigma_{x,J}\|_{sd}} &= \left\|\frac{(G_{\psi,J}^{-1/2}\psi_x^J)'}{\|\sigma_{x,J}\|_{sd}}((\widehat{G}^o_{b,J})^{-1/2}\widehat{S}^o_J)^-_l(\widehat{G}^o_{b,J})^{-1/2} - \frac{(G_{\psi,J}^{-1/2}\psi_x^J)'}{\|\sigma_{x,J}\|_{sd}}(S^o_J)^-_l\right\|_{\ell^2} \\ &\le \left\|\frac{(G_{\psi,J}^{-1/2}\psi_x^J)'}{\|\sigma_{x,J}\|_{sd}}(S^o_J)^-_l\right\|_{\ell^2}\left\|S^o_J\{((\widehat{G}^o_{b,J})^{-1/2}\widehat{S}^o_J)^-_l(\widehat{G}^o_{b,J})^{-1/2} - (S^o_J)^-_l\}\right\|_{\ell^2} \\ &\le C\left\|S^o_J\{((\widehat{G}^o_{b,J})^{-1/2}\widehat{S}^o_J)^-_l(\widehat{G}^o_{b,J})^{-1/2} - (S^o_J)^-_l\}\right\|_{\ell^2}. \end{aligned} \tag{43}$$

By Lemma C.4(ii) and the fact that $\tau_{\bar{J}_{\max}}\sqrt{\bar{J}_{\max}(\log \bar{J}_{\max})/n} \to 0$, we can deduce that

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}\left|\frac{\|\hat{\gamma}_{x,J}\|_{\ell^2}}{\|\sigma_{x,J}\|_{sd}} - 1\right| \le c\right) \ge \inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}\frac{\|\hat{\gamma}_{x,J} - \gamma_{x,J}\|_{\ell^2}}{\|\sigma_{x,J}\|_{sd}} \le c\right) \to 1 \tag{44}$$

holds for any $c > 0$. Finally, from Assumption 2(i) we may deduce that $\sup_{J,J_2\in\mathcal{T}}\|\Omega^o_{J,J_2}\|_{\ell^2} \le \sigma^2$ where σ^2 is such that $\mathbb{E}[u^2|W] \le \sigma^2$ (almost everywhere).

Control of T_{1,x,J,J_2}: We have

$$|T_{1,x,J,J_2}| = \left|(\hat{\gamma}_{x,J} - \gamma_{x,J})'\Omega^o_{J,J_2}\hat{\gamma}_{x,J_2}\right| \le \sigma^2\|\hat{\gamma}_{x,J} - \gamma_{x,J}\|_{\ell^2}\|\hat{\gamma}_{x,J_2}\|_{\ell^2}$$

It then follows by (43), (44), and Lemma C.4(ii) that

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J,J_2)\in\mathcal{X}\times\mathcal{J}_n\times\mathcal{J}_n}\frac{|T_{1,x,J,J_2}|}{\|\sigma_{x,J}\|_{sd}\|\sigma_{x,J_2}\|_{sd}} \le \sigma^2(1+c)CC_{C.4}\tau_{\bar{J}_{\max}}\sqrt{\frac{\bar{J}_{\max}\log\bar{J}_{\max}}{n}}\right) \to 1.$$

Control of T_{2,x,J,J_2}: Follows similarly to T_{1,x,J,J_2}.

Control of T_{3,x,J,J_2}: We have

$$|T_{3,J,J_2}| = \left|\hat{\gamma}'_{x,J}(\widehat{\Omega}^o_{J,J_2} - \Omega^o_{J,J_2})\hat{\gamma}'_{x,J_2}\right| \le \|\hat{\gamma}_{x,J}\|_{\ell^2}\|\hat{\gamma}_{x,J_2}\|_{\ell^2}\|\widehat{\Omega}^o_{J,J_2} - \Omega^o_{J,J_2}\|_{\ell^2}\,.$$

It then follows by (44) and Lemma C.5 that

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J,J_2) \in \mathcal{X} \times \mathcal{J}_n \times \mathcal{J}_n} \frac{|T_{3,x,J,J_2}|}{\|\sigma_{x,J}\|_{sd} \|\sigma_{x,J_2}\|_{sd}} \leq (1+c)^2 C_{C.5} \delta_n \right) \to 1$$

as required. ■

Proof of Lemma B.3. Part (i): By the triangle inequality, we have $\|\sigma_{x,J_2}\|_{sd} - \|\sigma_{x,J}\|_{sd} \leq \|\sigma_{x,J,J_2}\|_{sd} \leq \|\sigma_{x,J}\|_{sd} + \|\sigma_{x,J_2}\|_{sd}$ so the upper bound follows from Assumption 4(i) and (weak) monotonicity of $J \mapsto \tau_J$. For the lower bound, Assumption 4(ii) implies that there exists $N \in \mathbb{N}$ such that for all $J \geq N$, we have

$$\sup_{x \in \mathcal{X}, J_2 \in \mathcal{T} : J_2 > J} \frac{\|\sigma_{x,J}\|_{sd}}{\|\sigma_{x,J_2}\|_{sd}} \leq \gamma + \frac{1-\gamma}{2} < 1\,,$$

and so $\|\sigma_{x,J,J_2}\|_{sd} \geq \|\sigma_{x,J_2}\|_{sd} - \|\sigma_{x,J}\|_{sd} \geq \frac{1}{2}(1-\gamma)\|\sigma_{x,J_2}\|_{sd}$ for every $x \in \mathcal{X}$ and $J, J_2 \in \mathcal{T}$ with $J_2 > J \geq N$.

Part (ii): As $|a-1| \leq |a^2 - 1|$ for $a \geq 0$, it suffices to bound

$$\left| \frac{\|\hat{\sigma}_{x,J,J_2}\|^2_{sd}}{\|\sigma_{x,J,J_2}\|^2_{sd}} - 1 \right|,$$

where

$$\|\hat{\sigma}_{x,J,J_2}\|^2_{sd} = \left(\|\hat{\sigma}_{x,J}\|^2_{sd} + \|\hat{\sigma}_{x,J_2}\|^2_{sd} - 2\hat{\sigma}_{x,J,J_2} \right), \quad \|\sigma_{x,J,J_2}\|^2_{sd} = \left(\|\sigma_{x,J}\|^2_{sd} + \|\sigma_{x,J_2}\|^2_{sd} - 2\sigma_{x,J,J_2} \right).$$

By the triangle inequality and Lemma B.4, we have

$$\left| \|\hat{\sigma}_{x,J,J_2}\|^2_{sd} - \|\sigma_{x,J,J_2}\|^2_{sd} \right| \leq C_{B.4} \delta_n \Big(\|\sigma_{x,J}\|_{sd} + \|\sigma_{x,J_2}\|_{sd} \Big)^2 \qquad \forall\, (x, J, J_2) \in \mathcal{S}_n$$

holds with $\mathbb{P}_{h_0}$ probability approaching 1 (uniformly over $h_0 \in \cup_{p \in [\underline{p}, \overline{p}]} \mathcal{H}^p$). To complete the proof, note that by Part (i) and Assumption 4(i), we have

$$\frac{\|\sigma_{x,J}\|_{sd} + \|\sigma_{x,J_2}\|_{sd}}{\|\sigma_{x,J,J_2}\|_{sd}} \leq C_{B.3} \left(\frac{\|\sigma_{x,J}\|_{sd}}{\|\sigma_{x,J_2}\|_{sd}} + 1 \right) \leq C_{B.3} \left(\frac{\tau_J \sqrt{\overline{C} J}}{\tau_{J_2} \sqrt{\underline{c} J_2}} + 1 \right) \leq C_{B.3} \left(\sqrt{\overline{C}/\underline{c}} + 1 \right)$$

where the final inequality is by (weak) monotonicity of $J \mapsto \tau_J$. ■

C.1.3 Supplemental Results: Uniform Consistency of $\hat{J}_{\max}$

We first present a preliminary Lemma before proving Lemma B.5 and Remark B.1.

Lemma C.6 *Let Assumptions 1(i)(ii) hold and let $\bar{J}_{\max} \to \infty$ with $(\bar{J}_{\max} \log \bar{J}_{\max})/n \to 0$.*

Then: there is a finite positive constant $C_{C.6}$ such that

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J \in \mathcal{T} \cap [1, \bar{J}_{\max}]} |\hat{s}_J - s_J| \leq C_{C.6}\sqrt{(\bar{J}_{\max} \log \bar{J}_{\max})/n} \right) \to 1 \,.$$

Proof of Lemma C.6. By Weyl's Inequality and the triangle inequality, we have

$$\begin{aligned}
|\hat{s}_J - s_J| &\leq \|\widehat{G}_{b,J}^{-1/2} \widehat{S}_J \widehat{G}_{\psi,J}^{-1/2} - S_J^o\|_{\ell^2} \\
&\leq \|\widehat{G}_{b,J}^{-1/2} G_{b,J}^{1/2} \widehat{S}_J^o G_{\psi,J}^{1/2} \widehat{G}_{\psi,J}^{-1/2} - \widehat{S}_J^o\|_{\ell^2} + \|\widehat{S}_J^o - S_J^o\|_{\ell^2} \\
&\leq \|\widehat{G}_{b,J}^{-1/2} G_{b,J}^{1/2} - I\|_{\ell^2} \|\widehat{S}_J^o\|_{\ell^2} \|G_{\psi,J}^{1/2} \widehat{G}_{\psi,J}^{-1/2}\|_{\ell^2} + \|\widehat{S}_J^o\|_{\ell^2} \|G_{\psi,J}^{1/2} \widehat{G}_{\psi,J}^{-1/2} - I\|_{\ell^2} + \|\widehat{S}_J^o - S_J^o\|_{\ell^2} \,.
\end{aligned}$$

By Lemma C.3, there is $C_{C.3}$ such that

$$\|\widehat{S}_J^o - S_J^o\|_{\ell^2} \leq C_{C.3}\sqrt{\frac{\bar{J}_{\max} \log \bar{J}_{\max}}{n}} \quad \forall \;\; J \in [1, \bar{J}_{\max}] \cap \mathcal{T}$$

wpa1 $\mathcal{H}$-uniformly. As $\|S_J^o\|_{\ell^2} \leq 1$, this also implies that $\|S_J^o\|_{\ell^2} \leq 2$ holds wpa1 $\mathcal{H}$-uniformly. For the remaining terms, first note that $\widehat{G}_{b,J}$ and $\widehat{G}_{\psi,J}$ are invertible for all $J \in \mathcal{T} \cap [1, \bar{J}_{\max}]$ wpa1 $\mathcal{H}$-uniformly by Lemma C.3. Then by Lemma F.3 of Chen and Christensen (2018), we may deduce (with an identical argument for the terms involving $\widehat{G}_{\psi,J}$) that the following inequalities hold for all $J \in \mathcal{T} \cap [1, \bar{J}_{\max}]$ wpa1 $\mathcal{H}$-uniformly:

$$\begin{aligned}
\|\widehat{G}_{b,J}^{-1/2} G_{b,J}^{1/2} - I\|_{\ell^2} &= \|\widehat{G}_{b,J}^{-1/2} (\widehat{G}_{b,J}^{1/2} - G_{b,J}^{1/2})\|_{\ell^2} \\
&\leq \frac{1}{\sqrt{\lambda_{\min}(\widehat{G}_{b,J}) + \lambda_{\min}(G_{b,J})}} \|\widehat{G}_{b,J} - G_{b,J}\|_{\ell^2} \|\widehat{G}_{b,J}^{-1/2}\|_{\ell^2} \\
&\leq \frac{1}{\sqrt{\lambda_{\min}(\widehat{G}_{b,J}) + \lambda_{\min}(G_{b,J})}} \|G_{b,J}^{1/2}\|_{\ell^2}^2 \|\widehat{G}_{b,J}^{-1/2}\|_{\ell^2} \times \|\widehat{G}_{b,J}^o - I\|_{\ell^2} \,.
\end{aligned}$$

Lemma C.3 together with Lemmas A.1 and A.2 ensures that the term in the preceding display that is pre-multiplying $\|\widehat{G}_{b,J}^o - I\|_{\ell^2}$ is bounded by some finite constant C for all $J \in \mathcal{T} \cap [1, \bar{J}_{\max}]$ wpa1 $\mathcal{H}$-uniformly. The conclusion then follows by Lemma C.3. ■

Lemma C.7 *Let Assumptions 1 and 3(i) hold and let $\bar{J}_{\max} \to \infty$ with $\tau_{\bar{J}_{\max}} \sqrt{(\bar{J}_{\max} \log \bar{J}_{\max})/n} \to 0$. Then: there is a finite positive constant $C_{C.7} > 1$ such that*

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J \in \mathcal{T} \cap [1, \bar{J}_{\max}]} \tau_J^{-1} \left| \hat{s}_J^{-1} - \tau_J \right| \leq C_{C.7} \right) \to 1 \,.$$

Proof of Lemma C.7. First note that by Lemma C.6, display (20), and the conditions on $\bar{J}_{\max}$ in the statement of the Lemma, that for any $\varepsilon > 0$ the inequality $|\hat{s}_J - s_J| \leq \varepsilon s_J$ holds for all $J \in \mathcal{T} \cap [1, \bar{J}_{\max}]$ wpa1 $\mathcal{H}$-uniformly, and therefore $|\hat{s}_J^{-1} - s_J^{-1}| \leq 2\varepsilon s_J^{-1}$ holds for all $J \in \mathcal{T} \cap [1, \bar{J}_{\max}]$ wpa1 $\mathcal{H}$-uniformly. The result now follows by display (20). ■

Proof of Lemma B.5. Part (i): We show that there exists a R_1 sufficiently small so that $\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}(\bar{J}_{\max}(R_1) \leq \hat{J}_{\max}) \to 1$. For any $R_1 > 0$, let $\mathcal{T}(R_1) = \mathcal{T} \cap [1, \bar{J}_{\max}(R_1)]$. Note that $\tau_{\bar{J}_{\max}(R_1)}(\bar{J}_{\max}(R_1) \log \bar{J}_{\max}(R_1))/n \to 0$ and so by definition of $\hat{J}_{\max}$ and $\bar{J}_{\max}(R_1)$, and Lemma C.7, we have

$$
\begin{aligned}
&\sup_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\hat{J}_{\max} < \bar{J}_{\max}(R_1)\right) \\
&\leq \sup_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J \in \mathcal{T}(R_1)} J\sqrt{\log J}\big[(\log n)^4 \vee \hat{s}_J^{-1}\big] > 2M\sqrt{n}\right) \\
&\leq \sup_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{J \in \mathcal{T}(R_1)} J\sqrt{\log J}\big[(\log n)^4 \vee \hat{s}_J^{-1}\big] > \frac{2M}{R_1} \sup_{J \in \mathcal{T}(R_1)} J\sqrt{\log J}\big[(\log n)^4 \vee \tau_J\big]\right) \\
&\leq \sup_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(C_{C.7} \sup_{J \in \mathcal{T}(R_1)} J\sqrt{\log J}\big[(\log n)^4 \vee \tau_J\big] > \frac{M}{R_1} \sup_{J \in \mathcal{T}(R_1)} J\sqrt{\log J}\big[(\log n)^4 \vee \tau_J\big]\right) + o(1)\,,
\end{aligned}
$$

which converges to 0 provided R_1 is sufficiently small that $C_{C.7} R_1 < M$.

Part (ii): We now show $\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}(\hat{J}_{\max} \leq \bar{J}_{\max}(R_2)) \to 1$ holds for R_2 sufficiently large. If $\hat{J}_{\max} > \bar{J}_{\max}(R_2)$, then $\bar{J}_{\max}(R_2)$ must fail at least one of the two inequalities given in the definition of $\hat{J}_{\max}$ in (4). Suppose it fails the first, so that

$$
\bar{J}_{\max}(R_2)\sqrt{\log \bar{J}_{\max}(R_2)}\big[(\log n)^4 \vee \hat{s}^{-1}_{\bar{J}_{\max}(R_2)}\big] > 2M\sqrt{n}\,.
$$

This implies that either $\hat{J}_{\max} < \bar{J}_{\max}(R_2)$ (a contradiction), or

$$
\inf_{J \in \mathcal{T}(R_2)} J\sqrt{\log J}\big[(\log n)^4 \vee \hat{s}_J^{-1}\big] > 2M\sqrt{n}\,,
$$

which has probability tending to 0 ($\mathcal{H}$-uniformly) by Lemma C.7.

Now suppose $\bar{J}_{\max}(R_2)$ fails the second inequality defining $\hat{J}_{\max}$ so that

$$
\bar{J}_{\max}(R_2)^+\sqrt{\log \bar{J}_{\max}(R_2)^+}\big[(\log n)^4 \vee \hat{s}^{-1}_{\bar{J}_{\max}(R_2)^+}\big] \leq 2M\sqrt{n}\,. \tag{45}
$$

Note that irrespective of the degree of ill-posedness, we have $(\bar{J}^+_{\max} \log \bar{J}_{\max}(R_2)^+)/n \to 0$, so by Lemma C.6 we see that

$$
\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\hat{s}_{\bar{J}_{\max}(R_2)^+} \leq s_{\bar{J}_{\max}(R_2)^+} + C_{C.6}\sqrt{\frac{\bar{J}_{\max}(R_2)^+ \log \bar{J}_{\max}(R_2)^+}{n}}\right) \to 1\,.
$$

If the inequality $C_{C.6}\sqrt{(\bar{J}_{\max}(R_2)^+ \log \bar{J}_{\max}(R_2)^+)/n} \leq s_{\bar{J}_{\max}(R_2)^+}$ holds, then $\hat{s}^{-1}_{\bar{J}_{\max}(R_2)^+} \geq$

$0.5s^{-1}_{\bar{J}_{\max}(R_2)^+} \geq 0.5\tau_{\bar{J}_{\max}(R_2)^+}$ and so by (45) and definition of $\bar{J}_{\max}(R_2)$ in (24), we have

$$\begin{aligned}2M\sqrt{n} &\geq \bar{J}_{\max}(R_2)^+\sqrt{\log \bar{J}_{\max}(R_2)^+}\big[(\log n)^4 \vee \hat{s}^{-1}_{\bar{J}_{\max}(R_2)^+}\big] \\ &\geq 0.5\bar{J}_{\max}(R_2)^+\sqrt{\log \bar{J}_{\max}(R_2)^+}\big[(\log n)^4 \vee \tau^{-1}_{\bar{J}_{\max}(R_2)^+}\big] > 0.5R_2\sqrt{n}\,,\end{aligned}$$

which is impossible whenever $R_2 > 4M$.

Conversely, if $C_{C.6}\sqrt{(\bar{J}_{\max}(R_2)^+ \log \bar{J}_{\max}(R_2)^+)/n} > s_{\bar{J}_{\max}(R_2)^+}$ holds then

$$\hat{s}^{-1}_{\bar{J}^+_{\max}(R_2)} \geq \frac{1}{2C_{C.6}}\sqrt{\frac{n}{\bar{J}_{\max}(R_2)^+ \log \bar{J}_{\max}(R_2)^+}}$$

and so this would imply that, for all n sufficiently large that we can ignore the $(\log n)^4$ term, we have

$$\begin{aligned}2M\sqrt{n} &\geq \bar{J}_{\max}(R_2)^+\sqrt{\log \bar{J}_{\max}(R_2)^+}\big[(\log n)^4 \vee \hat{s}^{-1}_{\bar{J}_{\max}(R_2)^+}\big] \\ &\geq \frac{\bar{J}_{\max}(R_2)^+\sqrt{\log \bar{J}_{\max}(R_2)^+}}{2C_{C.6}}\sqrt{\frac{n}{\bar{J}_{\max}(R_2)^+ \log \bar{J}_{\max}(R_2)^+}} = \frac{\sqrt{\bar{J}_{\max}(R_2)^+}}{2C_{C.6}}\sqrt{n}\end{aligned}$$

which is impossible since $\bar{J}_{\max}(R_2)^+ \to \infty$. ∎

Proof of Remark B.1. By definition of $\bar{J}_{\max}(R)$ we have $\bar{J}_{\max}(R_1) \leq \bar{J}_{\max}(R_2)$. Suppose $\bar{J}_{\max}(R_2)$ is not bounded by a multiple of $\bar{J}_{\max}(R_1)$, in which case $\bar{J}_{\max}(R_2)/\bar{J}_{\max}(R_1) \to +\infty$. Then $\bar{J}_{\max}(R_2) > \bar{J}_{\max}(R_1)^+ =: \inf\{J \in \mathcal{T} : J > \bar{J}_{\max}(R_1)\}$ for all n sufficiently large, in which case

$$\begin{aligned}R_1\sqrt{n} &< \bar{J}_{\max}(R_1)^+\sqrt{\log \bar{J}_{\max}(R_1)^+}\big[(\log n)^4 \vee \tau_{\bar{J}_{\max}(R_1)^+}\big] \\ &< \bar{J}_{\max}(R_2)\sqrt{\log \bar{J}_{\max}(R_2)}\big[(\log n)^4 \vee \tau_{\bar{J}_{\max}(R_2)}\big] \leq R_2\sqrt{n}\,,\end{aligned}$$

which rearranges to give

$$\frac{\bar{J}_{\max}(R_2)\sqrt{\log \bar{J}_{\max}(R_2)}\big[(\log n)^4 \vee \tau_{\bar{J}_{\max}(R_2)}\big]}{\bar{J}_{\max}(R_1)^+\sqrt{\log \bar{J}_{\max}(R_1)^+}\big[(\log n)^4 \vee \tau_{\bar{J}_{\max}(R_1)^+}\big]} \leq \frac{R_2}{R_1}\,,$$

which in turn implies $\bar{J}_{\max}(R_2)/\bar{J}_{\max}(R_1)^+ \leq R_2/R_1$ because $J \mapsto \sqrt{\log J}\big[(\log n)^4 \vee \tau_J\big]$ is increasing in J for each n. As $\bar{J}_{\max}(R_1)^+ = (1+o(1))2^d\bar{J}_{\max}(R_1)$, this final bound contradicts $\bar{J}_{\max}(R_2)/\bar{J}_{\max}(R_1) \to +\infty$. ∎

C.1.4 Supplemental Results: Uniform-in-J Bounds for the Bootstrap

Before proving Lemmas B.6 and B.7, we first state and prove a preliminary lemma. Recall the definition of $\mathbb{Z}_n^*$ from (17) and $\widehat{\mathbb{Z}}_n$ from (18).

Lemma C.8 *Let Assumptions 1-4 hold. Let $\bar{J}_{\max} \equiv \bar{J}_{\max}(R)$ from (24) for any $R > 0$, let $J_{\min} \to \infty$ with $J_{\min} \leq \bar{J}_{\max}$, let $\mathcal{J}_n = \{J \in \mathcal{T} : J_{\min} \leq J \leq \bar{J}_{\max}\}$, and let δ_n from (25) be defined with this $J_{\min}$ and $\bar{J}_{\max}$. Then: given any $A > 0$, there exists $a > 0$ which satisfies:*

$$(i) \quad \inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\mathbb{P}^*\left(\sup_{(x,J) \in \mathcal{X} \times \mathcal{J}_n} \left|\mathbb{Z}_n^*(x,J) - \widehat{\mathbb{Z}}_n(x,J)\right| > a\delta_n \sqrt{\log \bar{J}_{\max}}\right) \leq \bar{J}_{\max}^{-A}\right) \to 1,$$

$$(ii) \quad \inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\mathbb{P}^*\left(\sup_{(x,J) \in \mathcal{X} \times \mathcal{J}_n} \left|\widehat{\mathbb{Z}}_n(x,J)\right| > a \sqrt{\log \bar{J}_{\max}}\right) \leq \bar{J}_{\max}^{-A}\right) \to 1,$$

$$(iii) \quad \inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\mathbb{P}^*\left(\sup_{(x,J) \in \mathcal{X} \times \mathcal{J}_n} |\mathbb{Z}_n^*(x,J)| > a \sqrt{\log \bar{J}_{\max}}\right) \leq \bar{J}_{\max}^{-A}\right) \to 1.$$

Proof of Lemma C.8. By Assumption 2, the eigenvalues of Ω_J^o are bounded away from 0 and ∞ uniformly in J. It follows by Lemma C.5 that the eigenvalues of $\widehat{\Omega}_J^o$ are bounded away from 0 and ∞ uniformly in $J \in \mathcal{J}_n$, wpa1 $\mathcal{H}$-uniformly. Whenever $\widehat{\Omega}_J^o$ is invertible,

$$\mathbb{Z}_n^*(x,J) = \frac{[G_{\psi,J}^{-1/2}\psi_x^J]'}{\|\widehat{\sigma}_{x,J}\|_{sd}}((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2}G_n^*\,,$$

$$\widehat{\mathbb{Z}}_n(x,J) = \frac{[G_{\psi,J}^{-1/2}\psi_x^J]'}{\|\sigma_{x,J}\|_{sd}}(S_J^o)_l^-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}G_n^*\,, \qquad \text{where} \quad G_n^* \sim \mathcal{N}(0, \widehat{\Omega}_J^o)\,.$$

Part (i): We have

$$\begin{aligned}
&|\mathbb{Z}_n^*(x,J) - \widehat{\mathbb{Z}}_n(x,J)| \\
&\leq \left|\frac{[G_{\psi,J}^{-1/2}\psi_x^J]'}{\|\sigma_{x,J}\|_{sd}}\Big(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2} - (S_J^o)_l^-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}\Big)G_n^*\right| \times \frac{\|\sigma_{x,J}\|_{sd}}{\|\widehat{\sigma}_{x,J}\|_{sd}} \\
&\quad + \left|\frac{[G_{\psi,J}^{-1/2}\psi_x^J]'}{\|\sigma_{x,J}\|_{sd}}(S_J^o)_l^-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}G_n^*\right| \left|1 - \frac{\|\sigma_{x,J}\|_{sd}}{\|\widehat{\sigma}_{x,J}\|_{sd}}\right| \quad =: T_{1,x,J} + T_{2,x,J}.
\end{aligned}$$

Control of $T_{1,x,J}$: In view of Lemma B.4, we have $\sup_{x \in \mathcal{X}, J \in \mathcal{J}_n} \|\sigma_{x,J}\|_{sd}/\|\widehat{\sigma}_{x,J}\|_{sd} \leq 2$ wpa1 $\mathcal{H}$-uniformly. It therefore suffices to control the process

$$\Delta\mathbb{Z}_n(x,J) := \frac{[G_{\psi,J}^{-1/2}\psi_x^J]'}{\|\sigma_{x,J}\|_{sd}}\Big(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2} - (S_J^o)_l^-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}\Big)G_n^*\,.$$

We first bound the terms $\|\Delta\mathbb{Z}_n(x,J)\|_{L^2(\mathbb{P}^*)}$ and $\|\Delta\mathbb{Z}_n(x_1,J) - \Delta\mathbb{Z}_n(x_2,J)\|_{L^2(\mathbb{P}^*)}$ for $x_1 \neq x_2$, where the $L^2(\mathbb{P}^*)$ norm is with respect to $(\varpi_i)_{i=1}^n$, conditional on $\mathcal{Z}^n$. Wpa1 $\mathcal{H}$-uniformly, we have

$$\begin{aligned}
&\|\sigma_{x,J}\|_{sd}\|\Delta\mathbb{Z}_n(x,J)\|_{L^2(\mathbb{P}^*)}\\
&=\left\|[G_{\psi,J}^{-1/2}\psi_x^J]'\Big(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2}-(S_J^o)_l^-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}\Big)(\widehat{\Omega}_J^o)^{1/2}\right\|_{\ell^2}\\
&=\left\|[G_{\psi,J}^{-1/2}\psi_x^J]'(S_J^o)_l^-S_J^o\Big(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2}-(S_J^o)_l^-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}\Big)(\widehat{\Omega}_J^o)^{1/2}\right\|_{\ell^2}.
\end{aligned}$$

Given the uniform bounds on the eigenvalues of Ω_J^o and $\widehat{\Omega}_J^o$ for $J\in\mathcal{J}_n$ wpa1 $\mathcal{H}$-uniformly, there exists a finite positive constant C for which

$$\begin{aligned}
\|\Delta\mathbb{Z}_n(x,J)\|_{L^2(\mathbb{P}^*)}\le\frac{C}{\|\sigma_{x,J}\|_{sd}}&\left\|[G_{\psi,J}^{-1/2}\psi_x^J]'(S_J^o)_l^-\right\|_{\ell^2}\\
&\times\left\|S_J^o\Big(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2}-(S_J^o)_l^-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}\Big)\right\|_{\ell^2}
\end{aligned}\tag{46}$$

holds for every $(x,J)\in\mathcal{X}\times\mathcal{J}_n$. By Assumption 2, there is a universal constant $C'>0$ such that $C'^{-1}\|\sigma_{x,J}\|\le\|\sigma_{x,J}\|_{sd}\le C'\|\sigma_{x,J}\|$ for all $(x,J)\in\mathcal{X}\times\mathcal{T}$, which yields

$$\|\Delta\mathbb{Z}_n(x,J)\|_{L^2(\mathbb{P}^*)}\le CC'\left\|S_J^o\Big(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2}-(S_J^o)_l^-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}\Big)\right\|_{\ell^2}.$$

We now show that the remaining term converges uniformly for $J\in\mathcal{J}_n$ at rate δ_n, uniformly for $h_0\in\mathcal{H}$. First note

$$\begin{aligned}
&\left\|S_J^o\Big(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2}-(S_J^o)_l^-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}\Big)\right\|_{\ell^2}\\
&\le\left\|S_J^o\Big(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2}-(S_J^o)_l^-\Big)\right\|_{\ell^2}+\left\|I-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}\right\|_{\ell^2}.
\end{aligned}$$

The first term on the r.h.s. can be bounded using Lemma C.4(ii) by a quantity which in turn can be bounded by a multiple of δ_n. For the second term on the r.h.s., we use Lemma C.5 and Lemmas F.2 and F.3 of Chen and Christensen (2018) to deduce that

$$\inf_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\sup_{J\in\mathcal{J}_n}\left\|I-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}\right\|_{\ell^2}\le C''\delta_n\right)\to 1,$$

for some constant $C''>0$, from which it follows that

$$\inf_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}\|\Delta\mathbb{Z}_n(x,J)\|_{L^2(\mathbb{P}^*)}\le C'''\delta_n\right)\to 1,\tag{47}$$

for some constant $C'''>0$.

By analogy with (46), for a finite positive constant C we have

$$\|\Delta\mathbb{Z}_n(x_1,J)-\Delta\mathbb{Z}_n(x_2,J)\|_{L^2(\mathbb{P}^*)} \le C\left\|\left(\frac{[G_{\psi,J}^{-1/2}\psi_{x_1}^J]'}{\|\sigma_{x_1,J}\|_{sd}}-\frac{[G_{\psi,J}^{-1/2}\psi_{x_2}^J]'}{\|\sigma_{x_2,J}\|_{sd}}\right)(S_J^o)_l^-\right\|_{\ell^2} \times\left\|S_J^o\left(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2}-(S_J^o)_l^-(\Omega_J^o)^{1/2}(\widehat{\Omega}_J^o)^{-1/2}\right)\right\|_{\ell^2}.$$

As above, the second norm on the right-hand side converges uniformly for $J\in\mathcal{J}_n$ at rate δ_n, uniformly for $h_0\in\mathcal{H}$. For the first term, by similar steps to the proof of part (ii) below it follows that

$$\inf_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\sup_{J\in\mathcal{J}_n}\sup_{x_1,x_2\in\mathcal{X}:x_1\neq x_2}\frac{\|\Delta\mathbb{Z}_n(x_1,J)-\Delta\mathbb{Z}_n(x_2,J)\|_{L^2(\mathbb{P})}}{J^{\omega-1/2}\|x_1-x_2\|_{\ell^2}^{\omega'}}\le C''''\delta_n\right)\to 1 \qquad (48)$$

where ω,ω' are the Hölder continuity constants for the sieve ψ^J and C'''' is a positive constant.

Finally, by (47) and (48), with $D=C'''\vee C''''$ we have that

$$\sup_{J\in\mathcal{J}_n}\sup_{\substack{x_1,x_2\in\mathcal{X}:\\x_1\neq x_2}}\frac{\|\Delta\mathbb{Z}_n(x_1,J)-\Delta\mathbb{Z}_n(x_2,J)\|_{L^2(\mathbb{P}^*)}}{\delta_n J^{\omega-1/2}\|x_1-x_2\|_{\ell^2}^{\omega'}},\ \sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}\|\Delta\mathbb{Z}_n(x,J)\|_{L^2(\mathbb{P}^*)}\le D\delta_n \qquad (49)$$

wpa1 $\mathcal{H}$-uniformly. The process $\mathcal{X}\times\mathcal{J}_n\ni(x,J)\to\Delta\mathbb{Z}_n(x,J)$ is sub-gaussian with respect to the pseudometric $\|\Delta\mathbb{Z}_n(x_1,J_1)-\mathbb{Z}_n(x_2,J_2)\|_{L^2(\mathbb{P}^*)}=\varphi_n[(x_1,J_1),(x_2,J_2)]$. If (49) holds then for any $\epsilon>0$ the ϵ-covering number $N(\mathcal{X}\times\mathcal{J}_n,\varphi_n,\epsilon)$ of $\mathcal{X}\times\mathcal{J}_n$ with respect to φ_n is

$$N(\mathcal{X}\times\mathcal{J}_n,\varphi_n,\epsilon)\lesssim\left(\frac{D\delta_n\bar{J}_{\max}}{\epsilon}\right)^v,$$

for some $v>0$. It follows by Theorem 2.3.6 of Giné and Nickl (2016) that

$$\mathbb{E}^*\left[\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}|\Delta\mathbb{Z}_n(x,J)|\right]\lesssim\int_0^{D\delta_n}\sqrt{\log\left(\frac{D\delta_n\bar{J}_{\max}}{\epsilon}\right)}\mathrm{d}\epsilon\le D'\delta_n\sqrt{\log\bar{J}_{\max}}$$

for some $D'>0$. For any $A>0$, it now follows from the above display and Theorem 2.5.8 of Giné and Nickl (2016) that wpa1 $\mathcal{H}$-uniformly, we have

$$\mathbb{P}^*\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}|\Delta\mathbb{Z}_n(x,J)|>(D'+\sqrt{2A}D)\delta_n\sqrt{\log\bar{J}_{\max}}\right)\le\bar{J}_{\max}^{-A}.$$

Control of $T_{2,x,J}$: The argument is similar to the above. Lemma B.4 yields a convergence rate of δ_n for the term $\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}|\|\hat{\sigma}_{x,J}\|_{sd}/\|\sigma_{x,J}\|_{sd}-1|$, while the $\sqrt{\log\bar{J}_{\max}}$ appears as the order of the supremum of a suitably normalized Gaussian process.

Part (ii): By analogy with (46), for any $J \in \mathcal{J}_n$ and $x_1, x_2 \in \mathcal{X}$, we have

$$\|\widehat{\mathbb{Z}}_n(x_1,J) - \widehat{\mathbb{Z}}_n(x_2,J)\|_{L^2(\mathbb{P}^*)} = \left\| \left(\frac{[G_{\psi,J}^{-1/2}\psi_{x_1}^J]'}{\|\sigma_{x_1,J}\|_{sd}} - \frac{G_{\psi,J}^{-1/2}[\psi_{x_2}^J]'}{\|\sigma_{x_2,J}\|_{sd}} \right) (S_J^o)_l^- (\Omega_J^o)^{1/2} \right\|_{\ell^2}.$$

Note the r.h.s. expression is of the form $|\frac{x}{\|x\|} - \frac{y}{\|y\|}|$, which we may bound using the fact that for any norm $\|\cdot\|$, we have

$$\left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\| \leq 2\frac{\|x-y\|}{\|x\| \vee \|y\|} \qquad \forall\, x,y \in \mathbb{R}^n \setminus \{0\}.$$

By Assumption 4, $(\|\sigma_{x_1,J}\|_{\ell^2} \vee \|\sigma_{x_2,J}\|_{\ell^2})^{-1} \leq (\sqrt{\underline{c}}\tau_J\sqrt{J})^{-1}$. For the numerator, we have

$$\left\| \left([G_{\psi,J}^{-1/2}\psi_{x_1}^J] - G_{\psi,J}^{-1/2}[\psi_{x_2}^J] \right)' (S_J^o)_l^- (\Omega_J^o)^{1/2} \right\|_{\ell^2} \leq C s_J^{-1} \|G_{\psi,J}^{-1/2}([\psi_{x_1}^J] - [\psi_{x_2}^J])\|_{\ell^2}$$

for some finite positive constant C, because the eigenvalues of Ω_J^o are uniformly bounded away from 0 and ∞. By Hölder continuity of the sieve ψ^J and $s_J^{-1} \leq a_\tau \tau_J$ (cf. (20)), we may deduce that there is a finite positive constant C' for which

$$\|\widehat{\mathbb{Z}}_n(x_1,J) - \widehat{\mathbb{Z}}_n(x_2,J)\|_{L^2(\mathbb{P}^*)} \leq C' J^{\omega-1/2} \|x_1 - x_2\|_{\ell^2}^{\omega'} \quad \forall \;\, J \in \mathcal{J}_n\,.$$

Note $\widehat{\mathbb{Z}}_n(x,J)$ is sub-Gaussian with respect to the metric $\varphi_n((x_1,J_1),(x_2,J_2)) = \|\widehat{\mathbb{Z}}_n(x_1,J) - \widehat{\mathbb{Z}}_n(x_2,J_2)\|_{L^2(\mathbb{P}^*)}$. As in the proof of part (i), we may deduce that the covering number $N(\mathcal{X} \times \mathcal{J}_n, \varphi_n, \epsilon)$ of $\mathcal{X} \times \mathcal{I}_n$ under φ_n at all $\epsilon > 0$ is of the form

$$N(\mathcal{X} \times \mathcal{J}_n, \varphi_n, \epsilon) \lesssim \left(\frac{\bar{J}_{\max}}{\epsilon} \right)^v$$

for some finite positive constant v. As $\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} \mathbb{E}^*[\widehat{\mathbb{Z}}_n(x,J)^2] = 1$, the result follows by applying Theorems 2.3.6 and 2.5.8 of Giné and Nickl (2016) as in the proof of part (i).

Part (iii): Follows by similar arguments to part (ii). ■

Proof of Lemma B.6. First note by Lemma B.5 that there exist constants $R_1, R_2 > 0$ such that

$$\bar{J}_{\max}(R_1) \leq \hat{J}_{\max} \leq \bar{J}_{\max}(R_2) \tag{50}$$

wpa1 $\mathcal{H}$-uniformly. This implies that $\hat{\mathcal{J}} \supseteq \{J \in \mathcal{T} : 0.1(\log \bar{J}_{\max}(R_2))^2 \leq J \leq \bar{J}_{\max}(R_1)\}$ wpa1 $\mathcal{H}$-uniformly. Also note that $\bar{J}_{\max}(R_1) \asymp \bar{J}_{\max}(R_2)$ by Remark B.1, which implies that $\hat{\mathcal{J}}$ contains at least two elements—and hence that $\hat{\mathcal{S}}$ is nonempty—wpa1 $\mathcal{H}$-uniformly.

For the lower bound, note that for every fixed $(x,J,J_2) \in \hat{\mathcal{S}}$, the quantity $\mathbb{Z}_n^*(x,J,J_2)$ is a $N(0,1)$ random variable under $\mathbb{P}^*$ and so it follows that $\theta^*(\hat{\alpha})$ is not smaller than the $(1-\hat{\alpha})$

quantile of the $N(0,1)$ distribution and, in view of (50), no smaller than the $(1-\bar{J}_{\max}(R_1)^{-1})$ quantile of the $N(0,1)$ distribution wpa1 $\mathcal{H}$-uniformly. Standard approximations to the $N(0,1)$ quantile function (e.g. DasGupta (2008), Example 8.13) yield

$$\Phi^{-1}((1-\bar{J}_{\max}(R_1)^{-1})) \asymp \sqrt{\log \bar{J}_{\max}(R_1)} = (1+o(1))\sqrt{\log \bar{J}_{\max}(R)}\,,$$

where the final equality is by Remark B.1. We have therefore shown that there exists a constant $C_4 > 0$ for which

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(C_4\sqrt{\log \bar{J}_{\max}(R)} \le \theta^*(\hat{\alpha})\right) \to 1\,.$$

For the upper bound, note by (50) that $\hat{\mathcal{J}} \subseteq \{J \in \mathcal{T} : 0.1(\log \bar{J}_{\max}(R_1))^2 \le J \le \bar{J}_{\max}(R_2)\}$ wpa1 $\mathcal{H}$-uniformly. Then by Lemmas B.3 and B.4 (with $J_{\min} = 0.1(\log \bar{J}_{\max}(R_1))^2$, $\bar{J}_{\max} = \bar{J}_{\max}(R_2)$, and $\mathcal{J}_n = \{J \in \mathcal{T} : J_{\min} \le J \le \bar{J}_{\max}\}$) and Assumption 4(i) imply that there is a finite positive constant C for which

$$\begin{aligned}
|\mathbb{Z}_n^*(x,J,J_2)| &= \left|\frac{1}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\hat{L}_{J,x}b_{W_i}^{K(J)}\hat{u}_{i,J} - \hat{L}_{J_2,x}b_{W_i}^{K(J_2)}\hat{u}_{i,J_2}\right)\varpi_i\right)\right| \\
&\le C\left(\left|\frac{1}{\|\hat{\sigma}_{x,J}\|_{sd}}\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{L}_{J,x}b_{W_i}^{K(J)}\hat{u}_{i,J}\varpi_i\right| + \left|\frac{1}{\|\hat{\sigma}_{x,J_2}\|_{sd}}\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{L}_{J_2,x}b_{W_i}^{K(J_2)}\hat{u}_{i,J_2}\varpi_i\right|\right) \\
&= C\big(\,|\mathbb{Z}_n^*(x,J)| + |\mathbb{Z}_n^*(x,J_2)|\,\big)
\end{aligned}$$

holds for all $(x,J,J_2) \in \mathcal{S}_n := \{(x,J,J_2) \in \mathcal{X}\times\mathcal{J}_n\times\mathcal{J}_n : J_2 > J\}$ wpa1 $\mathcal{H}$-uniformly. It follows by this inequality and (50) that

$$\theta^*(\hat{\alpha}) \le 2C \times (1-\bar{J}_{\max}(R_2)^{-1}) \text{ quantile of } \sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}|\mathbb{Z}_n^*(x,J)|$$

wpa1 $\mathcal{H}$-uniformly. The result now follows by Lemma C.8(iii) and Remark B.1. ■

Proof of Lemma B.7. First note by the proof of Lemma B.6 that $\hat{\mathcal{J}} \subseteq \overline{\mathcal{J}}_n := \{J \in \mathcal{T} : 0.1(\log \bar{J}_{\max}(R_1))^2 \le J \le \bar{J}_{\max}(R_2)\}$ wpa1 $\mathcal{H}$-uniformly. Therefore, wpa1 $\mathcal{H}$-uniformly we have $z_{1-\alpha}^* \le \overline{z}_{1-\alpha}^*$, the $1-\alpha$ quantiles of $\sup_{(x,J)\in\mathcal{X}\times\overline{\mathcal{J}}_n}|\mathbb{Z}_n^*(x,J)|$.

Let φ_n denote the pseudometric on $\mathcal{X}\times\overline{\mathcal{J}}_n$ given by $\|\mathbb{Z}_n^*(x_1,J_1) - \mathbb{Z}_n^*(x_2,J_2)\|_{L^2(\mathbb{P}^*)} = \varphi_n[(x_1,J_1),(x_2,J_2)]$. It follows by a similar argument to Lemma C.10 that

$$N(\mathcal{X}\times\overline{\mathcal{J}}_n, \varphi_n, \epsilon) \le D\left(\frac{\bar{J}_{\max}(R_2)}{\epsilon}\right)^v$$

for constants $D, v > 0$. It now follows by Theorems 2.3.6 and 2.5.8 of Giné and Nickl (2016) that

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\overline{\mathcal{J}}_n}|\mathbb{Z}_n^*(x,J)| \le C\sqrt{\log \bar{J}_{\max}(R_2)}\right) \to 1$$

for some constant $C > 0$ is some universal constant. As $\bar{J}_{\max}(R) \asymp \bar{J}_{\max}(R_2)$ by Remark B.1, we have $\bar{z}^*_{1-\alpha} \leq C'\sqrt{\log \bar{J}_{\max}}$ wpa1 $\mathcal{H}$-uniformly for some constant $C' > 0$. ■

C.1.5 Supplemental Results: Uniform Consistency for the Bootstrap

Before proving this result, we will first state and prove several preliminary results. Define

$$\mathcal{F}_n = \left\{ g : g(w,u) = \pm \|\sigma_{x,J}\|_{sd}^{-1} L_{J,x} b_w^{K(J)} u \quad \text{for some} \quad (x,J) \in \mathcal{X} \times \mathcal{J}_n \right\} \cup \{0\} . \tag{51}$$

Also let $\mathcal{F}_n - \mathcal{F}_n = \{g_1 - g_2 : g_1, g_2 \in \mathcal{F}_n\}$, $\mathcal{F}_n^2 = \{g_1 g_2 : g_1, g_2 \in \mathcal{F}_n\}$, and $(\mathcal{F}_n - \mathcal{F}_n)^2 = \{(g_1 - g_2)(g_3 - g_4) : g_1, g_2, g_3, g_4 \in \mathcal{F}_n\}$. We say a class of functions $\mathcal{F}$ of (w,u) is *VC(F, M)* if $|f| \leq F$ for all $f \in \mathcal{F}$ (i.e., F is an envelope for $\mathcal{F}_n$) and there exist constants $D, v > 0$ for which

$$\sup_Q N(\mathcal{F}, \|\cdot\|_{L^2(Q)}, \epsilon \|F\|_{L^2(Q)}) \leq D \left(\frac{M}{\epsilon} \right)^v \quad \forall \epsilon \in (0,1] ,$$

where the supremum is over all discrete probability measures on $\mathcal{W} \times \mathbb{R}$ for which $\|F\|_{L^2(Q)} > 0$, and $N(\mathcal{F}, d, \epsilon)$ denotes the ϵ-covering number of the set $\mathcal{F}$ under the pseudometric d.

Lemma C.9 *Let Assumptions 1, 2, 3(i) and 4(i) hold. Then: there exists sufficiently large positive constant C for which $F(w,u) = C\sqrt{\bar{J}_{\max}}|u|$ is an envelope for $\mathcal{F}_n$, and*

(i) $\mathcal{F}_n$ is VC($F, \bar{J}_{\max}$);

(ii) $\mathcal{F}_n^2$ is VC($F^2, \bar{J}_{\max}$);

(iii) $\mathcal{F}_n - \mathcal{F}_n$ is VC($2F, \bar{J}_{\max}$);

(iv) $(\mathcal{F}_n - \mathcal{F}_n)^2$ is VC($4F^2, \bar{J}_{\max}$).

Proof of Lemma C.9. First, by Assumptions 2 and 4(i), we may deduce that there is a finite positive constant C such that $|g(w,u)| \leq C\sqrt{J}|u| \leq C\sqrt{\bar{J}_{\max}}|u|$ for each $g \in \mathcal{F}_n$. We therefore let

$$F(w,u) = C\sqrt{\bar{J}_{\max}}|u|$$

denote the envelope of $\mathcal{F}_n$.

Part (i): Let

$$\mathcal{F}_{n,J} = \left\{ g : g(w,u) = \|\sigma_{x,J}\|_{sd}^{-1} L_{J,x} b_w^{K(J)} u \quad \text{for some} \quad x \in \mathcal{X} \right\} \cup \{0\} ,$$

and note that $\mathcal{F}_n = \cup_{J \in \mathcal{J}_n} (\mathcal{F}_{n,J} \cup -\mathcal{F}_{n,J})$, where $-\mathcal{F}_{n,J} = \{-g : g \in \mathcal{F}_{n,J}\}$.

Also let $A_J = (\Omega_J^o)^{1/2} S_J^o [(S_J^o)' S_J^o]^{-1}$ and observe that

$$L_{J,x} b_w^{K(J)} u = [G_{\psi,J}^{-1/2} \psi_x^J]' A_J' (\Omega_J^o)^{-1/2} (G_{b,J}^{-1/2} b_w^{K(J)}) u \,,$$

where the minimum and maximum eigenvalues of Ω_J^o are uniformly (in J and h_0) bounded away from 0 and $+\infty$ by Assumption 2. Also note $\sup_{h_0 \in \mathcal{H}} \sup_w \|G_{b,J}^{-1/2} b_w^{K(J)}\|_{\ell^2} \lesssim \sqrt{\bar{J}_{\max}}$ by Assumption 1(ii). Now, for any $x_0, x_1 \in \mathcal{X}$, we have

$$\begin{aligned} \left| \frac{L_{J,x_0} b_w^{K(J)} u}{\|\sigma_{x_0,J}\|_{sd}} - \frac{L_{J,x_1} b_w^{K(J)} u}{\|\sigma_{x_1,J}\|_{sd}} \right| &\lesssim \left\| \frac{A_J [G_{\psi,J}^{-1/2} \psi_{x_0}^J]}{\|A_J [G_{\psi,J}^{-1/2} \psi_{x_0}^J]\|_{\ell^2}} - \frac{A_J [G_{\psi,J}^{-1/2} \psi_{x_1}^J]}{\|A_J [G_{\psi,J}^{-1/2} \psi_{x_1}^J]\|_{\ell^2}} \right\|_{\ell^2} \sqrt{\bar{J}_{\max}} |u| \\ &\leq 2 \left\| \frac{A_J ([G_{\psi,J}^{-1/2} \psi_{x_0}^J] - [G_{\psi,J}^{-1/2} \psi_{x_1}^J])}{\|A_J [G_{\psi,J}^{-1/2} \psi_{x_0}^J]\|_{\ell^2} \vee \|A_J [G_{\psi,J}^{-1/2} \psi_{x_1}^J]\|_{\ell^2}} \right\|_{\ell^2} \sqrt{\bar{J}_{\max}} |u| \end{aligned}$$

where the second line uses the fact that for any norm $\|\cdot\|$ on $\mathbb{R}^n$ we have

$$\left\| \frac{x}{\|x\|} - \frac{y}{\|y\|} \right\| \leq 2 \frac{\|x-y\|}{\|x\| \vee \|y\|} \qquad \forall\, x, y \in \mathbb{R}^n \setminus \{0\}.$$

Note $\|A_J\|_{\ell_2} \lesssim s_J^{-1} \leq a_\tau \tau_J$. As the sieve ψ^J is Hölder continuous, i.e. $\|G_{\psi,J}^{-1/2} \psi_{x_0}^J - G_{\psi,J}^{-1/2} \psi_{x_1}^J\|_{\ell^2} \lesssim J^\omega \|x_0 - x_1\|_{\ell^2}^{\omega'}$ uniformly in J for some positive constants ω, ω', we have

$$\left| \frac{L_{J,x_0} b_w^{K(J)} u}{\|\sigma_{x_0,J}\|_{sd}} - \frac{L_{J,x_1} b_w^{K(J)} u}{\|\sigma_{x_1,J}\|_{sd}} \right| \lesssim \frac{\tau_J J^\omega \|x_0 - x_1\|_{\ell^2}^{\omega'} \sqrt{\bar{J}_{\max}} |u|}{\|\sigma_{x_0,J}\| \vee \|\sigma_{x_1,J}\|} \,.$$

It now follows by Assumption 4(i) that

$$\left| \frac{L_{J,x_0} b_w^{K(J)} u}{\|\sigma_{x_0,J}\|_{sd}} - \frac{L_{J,x_1} b_w^{K(J)} u}{\|\sigma_{x_1,J}\|_{sd}} \right| \lesssim J^\omega \|x_0 - x_1\|_{\ell^2}^{\omega'} |u| \lesssim \bar{J}_{\max}^{\omega - \frac{1}{2}} \|x_0 - x_1\|_{\ell^2}^{\omega'} F \,. \tag{52}$$

It now follows by compactness of $\mathcal{X}$ (Assumption 1(i)) that there exists finite positive constants D, v for which

$$\sup_Q N(\mathcal{F}_{n,J}, \|\cdot\|_{L^2(Q)}, \epsilon \|F\|_{L^2(Q)}) \leq D \left(\frac{\bar{J}_{\max}}{\epsilon} \right)^v \quad \forall\, \epsilon \in (0,1] \,,$$

where the supremum is over all discrete probability measures Q on $\mathcal{W} \times \mathbb{R}$ with $\|F\|_{L^2(Q)} > 0$. The result now follows (with a suitable modification of D and v) by noting that the covering numbers for $\mathcal{F}_n = \cup_{J \in \mathcal{J}_n} (\mathcal{F}_{n,J} \cup -\mathcal{F}_{n,J})$ are at most twice the sum of the covering numbers for each respective $\mathcal{F}_{n,J}$, and because $|\mathcal{J}_n| \lesssim \log(\bar{J}_{\max})$.

Part (ii): Fix any $J_0, J_1 \in \mathcal{J}_n$. Let $(x_o)_{o \in \mathcal{O}}$ be an ϵ-cover for $\mathcal{X}$. Choose $o_0, o_1 \in \mathcal{O}$ for which $\|x_i - x_{o_i}\|_{\ell_2} \le \epsilon$ for $i = 0, 1$. Then by (52), we have

$$\left| \frac{L_{J,x_0} b_w^{K(J)} u}{\|\sigma_{x_0,J}\|_{sd}} \frac{L_{J,x_1} b_w^{K(J)} u}{\|\sigma_{x_1,J}\|_{sd}} - \frac{L_{J,x_{o_0}} b_w^{K(J)} u}{\|\sigma_{x_{o_0},J}\|_{sd}} \frac{L_{J,x_{o_1}} b_w^{K(J)} u}{\|\sigma_{x_{o_1},J}\|_{sd}} \right| \lesssim \bar{J}_{\max}^{\omega - \frac{1}{2}} \epsilon^{\omega'} F^2 \,.$$

The result now follows similarly to part (i).

Parts (iii) and (iv): These follow now from parts (i) and (ii) by standard arguments. ■

For the following lemma, recall the process $\mathbb{Z}_n$ defined in (19). Define the pseudometric φ_n on $\mathcal{X} \times \mathcal{J}_n$ by $\varphi_n((x_0, J_0), (x_1, J_1))^2 = \mathbb{E}_{h_0}[(\mathbb{Z}_n(x_0, J_0) - \mathbb{Z}_n(x_1, J_1))^2]$.

Lemma C.10 *Let Assumptions 1, 2, 3(i) and 4(i) hold. Then: there exist constants $D, v > 0$ for which*

$$N(\mathcal{X} \times \mathcal{J}_n, \varphi_n, \epsilon) \le D \left(\frac{\bar{J}_{\max}}{\epsilon} \right)^v \quad \forall \;\; \epsilon > 0 \,,$$

uniformly for $h_0 \in \mathcal{H}$.

Proof of Lemma C.10. Fix $J \in \mathcal{J}_n$ and let $x_0, x_1 \in \mathcal{X}$. Then by similar arguments and notation to the proof of Lemma C.9(i), we obtain

$$\begin{aligned} \varphi_n((x_0, J), (x_1, J)) &= \left\| \frac{A_J[G_{\psi,J}^{-1/2} \psi_{x_0}^J]}{\|A_J[G_{\psi,J}^{-1/2} \psi_{x_0}^J]\|_{\ell^2}} - \frac{A_J[G_{\psi,J}^{-1/2} \psi_{x_1}^J]}{\|A_J[G_{\psi,J}^{-1/2} \psi_{x_1}^J]\|_{\ell^2}} \right\|_{\ell^2} \\ &\le 2 \left\| \frac{A_J([G_{\psi,J}^{-1/2} \psi_{x_0}^J] - [G_{\psi,J}^{-1/2} \psi_{x_1}^J])}{\|A_J[G_{\psi,J}^{-1/2} \psi_{x_0}^J]\|_{\ell^2} \vee \|A_J[G_{\psi,J}^{-1/2} \psi_{x_1}^J]\|_{\ell^2}} \right\|_{\ell^2} \lesssim J^{\omega - \frac{1}{2}} \|x_0 - x_1\|_{\ell^2}^{\varpi'} \,. \end{aligned}$$

The compactness of $\mathcal{X}$ (Assumption 1(i)) now implies that there are constants $D, v > 0$ for which

$$N(\mathcal{X} \times \{J\}, \varphi_n, \epsilon) \le D \left(\frac{J}{\epsilon} \right)^v \quad \forall \;\; \epsilon > 0 \,,$$

holds uniformly in h_0. The result follows (with possibly different D and v) by noting that the covering numbers of $\mathcal{X} \times \mathcal{J}_n$ are at most the sum of these individual covering numbers over $J \in \mathcal{J}_n$ and $|\mathcal{J}_n| \lesssim \log(\bar{J}_{\max})$. ■

Recall the definition of $\widehat{\mathbb{Z}}_n$ from (18) and $\mathbb{Z}_n$ from (19). Also let

$$\Upsilon_n = \sqrt{\frac{\bar{J}_{\max} \sqrt{\log \bar{J}_{\max}} (\log n)^2}{\sqrt{n}}} + \left(\frac{\sqrt{\bar{J}_{\max} \log \bar{J}_{\max}} (\log n)^2}{\sqrt{n}} \right)^{1/3} .$$

Enriching the original probability spaces as necessary, we also define on this space a (tight) Gaussian processes $\mathbb{B}_n \in \ell^\infty(\mathcal{X} \times \mathcal{J}_n)$ which is independent of the data $\mathcal{Z}_n$ and has identical covariance function to $\mathbb{Z}_n$. We also enrich the bootstrap probability space as necessary to support a random variable $\overline{B}_n^* =_d \overline{B}_n$

Lemma C.11 *Let Assumptions 1, 2, and 4(i) hold. Then: there exists finite positive constants D, D' such that:*

(i) There exists a random variable $\overline{B}_n =_d \sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} |\mathbb{B}_n(x,J)|$ which satisfies

$$\sup_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\left|\overline{B}_n - \sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} |\mathbb{Z}_n(x,J)|\right| > D\Upsilon_n\right) < D'\left(\frac{1}{\sqrt{\log(\bar{J}_{\max})}} + \frac{\log n}{n}\right).$$

(ii) There exists a sequence of sets Ω_n with $\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}(\mathcal{Z}^n \in \Omega_n) \to 1$ such that for every $\mathcal{Z}^n \in \Omega_n$, there exists a random variables $\overline{B}_n^* =_d \overline{B}_n$ which satisfies

$$\mathbb{P}^*\left(\left|\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} \left|\widehat{\mathbb{Z}}_n(x,J)\right| - \overline{B}_n^*\right| > D\left[\sqrt{\frac{\bar{J}_{\max}\sqrt{\log \bar{J}_{\max}}\log n}{\sqrt{n}}} + \frac{1}{\sqrt{u_n \log n}}\right]\right)$$
$$< D'\sqrt{\frac{\bar{J}_{\max}\sqrt{\log \bar{J}_{\max}}(\log n)^2 u_n}{\sqrt{n}}},$$

where $u_n \to \infty$ is a sequence which is $o(\log n)$.

Proof of Lemma C.11. Part (i): We very the conditions of Corollary 2.2 in Chernozhukov, Chetverikov, and Kato (2014b). Consider the set $\mathcal{F}_n$ defined in (51) and note that $F(w,u) = C\sqrt{\bar{J}_{\max}}|u|$ with sufficiently large C is an envelope for $\mathcal{F}_n$ (see Lemma C.9). We may deduce by similar arguments to the proof of Lemma C.8 that there is a constant $C' > 0$ for which

$$\|\mathbb{Z}_n(x,J)\|_{L^2(\mathbb{P}_{h_0})} \le \|\sigma_{x,J}\|_{sd}^{-1} \left\|[G_{\psi,J}^{-1/2}\psi_x^J]'(S_J^o)_l^-\right\|_{\ell^2} \|(\Omega_J^o)^{1/2}\|_{\ell^2} \le C'$$

for all $(x,J) \in \mathcal{X}\times\mathcal{T}$ and $h_0 \in \mathcal{H}$, in which case $\sup_{h_0\in\mathcal{H}}\sup_{g\in\mathcal{F}_n} \mathbb{E}_{h_0}[g(W,u)^2] \le C'$. Also note $|g(w,u)|^3 \le F(u)g(w,u)^2$, and so it follows by the preceding display and Assumption 2 that there is a constant $C'' > 0$ for which

$$\sup_{h_0\in\mathcal{H}}\sup_{g\in\mathcal{F}_n} \mathbb{E}_{h_0}\left[|g(W,u)|^3\right] \le C''\sqrt{\bar{J}_{\max}}.$$

As shown in Lemma C.9(i), there exists constants $C''', v > 0$ which satisfy

$$\sup_Q N(\mathcal{F}_n, \|\cdot\|_{L^2(Q)}, \epsilon\|F\|_{L^2(Q)}) \le C'''\left(\frac{\bar{J}_{\max}}{\epsilon}\right)^v, \quad \epsilon \in (0,1].$$

where the supremum is over all discrete probability measures Q on $\mathcal{W}\times\mathbb{R}$. Finally, note that by definition we have $\log(\bar{J}_{\max}) = O(\log n)$. Applying Corollary 2.2 of Chernozhukov et al. (2014b) with $b_n \lesssim \sqrt{\bar{J}_{\max}}$, $\sigma^2 \lesssim 1$, $K_n \lesssim \log n$, $\sup_{h_0\in\mathcal{H}} \mathbb{E}_{h_0}[|F(u)|^4]^{1/4} \lesssim \sqrt{\bar{J}_{\max}}$, and $\gamma_n^{-1} = \sqrt{\log \bar{J}_{\max}}$ yields the desired result.

Part (ii), step 1: To prove part (ii) we adapt arguments from the proof of Theorem A.2 in

Chernozhukov et al. (2014a). Recall the sets $\mathcal{F}_n$, $\mathcal{F}_n^2$, and $(\mathcal{F}_n - \mathcal{F}_n)^2$ defined in Lemma C.9. By part (i), we have $|g(w,u)| \le C^2 \bar{J}_{\max} u^2$ for $g \in \mathcal{F}_n^2$. Moreover, as $g(w,u)^2 \le F^2 |g(w,u)|$ for $g \in \mathcal{F}_n^2$, we may deduce by similar arguments to part (i) that there is a constant $C' > 0$ for which $\sup_{h_0 \in \mathcal{H}} \sup_{g \in \mathcal{F}_n^2} \mathbb{E}_{h_0}[g(W,u)^2] \le C' \bar{J}_{\max}$.

Let $M_n = (n/(\log \bar{J}_{\max}))^{1/4}$ and note that

$$
\begin{aligned}
\sup_{g \in \mathcal{F}_n^2} |(\mathbb{E}_n - \mathbb{E}_{h_0}) g| &\le \sup_{g \in \mathcal{F}_n^2} \left|(\mathbb{E}_n - \mathbb{E}_{h_0})\big(g \mathbb{1}_{|u| \le M_n}\big)\right| + \sup_{g \in \mathcal{F}_n^2} \left|(\mathbb{E}_n - \mathbb{E}_{h_0})\big(g \mathbb{1}_{|u| > M_n}\big)\right| \\
&=: T_{1,n} + T_{2,n}.
\end{aligned}
$$

Control of $T_{2,n}$: Note $g \mathbb{1}_{|u|>M_n} \le F^2 \mathbb{1}_{|u|>M_n}$ and so we have

$$
\sup_{g \in \mathcal{F}_n^2} \left|(\mathbb{E}_n - \mathbb{E}_{h_0})\big(g \mathbb{1}_{|u|>M_n}\big)\right| \le C^2 \bar{J}_{\max} \left(\mathbb{E}_n[u^2 \mathbb{1}_{|u|>M_n}] + \mathbb{E}_{h_0}[u^2 \mathbb{1}_{|u|>M_n}]\right) ,
$$

from which it follows by Markov's inequality that

$$
\sup_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(T_{2,n} > \frac{\bar{J}_{\max}\sqrt{\log \bar{J}_{\max}}}{\sqrt{n}} \right) \le \frac{2C^2\sqrt{n}}{\sqrt{\log \bar{J}_{\max}}} \frac{\sup_{h_0 \in \mathcal{H}} \mathbb{E}_{h_0}[u^4 \mathbb{1}_{|u|>M_n}]}{M_n^2} \to 0 \tag{53}
$$

by Assumption 2(i) and definition of M_n.

Control of $T_{1,n}$: Let $\mathcal{F}_n^2 \mathbb{1}_{|u|\le M_n} = \{h : h(w,u) = g(w,u)\mathbb{1}_{|u|\le M_n} , g \in \mathcal{F}_n^2\}$. It follows from Lemma C.9(ii) that $\mathcal{F}_n^2 \mathbb{1}_{|u|\le M_n}$ is $VC(F^2, \bar{J}_{\max})$. Standard algebraic manipulations then yield

$$
\int_0^{\sqrt{\sup_{g \in \mathcal{F}_n^2} \mathbb{E}_n[g^2 \mathbb{1}_{|u|\le M_n}]}} \sqrt{\log N(\mathcal{F}_n^2 \mathbb{1}_{|u|\le M_n}, \|\cdot\|_{L^2(\mathbb{P}_n)}, \epsilon)}\, d\epsilon \lesssim \|F^2\|_{L^2(\mathbb{P}_n)} \sqrt{\log \bar{J}_{\max}}
$$

uniformly for $h_0 \in \mathcal{H}$. By Jensen's inequality we also have

$$
\sup_{h_0 \in \mathcal{H}} \mathbb{E}_{h_0}\left[\|F^2\|_{L^2(\mathbb{P}_n)}\right] \le \sup_{h_0 \in \mathcal{H}} \mathbb{E}_{h_0}\left[F^4\right]^{1/2} \lesssim \bar{J}_{\max}
$$

by Assumption 2(i). It follows by the preceding two displays and Theorem 3.5.1 of Giné and Nickl (2016) that there is a constant $C''' > 0$ for which

$$
\mathbb{E}_{h_0}\left[\sup_{g \in \mathcal{F}_n^2} \left|\mathbb{E}_n[g \mathbb{1}_{|u|\le M_n}] - \mathbb{E}_{h_0}[g \mathbb{1}_{|u|\le M_n}]\right| \right] \le C''' \frac{\bar{J}_{\max}\sqrt{\log \bar{J}_{\max}}}{\sqrt{n}} .
$$

Theorem 3.3.9 of Giné and Nickl (2016) then implies that for some $C'''' > 0$ we have

$$
\sup_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(T_{1,n} \ge (1 + C''') \frac{\bar{J}_{\max}\sqrt{\log \bar{J}_{\max}}}{\sqrt{n}} \right) \le \exp\left(- C'''' \frac{\bar{J}_{\max}^2 (\log \bar{J}_{\max})/n}{(\bar{J}_{\max}^2/n + \bar{J}_{\max}/n)} \right) \to 0 . \tag{54}
$$

Combining (53) and (54), we have therefore shown that there is a constant $D > 0$ for which

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{g \in \mathcal{F}_n^2} |\mathbb{E}_n[g] - \mathbb{E}_{h_0}[g]| \leq D \frac{\bar{J}_{\max} \sqrt{\log \bar{J}_{\max}}}{\sqrt{n}} \right) \to 1 \,.$$

A analogous argument (increasing D if required) shows that

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{g \in (\mathcal{F}_n - \mathcal{F}_n)^2} |\mathbb{E}_n[g] - \mathbb{E}_{h_0}[g]| \leq D \frac{\bar{J}_{\max} \sqrt{\log \bar{J}_{\max}}}{\sqrt{n}} \right) \to 1 \,.$$

Part (ii), step 2: As every $g \in \mathcal{F}_n$ can be uniquely identified to some $(x, J) \in \mathcal{X} \times \mathcal{J}_n$, we shall view the bootstrap process $\widehat{\mathbb{Z}}_n$ from (18), empirical process $\mathbb{Z}_n$ from (19), and Gaussian process $\mathbb{B}_n$ in the statement of the lemma as processes over the index set $\mathcal{F}_n$ in what follows. By Lemma C.10, can choose a finite subset $T_n \subseteq \mathcal{F}_n$ for which

$$\sup_{f \in \mathcal{F}_n} \inf_{g \in T_n} \mathbb{E}_{h_0}\big[(f(W,u) - g(W,u))^2 \big] \leq D \bar{J}_{\max} \sqrt{(\log \bar{J}_{\max})/n}$$

and $\log(|T_n|) \lesssim \log(n \bar{J}_{\max}) \lesssim \log n$ both hold uniformly for $h_0 \in \mathcal{H}$.

Define

$$\widetilde{Z}_n = \max_{g \in T_n} |\widehat{\mathbb{Z}}_n(g)| \,, \qquad \overline{Z}_n = \sup_{g \in \mathcal{F}_n} |\widehat{\mathbb{Z}}_n(g)| \,, \qquad \widetilde{B}_n = \max_{g \in T_n} |\mathbb{B}_n(g)| \,, \qquad \overline{B}_n = \sup_{g \in \mathcal{F}_n} |\mathbb{B}_n(g)| \,.$$

By Theorem 3.2 in (Dudley, 2014), we extend $\mathbb{B}_n(\cdot)$ as a process over the linear span of $\mathcal{F}_n$ in a way that ensures it has linear sample paths. We similarly extend $\widehat{\mathbb{Z}}_n$ over the linear span of $\mathcal{F}_n$. It then follows that $|\widetilde{Z}_n - \overline{Z}_n| \leq \sup_{g \in \mathcal{E}_n} |\widehat{\mathbb{Z}}_n(g)|$ and $|\widetilde{B}_n - \overline{B}_n| \leq \sup_{g \in \mathcal{E}_n} |\mathbb{B}_n(g)|$, where

$$\mathcal{E}_n = \left\{ g = g_1 - g_2 : g_1, g_2 \in \mathcal{F}_n \,,\ \mathbb{E}_{h_0}\big[(g_1(W,u) - g_2(W,u))^2 \big] \leq D \bar{J}_{\max} \sqrt{(\log \bar{J}_{\max})/n} \right\}.$$

To control $\sup_{g \in \mathcal{E}_n} |\widehat{\mathbb{Z}}_n(g)|$ and $\sup_{g \in \mathcal{E}_n} |\mathbb{B}_n(g)|$, let $\sigma_n^2 = 2D \bar{J}_{\max} \sqrt{(\log \bar{J}_{\max})/n}$ and note that the definition of $\mathcal{E}_n$ and step 1 imply $\sup_{g \in \mathcal{E}_n} \left|\mathbb{E}_n[g^2]\right| \leq \sigma_n^2$ wpa1 $\mathcal{H}$-uniformly. It follows by Theorem 2.5.8 of Giné and Nickl (2016) that

$$\mathbb{P}^*\left(\sup_{g \in \mathcal{E}_n} |\widehat{\mathbb{Z}}_n(g)| \leq \mathbb{E}^*\left[\sup_{g \in \mathcal{E}_n} |\widehat{\mathbb{Z}}_n(g)| \right] + \sqrt{2\sigma_n^2 \log n} \right) \geq 1 - \frac{1}{n} \,,$$

wpa1 $\mathcal{H}$-uniformly. As $|g| \leq 2F$ for each $h \in \mathcal{E}_n$, it follows by Theorem 2.3.6 of Giné and Nickl

(2016) and Lemma C.9 that for any $h_0 \in \mathcal{H}$,

$$\mathbb{E}^*\left[\sup_{g\in\mathcal{E}_n}\left|\widehat{\mathbb{Z}}_n(g)\right|\right] \lesssim 2\|F\|_{L^2(\mathbb{P}_n)}\int_0^{\sigma_n/(2\|F\|_{L^2(\mathbb{P}_n)})}\sqrt{\log N(\mathcal{E}_n, \|\cdot\|_{L^2(\mathbb{P}_n)}, 2\epsilon\|F\|_{L^2(\mathbb{P}_n)})}d\epsilon$$
$$\lesssim \sigma_n\sqrt{\log\left(\frac{\bar{J}_{\max}\|F\|_{L^2(\mathbb{P}_n)}}{\sigma_n}\right)},$$

wpa1 $\mathcal{H}$-uniformly. The definitions of F, σ_n, $\bar{J}_{\max}$ and the fact that $\sup_{h_0\in\mathcal{H}}\mathbb{E}_{h_0}[|u|^2] < \infty$ imply that

$$\inf_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\Big(\log\left(\bar{J}_{\max}\|F\|_{L^2(\mathbb{P}_n)}/\sigma_n\right) \le 2\log n\Big) \to 1\,.$$

Combining the preceding three displays, we see that there is a finite positive constant C'' for which

$$\mathbb{P}^*\left(\sup_{g\in\mathcal{E}_n}\left|\widehat{\mathbb{Z}}_n(g)\right| \le C''\sigma_n\sqrt{\log n}\right) \ge 1 - \frac{1}{n}\,.$$

holds wpa1 $\mathcal{H}$-uniformly.

We now prove an analogous bound for the Gaussian process $\mathbb{B}_n$. By Theorem 2.5.8 of Giné and Nickl (2016) and definition of σ_n, we have

$$\inf_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\sup_{g\in\mathcal{E}_n}|\mathbb{B}_n(g)| \le \mathbb{E}_{h_0}\left[\sup_{g\in\mathcal{E}_n}|\mathbb{B}_n(g)|\right] + \sqrt{\sigma_n^2\log n}\right) \ge 1 - \frac{1}{n}\,. \tag{55}$$

Equip the linear span of $\mathcal{F}_n$ with the pseudometric $d_n(g,g') = \|\mathbb{B}_n(g) - \mathbb{B}_n(g')\|_{L^2(\mathbb{P}_{h_0})}$ and note $d_n(g,g') \le \sqrt{2}\sigma_n$ hold for $g, g' \in \mathcal{E}_n$ by construction. Also note that $N(\mathcal{E}_n, d_n, \epsilon) \le N(\mathcal{F}_n, d_n, \epsilon/2)^2$. As $\mathbb{B}_n(\cdot)$ (restricted $\mathcal{F}_n$) has the same covariance kernel as $\mathbb{Z}_n$ from (19) on $\mathcal{X}\times\mathcal{J}_n$, we have $N(\mathcal{F}_n, d_n, \epsilon) \lesssim (\bar{J}_{\max}/\epsilon)^v$ for some positive constant v by Lemma C.10. Now by Theorem 2.3.6 of Giné and Nickl (2016), we may deduce

$$\sup_{h_0\in\mathcal{H}}\mathbb{E}_{h_0}\left[\sup_{g\in\mathcal{E}_n}|\mathbb{B}_n(g)|\right] \lesssim \int_0^{\sqrt{2}\sigma_n}\sqrt{\log N(\mathcal{E}_n, d_n, \epsilon)}\,d\epsilon \lesssim \sigma_n\sqrt{\log n}\,,$$

where we used the fact that $\log(\bar{J}_{\max}/\sigma_n) \lesssim \log n$. Combining this with (55) yields

$$\inf_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\sup_{g\in\mathcal{E}_n}|\mathbb{B}_n(g)| \le C'''\sigma_n\sqrt{\log n}\right) \ge 1 - \frac{1}{n}\,,$$

for some finite positive constant C'''.

Part (ii), step 3: By part (ii), steps 1 and 2, the following three inequalities hold wpa1

$\mathcal{H}$-uniformly:

$$\mathbb{P}^*\left(\left|\widetilde{Z}_n - \overline{Z}_n\right| \le \zeta_n\right) \ge 1 - \frac{1}{n}\,, \quad \mathbb{P}_{h_0}\left(\left|\widetilde{B}_n - \overline{B}_n\right| \le \zeta_n\right) \ge 1 - \frac{1}{n}\,, \quad \sup_{g \in \mathcal{F}_n^2} |\mathbb{E}_n[g] - \mathbb{E}_{h_0}[g]| \lesssim \sigma_n^2,$$

where $\zeta_n = D'\sigma_n\sqrt{\log n}$ for some finite constant $D' > 0$. We work for the remainder of the proof of this part on the sequence of events upon which the above three inequalities hold.

Let $u_n \to \infty$ be a sequence which is $o(\log n)$. Set $\varrho_n = (u_n \log n)^{-1/2}$, noting that $\varrho_n^{-1} = o(\log n)$. Fix any Borel subset $A \subseteq \mathbb{R}$. Let $A^\epsilon = \{x \in \mathbb{R} : \inf_{y \in A} d(x,y) \le \epsilon\}$ denote the ϵ-enlargement. By Lemma 4.2 of Chernozhukov et al. (2014b) (with the choices $\beta = \delta^{-1}\sqrt{\log n}$ and $\delta = \varrho_n$ in the notation of Chernozhukov et al. (2014b)), there exists $q_n \in C^\infty(\mathbb{R})$ which satisfies

$$(1-\epsilon_n)\mathbb{1}_{B^{\zeta_n}} \le q_n(t) \le \epsilon_n + (1-\epsilon_n)\mathbb{1}_{B^{\zeta_n + 3\varrho_n}}\,, \quad \epsilon_n = \sqrt{\frac{e \log n}{n}}\,,$$

where the firs three derivatives of q_n satisfy $\|q_n'\|_\infty \le \varrho_n^{-1}$, $\|q_n''\|_\infty \lesssim \varrho_n^{-2}\sqrt{\log n}$, and $\|q_n'''\|_\infty \lesssim \varrho_n^{-3}\sqrt{\log n}$. Also note that

$$\max_{g_1,g_2 \in T_n} |\mathbb{E}_n[g_1 g_2] - \mathbb{E}_{h_0}[g_1 g_2]| \le \sup_{g \in \mathcal{F}_n^2} |\mathbb{E}_n[g] - \mathbb{E}_{h_0}[g]| \lesssim \sigma_n^2\,,$$

uniformly for $h_0 \in \mathcal{H}$. Hence by Comment 1 of Chernozhukov, Chetverikov, and Kato (2015), noting that $\log T_n \lesssim \log n$ and that $\varrho_n^{-1}\sigma_n \lesssim 1$ by definition of $\bar{J}_{\max}$, we have

$$\left|\mathbb{E}^*\left[q_n(\widetilde{Z}_n)\right] - \mathbb{E}_{h_0}\left[q_n(\widetilde{B}_n)\right]\right| \lesssim \varrho_n^{-2}\sigma_n^2\sqrt{\log n} + \varrho_n^{-1}\sigma_n\sqrt{\log n} \lesssim \varrho_n^{-1}\sigma_n\sqrt{\log n}\,,$$

uniformly for $h_0 \in \mathcal{H}$. We then have for finite constants $C, C' > 0$ that

$$\begin{aligned}
\mathbb{P}^*\left(\overline{Z}_n \in A\right) &\le \mathbb{P}^*\left(\widetilde{Z}_n \in A^{\zeta_n}\right) + n^{-1} \\
&\le (1-\epsilon_n)^{-1}\mathbb{E}^*\left[q_n(\widetilde{Z}_n)\right] + n^{-1} \\
&\le (1-\epsilon_n)^{-1}\mathbb{E}_{h_0}\left[q_n(\widetilde{B}_n)\right] + n^{-1} + C\varrho_n^{-1}\sigma_n\sqrt{\log n} \\
&\le (1-\epsilon_n)^{-1}\epsilon_n + \mathbb{P}_{h_0}\left(\widetilde{B}_n \in A^{\zeta_n + 3\varrho_n}\right) + n^{-1} + C\varrho_n^{-1}\sigma_n\sqrt{\log n} \\
&\le (1-\epsilon_n)^{-1}\epsilon_n + \mathbb{P}_{h_0}\left(\overline{B}_n \in A^{2\zeta_n + 3\varrho_n}\right) + 2n^{-1} + C\varrho_n^{-1}\sigma_n\sqrt{\log n} \\
&\le \mathbb{P}_{h_0}\left(\overline{B}_n \in A^{2\zeta_n + 3\varrho_n}\right) + C'\varrho_n^{-1}\sigma_n\sqrt{\log n}\,,
\end{aligned}$$

uniformly for $h_0 \in \mathcal{H}$. It now follows from the preceding inequality by Lemma 4.1 of Chernozhukov et al. (2014b) that there exists a random variable $\overline{B}_n^* =_d \overline{B}_n$ which satisfies

$$\mathbb{P}^*\left(\left|\overline{Z}_n - \overline{B}_n^*\right| > 2\zeta_n + 3\varrho_n\right) < C'\varrho_n^{-1}\sigma_n\sqrt{\log n}\,,$$

uniformly for $h_0 \in \mathcal{H}$. ∎

Proof of Theorem B.2. Part (i), Step 1: Recalling the definition of $\mathbb{Z}_n$ from (19), note that

$$\begin{aligned}
&\left|\sqrt{n}\frac{\hat{h}_J(x)-\tilde{h}_J(x)}{\|\hat{\sigma}_{x,J}\|_{sd}}-\mathbb{Z}_n(x,J)\right| \\
&\leq \left|\sqrt{n}\frac{\hat{h}_J(x)-\tilde{h}_J(x)}{\|\sigma_{x,J}\|_{sd}}-\mathbb{Z}_n(x,J)\right| + \left|\frac{\|\sigma_{x,J}\|_{sd}}{\|\hat{\sigma}_{x,J}\|_{sd}}-1\right|\left|\sqrt{n}\frac{\hat{h}_J(x)-\tilde{h}_J(x)}{\|\sigma_{x,J}\|_{sd}}\right| \quad =: T_{1,x,J}+T_{2,x,J}.
\end{aligned}$$

Control of $T_{1,x,J}$: Expanding the definition of $\hat{h}_J(x)-\tilde{h}_J(x)$ yields

$$\begin{aligned}
T_{1,x,J} &= \left|\frac{[G_{\psi,J}^{-1/2}\psi_x^J]'}{\|\sigma_{x,J}\|_{sd}}(S_J^o)_l^- S_J^o\big(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2}-(S_J^o)_l^-\big)\big(G_{b,J}^{-1/2}B_J'u/\sqrt{n}\big)\right| \\
&\leq \left\|\frac{[G_{\psi,J}^{-1/2}\psi_x^J]'}{\|\sigma_{x,J}\|_{sd}}(S_J^o)_l^-\right\|_{\ell^2}\left\|S_J^o\big(((\widehat{G}_{b,J}^o)^{-1/2}\widehat{S}_J^o)_l^-(\widehat{G}_{b,J}^o)^{-1/2}-(S_J^o)_l^-\big)\right\|_{\ell^2}\left\|G_{b,J}^{-1/2}B_J'u/\sqrt{n}\right\|_{\ell^2}.
\end{aligned}$$

Assumption 2 implies that $\|\sigma_{x,J}\|_{sd}\asymp\|\sigma_{x,J}\|=\|[G_{\psi,J}^{-1/2}\psi_x^J]'(S_J^o)_l^-\|_{\ell^2}$ uniformly for $(x,J)\in\mathcal{X}\times\mathcal{T}$. It then follows by Lemmas C.2 and C.4 that there is a constant $C>0$ for which

$$\inf_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}T_{1,x,J}\leq C\frac{\tau_{\bar{J}_{\max}}\bar{J}_{\max}\sqrt{\log\bar{J}_{\max}}}{n}\right)\to 1\,. \tag{56}$$

Control of $T_{2,x,J}$: By Lemma B.4 and the fact that $|1-a|\leq|1-a^2|$ for $a\geq 0$, we have

$$\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}T_{2,x,J}\leq C_{B.4}\delta_n\times\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}\left|\sqrt{n}\frac{\hat{h}_J(x)-\tilde{h}_J(x)}{\|\hat{\sigma}_{x,J}\|_{sd}}\right|$$

wpa1 $\mathcal{H}$-uniformly. Recall $\overline{B}_n$ from Lemma C.11. By (56) and definition of $\bar{J}_{\max}$, we have

$$\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}T_{2,x,J}\leq C_{B.4}\delta_n\times\left(1+\left|\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}|\mathbb{Z}_n(x,J)|-\overline{B}_n\right|+\overline{B}_n\right)$$

wpa1 $\mathcal{H}$-uniformly. By Lemma C.11(i) it suffices to control $\overline{B}_n =_d \sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n}|\mathbb{B}_n(x,J)|$, where $\mathbb{B}_n$ is a Gaussian process with the same covariance function as $\mathbb{Z}_n$. We may deduce by similar arguments to the proof of Lemma C.8 that there is a constant $C>0$ for which

$$\|\mathbb{B}_n(x,J)\|_{L^2(\mathbb{P}_{h_0})}=\|\mathbb{Z}_n(x,J)\|_{L^2(\mathbb{P}_{h_0})}\leq\|\sigma_{x,J}\|_{sd}^{-1}\left\|[G_{\psi,J}^{-1/2}\psi_x^J]'(S_J^o)_l^-\right\|_{\ell^2}\|(\Omega_J^o)^{1/2}\|_{\ell^2}\leq C$$

for all $(x,J)\in\mathcal{X}\times\mathcal{T}$ and $h_0\in\mathcal{H}$. Applying Theorem 2.5.8 of Giné and Nickl (2016) yields

$$\sup_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\overline{B}_n>\mathbb{E}_{h_0}[\overline{B}_n]+\sqrt{2C^2C'\log\bar{J}_{\max}}\right)\leq\bar{J}_{\max}^{-C'}\to 0\,,$$

for any constant $C' > 0$. The remaining term $\mathbb{E}_{h_0}[\overline{B}_n]$ may be shown to be $O(\sqrt{\bar{J}_{\max}})$ uniformly for $h_0 \in \mathcal{H}$ using Theorem 2.3.6 of Giné and Nickl (2016) and similar arguments to the proof of Lemma C.8. It follows that

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} T_{2,x,J} \leq C'' \delta_n \sqrt{\log \bar{J}_{\max}} \right) \to 1$$

for some constant $C'' > 0$, and hence that

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} \left| \sqrt{n}\frac{\hat{h}_J(x) - \tilde{h}_J(x)}{\|\hat{\sigma}_{x,J}\|_{sd}} - \mathbb{Z}_n(x,J) \right| \leq C''' \delta_n \sqrt{\log \bar{J}_{\max}} \right) \to 1\,,$$

for some constant $C''' > 0$. Note $\delta_n \sqrt{\log \bar{J}_{\max}} \to 0$ by definition of $\bar{J}_{\max}$ and because $J_{\min} \asymp (\log \bar{J}_{\max})^2$ and $2\underline{p} > d$.

Part (i), Step 2: By step 1 and Lemma C.11(i), we have

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\left| \sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} \left| \sqrt{n}\frac{\hat{h}_J(x) - \tilde{h}_J(x)}{\|\hat{\sigma}_{x,J}\|_{sd}} \right| - \overline{B}_n \right| \leq C\chi_n \right) \to 1\,,$$

for some constant $C > 0$, where $\chi_n = \Upsilon_n + \delta_n \sqrt{\log \bar{J}_{\max}}$ with Υ_n from Lemma C.11(i). Lemma 2.3 in Chernozhukov et al. (2014b) (noting $\sup_{h_0\in\mathcal{H}} \mathbb{E}_{h_0}[\overline{B}_n] \lesssim \sqrt{\log \bar{J}_{\max}}$ from step 1) yields

$$\begin{aligned} \sup_{h_0\in\mathcal{H}} \sup_{s\in\mathbb{R}} \left| \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} \left| \sqrt{n}\frac{\hat{h}_J(x) - \tilde{h}_J(x)}{\|\hat{\sigma}_{x,J}\|_{sd}} \right| \leq s \right) - \mathbb{P}_{h_0}(\overline{B}_n \leq s) \right| & \\ \lesssim \chi_n \left(\sqrt{\log \bar{J}_{\max}} + \log(\chi_n^{-1}) \right) + o(1)\,. & \quad (57) \end{aligned}$$

We now apply a similar argument to the bootstrapped process. By the proof of Lemma B.6 we have $\mathbb{E}^*[\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} |\widehat{\mathbb{Z}}_n(x,J)|] \lesssim \sqrt{\log \bar{J}_{\max}}$ wpa1 $\mathcal{H}$-uniformly. Then by Lemma C.11(ii) and Lemma 2.3 in Chernozhukov et al. (2014b), we have for some $C'' > 0$ that

$$\begin{aligned} \inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{s\in\mathbb{R}} \left| \mathbb{P}^*\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} \left| \widehat{\mathbb{Z}}_n(x,J) \right| \leq s \right) - \mathbb{P}_{h_0}(\overline{B}_n \leq s) \right| \right. & \\ \left. \leq C'' \left(\beta_n \left(\sqrt{\log \bar{J}_{\max}} + \log(\beta_n^{-1}) \right) + \gamma_n \right) \right) \to 1\,, & \quad (58) \end{aligned}$$

where

$$\beta_n = \sqrt{\frac{\bar{J}_{\max}\sqrt{\log \bar{J}_{\max}}}{\sqrt{n}} \log n + \frac{1}{\sqrt{u_n \log n}}}\,, \qquad \gamma_n = \sqrt{\frac{\bar{J}_{\max}\sqrt{\log \bar{J}_{\max}}}{\sqrt{n}} (\log n)^2 u_n}\,,$$

and $u_n \to \infty$ is a sequence which is $o(\log n)$. Finally, by an anti-concentration inequality (Chernozhukov et al., 2014a, Corollary 2.1) wpa1 $\mathcal{H}$-uniformly, for all $a > 0$ we have

$$\sup_{s \in \mathbb{R}} \left| \mathbb{P}^* \left(\left| \sup_{(x,J) \in \mathcal{X} \times \mathcal{J}_n} |\widehat{\mathbb{Z}}_n(x,J)| - s \right| \leq a\delta_n \sqrt{\log \bar{J}_{\max}} \right) \right| \lesssim a\delta_n \log \bar{J}_{\max} \,.$$

In view of Lemma C.8(i), we therefore have for some $C''' > 0$ that

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(\sup_{s \in \mathbb{R}} \left| \mathbb{P}^* \left(\sup_{(x,J) \in \mathcal{X} \times \mathcal{J}_n} |\widehat{\mathbb{Z}}_n(x,J)| \leq s \right) - \mathbb{P}^* \left(\sup_{(x,J) \in \mathcal{X} \times \mathcal{J}_n} |\mathbb{Z}^*_n(x,J)| \leq s \right) \right| \right.$$
$$\left. \leq C''' \delta_n \log \bar{J}_{\max} \right) \to 1 \quad (59)$$

The desired result now follows from combining (57), (58), and (59) and observing each of $\chi_n \sqrt{\log \bar{J}_{\max}}$, $\beta_n \sqrt{\log \bar{J}_{\max}}$, and γ_n are $o(1)$ because $J_{\min} \asymp (\log \bar{J}_{\max})^2$ and $2\underline{p} > d$.

Part (ii): Follows by a similar argument to part (i). ■

C.2 Supplemental Results: UCBs for Derivatives

Here we present supplemental results for the proofs of Theorems 4.4 and 4.5. Throughout this subsection, for any fixed $R > 0$, let $\bar{J}_{\max} \equiv \bar{J}_{\max}(R)$. Also let $J_{\min} \to \infty$ as $n \to \infty$ with $J_{\min} \leq \bar{J}_{\max}$. Define $\mathcal{J}_n = \{J \in \mathcal{T} : J_{\min} \leq J \leq \bar{J}_{\max}\}$. Also recall δ_n from (25).

Lemma C.12 *Let Assumptions 1-3 hold. Then: there is a universal constant $C_{C.12} > 0$ such that*

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(\sup_{(x,J) \in \mathcal{X} \times \mathcal{J}_n} \left| \frac{\|\hat{\sigma}^a_{x,J}\|^2_{sd}}{\|\sigma^a_{x,J}\|^2_{sd}} - 1 \right| \leq C_{B.4} \delta_n \right) \to 1 \,.$$

Proof of Lemma C.12. The proof follows by identical arguments to the proof of Lemma B.4, replacing $\hat{\gamma}_{x,J}$ and $\gamma_{x,J}$ with $\hat{\gamma}^a_{x,J} := (\widehat{G}^o_{b,J})^{-1} \widehat{S}^o_J [(\widehat{S}^o_J)'(\widehat{G}^o_{b,J})^{-1} \widehat{S}^o_J]^{-1} G^{-1/2}_{\psi,J} (\partial^a \psi^J_x)$ and $\gamma^a_{x,J} := S^o_J [(S^o_J)' S^o_J]^{-1} G^{-1/2}_{\psi,J} (\partial^a \psi^J_x)$, respectively. ■

Lemma C.13 *Let Assumptions 1-4 hold. For a given $\alpha \in (0,1)$, let $z^{a*}_{1-\alpha}$ denote the $1-\alpha$ quantile of $\sup_{(x,J) \in \mathcal{X} \times \hat{\mathcal{J}}} |\mathbb{Z}^{a*}_n(x,J)|$. Then: with $\bar{J}_{\max}(R)$ as defined in (24) for any $R > 0$,there exists a constant $C_{C.13} > 0$ for which*

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(z^{a*}_{1-\alpha} \leq C_{C.13} \sqrt{\log \bar{J}_{\max}(R)} \right) \to 1 \,.$$

Proof of Lemma C.13. Follows by identical arguments to the proof of Lemma B.7, noting that $\partial^a \psi^J_x$ is Hölder continuous provided $\underline{p} > |a|$. ■

Lemma C.14 *Let Assumptions 1-4 hold and let $J_{\min} \asymp (\log \bar{J}_{\max})^2$. Then: there exists a sequence $\gamma_n \downarrow 0$ for which*

$$\sup_{s\in\mathbb{R}} \left| \mathbb{P}_{h_0}\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} \left| \sqrt{n}\frac{\partial^a \hat{h}_J(x) - \partial^a \tilde{h}_J(x)}{\|\hat{\sigma}^a_{x,J}\|_{sd}} \right| \le s \right) - \mathbb{P}^*\left(\sup_{(x,J)\in\mathcal{X}\times\mathcal{J}_n} |\mathbb{Z}^{a*}_n(x,J)| \le s \right) \right| \le \gamma_n$$

holds wpa1 $\mathcal{H}$-uniformly.

Proof of Lemma C.14. The proof follows by similar arguments to the proof of Theorem B.2(i), replacing $\mathcal{F}_n$ with

$$\mathcal{F}^a_n := \left\{ g : g(w,u) = \pm \|\sigma^a_{x,J}\|^{-1}_{sd} L^a_{J,x} b^{K(J)}_w u \quad \text{for some} \quad (x,J) \in \mathcal{X}\times\mathcal{J}_n \right\} \cup \{0\},$$

where $L^a_{J,x} = (\partial^a \psi^J_x)'[S'_J G^{-1}_{b,J} S_J]^{-1} S'_J G^{-1}_{b,J}$. As $\sup_{x\in\mathcal{X}} \|G^{-1/2}_{\psi,J}\partial^a \psi^J_x\| \lesssim J^{1/2+|a|/d}$, it follows by Assumption 4(iii) that $F(w,u) = C\sqrt{\bar{J}_{\max}}|u|$ is an envelope for $\mathcal{F}^a_n$ for sufficiently large C. The VC properties of $\mathcal{F}^a_n$ and related classes then follow similarly to Lemma C.9 by Hölder continuity of $\partial^a \psi^J_x$. The empirical process

$$\mathbb{Z}^a_n(x,J) := \frac{1}{\|\sigma^a_{x,J}\|_{sd}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n L^a_{J,x} b^{K(J)}_{W_i} u_i \right)$$

induces a pseudometric over $\mathcal{X}\times\mathcal{J}_n$ whose covering number behaves as in Lemma C.10, again by Hölder continuity of $\partial^a \psi^J_x$. We may then establish a result analogous to Lemma C.11 for Gaussian approximation to the supremum of the process $\mathbb{Z}^a_n$ and its bootstrap analogue

$$\widehat{\mathbb{Z}}^a_n(x,J) = \frac{1}{\|\sigma^a_{x,J}\|_{sd}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n L^a_{J,x} b^{K(J)}_{W_i} u_i \varpi_i \right)$$

over $\mathcal{X}\times\mathcal{J}_n$. The remainder of the proof now similarly to the proof of Theorem B.2(i). ■

C.3 Proofs of Theorems 4.4 and 4.5 on UCBs for Derivatives

Proof of Theorem 4.4. The proof follows similar arguments to the proof of Theorem 4.2. Here we state the necessary modifications.

Part (i), step 1: Identical to part (i), step 1 of the proof of Theorem 4.2.

Part (i), step 2: Note that by Theorem B.1 and a similar argument to the proof of Corollary 3.1 of Chen and Christensen (2018), we have

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\|\partial^a \tilde{h}_J - \partial^a h_0\|_\infty \le C_6 J^{(|a|-p)/d} \quad \forall\, J \in [1, \bar{J}_{\max}] \cap \mathcal{T} \right) \to 1$$

for some constant $C_6 > 0$. Moreover, by Lemma C.12 and Assumption 4(iii) there is a constant $C_7 > 0$ for which

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J) \in \mathcal{X} \times \hat{\mathcal{J}}} \frac{\tau_J J^{1/2+|a|/d}}{\|\hat{\sigma}^a_{x,J}\|_{sd}} \le C_7 \right) \to 1\,.$$

It now follows by (34) that

$$\frac{|\partial^a \tilde{h}_{\hat{J}}(x) - \partial^a h_0(x)|}{\|\hat{\sigma}^a_{x,\hat{J}}\|_{sd}} \le C_6 C_7 \frac{\hat{J}^{-p/d}}{\tau_{\hat{J}} \sqrt{\hat{J}}} \le C_6 C_7 \beta^{-\overline{p}/d} 2^{\overline{p}} \frac{(2^d J_0(p,E))^{-p/d}}{\tau_{\lceil \beta J_0(p,E) \rceil} \sqrt{\beta J_0(p,E)}}\,,$$

wpa1 uniformly for $h_0 \in \mathcal{G}^p$ and $p \in [\underline{p}, \overline{p}]$ and $x \in \mathcal{X}$. The remainder of the proof of this part now follows by identical arguments to part (i), step 2 of the proof of Theorem 4.2, using Lemma C.14 in place of Theorem B.2(i).

Part (ii): By Lemmas B.6, C.12, and C.13 and Assumption 4(iii) we have

$$\sup_{x \in \mathcal{X}} |C^a_n(x,A)| \lesssim (1+A)\tau_{\hat{J}} \hat{J}^{1/2+|a|/d} \sqrt{(\log \bar{J}_{\max})/n}$$

wpa1 $\mathcal{H}$-uniformly. Then by display (34), with $J_0 = J_0(p,D)$ we have that

$$\begin{aligned} \sup_{x \in \mathcal{X}} |C^a_n(x,A)| &\lesssim (1+A)\tau_{J_0^+}(J_0^+)^{1/2+|a|/d} \sqrt{(\log \bar{J}_{\max})/n} \\ &\lesssim (1+A)\tau_{J_0} J_0^{|a|/d} \sqrt{(J_0 \log \bar{J}_{\max})/n} \lesssim (1+A)\frac{\sqrt{\log \bar{J}_{\max}}}{\theta^*(\hat{\alpha})} J_0^{(|a|-p)/d} \end{aligned}$$

holds wpa1 uniformly for $h_0 \in \mathcal{G}^p$ and $p \in [\underline{p}, \overline{p}]$, where the second inequality follows from the fact that the model is mildly ill-posed and the third is by definition (21). The result now follows by similar arguments to part (ii) of the proof of Theorem 4.2. ■

Proof of Theorem 4.5. The proof follows similar arguments to the proof of Theorem 4.3. Here we state the necessary modifications.

Part (i): By Lemma B.5, Theorem B.1, and similar arguments to the proof of Corollary 3.1 of Chen and Christensen (2018), there exists a constant $A_0 > 0$ for which

$$|\partial^a \hat{h}_{\tilde{J}}(x) - \partial^a h_0(x)| \le |\partial^a \hat{h}_{\tilde{J}}(x) - \partial^a \tilde{h}_{\tilde{J}}(x)| + A_0 \tilde{J}^{(|a|-\underline{p})/d}$$

holds for all $x \in \mathcal{X}$ wpa1 $\mathcal{H}$-uniformly. Then for any $A \ge A_0$, we have

$$\inf_{h_0 \in \mathcal{G}} \mathbb{P}_{h_0}\big(\partial^a h_0(x) \in C^a_n(x,A) \;\; \forall\, x \in \mathcal{X}\big) \ge \inf_{h_0 \in \mathcal{G}} \mathbb{P}_{h_0}\left(\sup_{x \in \mathcal{X}} \left| \frac{\partial^a \hat{h}_{\tilde{J}}(x) - \partial^a \tilde{h}_{\tilde{J}}(x)}{\|\hat{\sigma}^a_{x,\tilde{J}}\|_{sd}} \right| \le z^{a*}_{1-\alpha} \right) + o(1)\,.$$

The remainder of the proof now follows similarly to the proof of Theorem 4.3, using Lemma C.14 in place of Theorem B.2(i).

Part (ii): By Lemmas B.2, B.6, C.12, and C.13 and Assumption 4(iii), there exist constants $C, C' > 0$ for which

$$\sup_{x \in \mathcal{X}} |C_n^a(x, A)| \le C(1 + A)\tau_{\tilde{J}} \tilde{J}^{1/2+|a|/d} \sqrt{\log(\bar{J}^*_{\max}(R_2))/n} + A\tilde{J}^{(|a|-\underline{p})/d}$$
$$\le C'(1 + A)(J^*_{\max}(R_2))^{(|a|-p)/d} + A\tilde{J}^{(|a|-\underline{p})/d}$$

holds wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$. The remainder of the proof now follows similarly to the proof of Theorem 4.3. ■

References

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.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014a). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics 42*(5), 1787–1818.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014b). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics 42*(4), 1564–1597.

Chernozhukov, V., D. Chetverikov, and K. Kato (2015). Comparison and anti-concentration bounds for maxima of gaussian random vectors. *Probability Theory and Related Fields 162*(1-2), 47–70.

DasGupta, A. (2008). *Asymptotic theory of statistics and probability.* Springer.

Dudley, R. M. (2014). *Uniform central limit theorems.* Cambridge University Press.

Giné, E. and R. Nickl (2016). *Mathematical foundations of infinite-dimensional statistical models.* Cambridge University Press.

Tropp, J. A. (2012). User-friendly tail bounds for sums of random matrices. *Foundations of computational mathematics 12*(4), 389–434.