Online Appendix to "Adaptive Estimation and Uniform Confidence Bands for Nonparametric Structural Functions and Elasticities"

Xiaohong Chen Timothy Christensen Sid Kankanala

C Additional Simulation: Engel Curves

In this appendix we present additional simulation results for estimating a nonparametric structural function in an empirically calibrated Engel curve setting. The design is based on the British Family Expenditure Survey data used in Blundell et al. (2007). We draw household expenditure X and household income W from a bivariate normal density with correlation $\rho = 0.52$, which is the sample correlation of the expenditure and income data used in Blundell et al. (2007). We then transform X and W to have Uniform$[0,1]$ marginals using their respective inverse marginal CDFs. As a consequence, X and W are linked via a Gaussian copula and the design is severely ill-posed.[16] We then set $h_0(x) = \Phi(5x - 2.5)$ and set $u = h_0(X) - \mathbb{E}[h_0(X)|W] + v$ for $v \sim N(0, 0.01)$. The implementation is the same as the other Monte Carlos from Section 5. For each simulated data set we compute our data-driven estimator $\hat{h}_{\tilde{J}}$ and UCBs from (16). We compare these with estimators and UCBs using deterministic choices of sieve dimensions for $J = 4$, 5, 7, and 11 (the first few dimensions over which our procedure searches). We again use a cubic B-spline basis to approximate h_0 and a quartic B-spline basis for the reduced form.

Turning first to the simulation results presented in Table 9, we see that the average sup-norm loss of our data-driven estimator is similar to that of an estimator $\hat{h}_J$ for deterministic J with $J = 4$ and several multiples smaller than that with $J = 5$, 7, or 11. This is to be expected, as the design is severely ill-posed and the true function is very smooth, so a very small choice of J is appropriate. Of course, in practice the researcher does not know the degree of ill-posedness or the degree of smoothness of the structural function.

The second panel of Table 9 shows our data-driven UCBs have valid, albeit conservative, coverage across all sample sizes. By contrast, undersmoothed UCBs with $J = 4$ and $J = 5$ under-cover for $n = 2500$, 5000, and 10000. Undersmoothed UCBs with $J = 7$ have valid but conservative coverage, but these are 40% (with $n = 1250$) to 250% (with $n = 10000$) wider than our data-driven UCBs. It is important to note that although the

[16] This follows from, e.g., Beare (2010), equation (3.3).

Table 9: Simulation Results for the Engel Curve Design.

	Data-driven		Deterministic							
			$J = 4$		$J = 5$		$J = 7$		$J = 11$	
	Sup-norm Loss									
n	mean	med.	mean	med.	mean	med.	mean	med.	mean	med.
1250	0.221	0.183	0.218	0.180	0.337	0.287	0.450	0.400	0.558	0.488
2500	0.167	0.139	0.164	0.138	0.285	0.240	0.417	0.369	0.526	0.468
5000	0.115	0.094	0.113	0.094	0.233	0.197	0.361	0.318	0.484	0.432
10000	0.083	0.068	0.080	0.068	0.173	0.148	0.322	0.299	0.448	0.414
	UCB Coverage									
	90%	95%	90%	95%	90%	95%	90%	95%	90%	95%
1250	0.998	0.999	0.917	0.961	0.903	0.952	0.934	0.972	0.916	0.968
2500	0.998	0.999	0.868	0.931	0.867	0.943	0.950	0.980	0.941	0.982
5000	0.998	0.999	0.833	0.896	0.884	0.939	0.967	0.989	0.968	0.987
10000	0.991	0.994	0.700	0.826	0.826	0.904	0.956	0.988	0.964	0.992
			95% UCB Relative Width (Deterministic/Data-driven)							
			mean	med.	mean	med.	mean	med.	mean	med.
1250			0.653	0.658	1.056	0.978	1.431	1.351	1.798	1.718
2500			0.655	0.661	1.213	1.120	1.797	1.692	2.372	2.270
5000			0.656	0.661	1.458	1.331	2.219	2.107	3.082	2.991
10000			0.658	0.664	1.577	1.478	2.712	2.574	3.962	3.792

design is severely ill-posed, we are reporting coverage of our UCBs (16). In each simulated data set we have $\hat{J} = \tilde{J}$ irrespective of the sample size n, so the critical value is effectively $z^*_{1-\alpha} + \hat{A}\theta^*_{1-\hat{\alpha}}$. While Theorem 4.2 does not formally establish coverage guarantees of this band in the severely ill-posed case, these simulation results show that the band nevertheless has good coverage in this empirically relevant design.

Figure 7 presents plots of data-driven estimates and UCBs for h_0 and its derivative for a sample of size 2500, alongside deterministic-J estimates and UCBs. In this sample, $\tilde{J} = 4$ and our data-driven UCBs contain the true structural function. As with the other simulations, the data-driven bands are narrower and more accurately convey the shape of h_0 than the $J = 7$ bands, which are much more wiggly. Our bands are also slightly narrower than the $J = 5$ bands. Panel (d) of Figure 3 also presents data-driven estimates and UCBs for the conditional mean of Y given X. Evidently, the true structural function falls outside the UCBs for the conditional mean function over almost all of the support of X, again highlighting the importance of estimating h_0 using IV methods in this design.

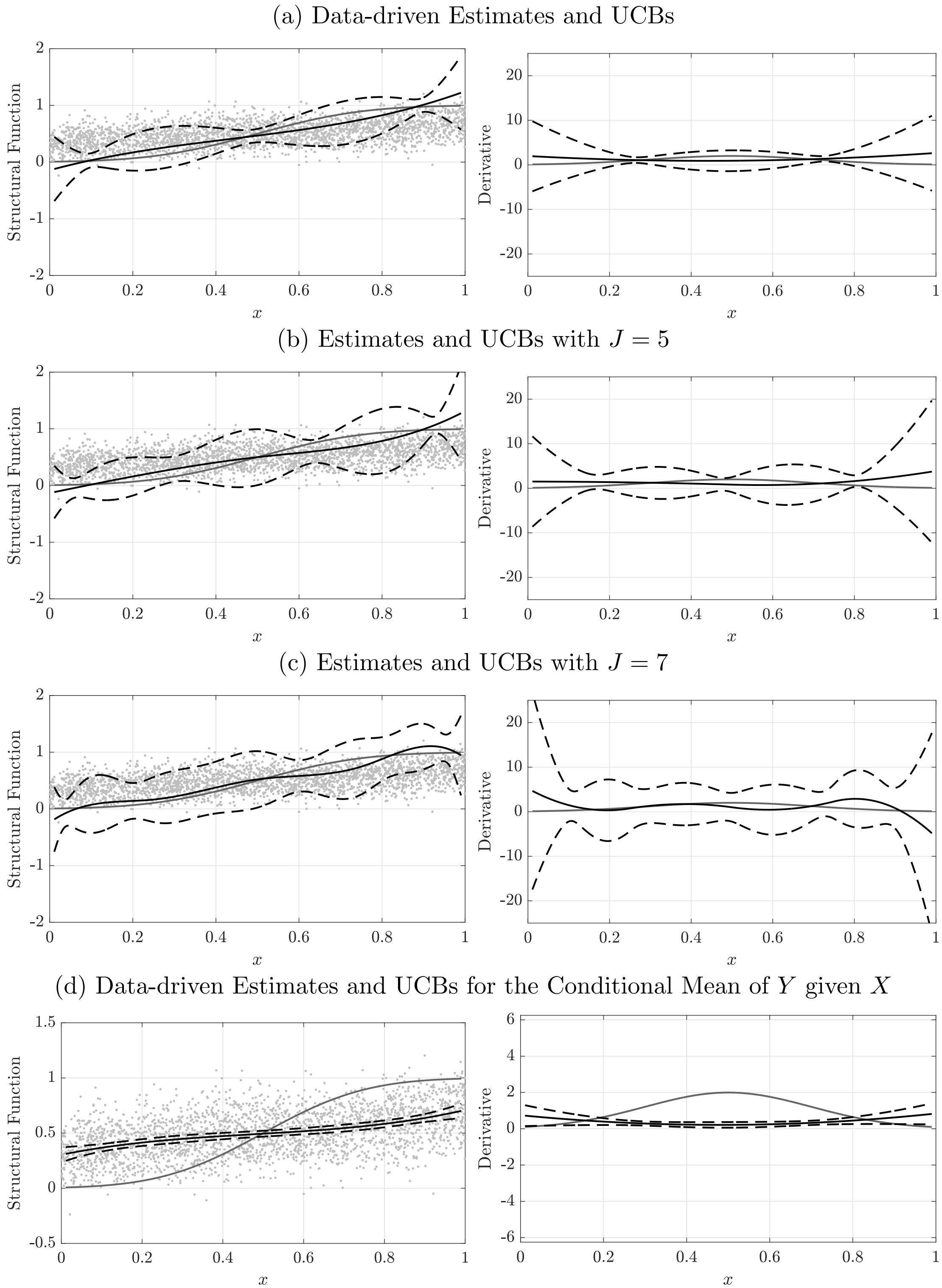

Figure 7: Engel curve design: Plots for a sample of size $n = 2500$. Left panels correspond to the structural function, right panels correspond to its derivative. *Note:* Solid grey lines are the true structural function and derivative; solid black lines are estimates, dashed black lines are 95% UCBs.

D Basis Functions and Hölder Classes

Let Ψ_J denote the closed linear subspace of L^2_X spanned by a basis $\{\psi_{J1}, \ldots, \psi_{JJ}\}$. We use the following notation for vectors and matrices formed from the basis functions

$$
\begin{aligned}
\psi^J_x &= (\psi_{J1}(x), \ldots, \psi_{JJ}(x))' , & b^K_w &= (b_{K1}(w), \ldots, b_{KK}(w))' , \\
\zeta_{\psi,J} &= \sup_{x \in [0,1]^d} \|G_{\psi,J}^{-1/2} \psi^J_x\|_{\ell^2} , & \zeta_{b,J} &= \sup_{w \in [0,1]^{d_w}} \|G_{b,J}^{-1/2} b^{K(J)}_w\|_{\ell^2} , \\
G_{\psi,J} &= \mathbb{E}\big[\psi^J_X (\psi^J_X)'\big] , & G_{b,J} &= \mathbb{E}\big[b^{K(J)}_W (b^{K(J)}_W)'\big] , \\
S_J &= \mathbb{E}\big[b^{K(J)}_W (\psi^J_X)'\big] , & S^o_J &= G_{b,J}^{-1/2} \mathbb{E}\big[b^{K(J)}_W (\psi^J_X)'\big] G_{\psi,J}^{-1/2} .
\end{aligned}
$$

Let s_J be the smallest singular value of $(G_{b,J})^{-1/2} S_J (G_{\psi,J})^{-1/2}$. By Lemma A.1 of Chen and Christensen (2018), under Assumptions 1 and 3(i) there is a finite positive constant a_τ such that

$$
a_\tau^{-1} s_J^{-1} \leq \tau_J \leq s_J^{-1} \quad \text{for all} \quad J \in \mathcal{T} . \tag{33}
$$

D.1 B-splines

The construction of univariate B-spline bases supported on $[0,1]$ follows Chapter 12.3 of DeVore and Lorentz (1993). The basis is characterized by an *order* $r \in \mathbb{N}$ (or *degree* $r-1$) and a *resolution level* $l \in \mathbb{N} \cup \{0\}$. 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[17] 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 take tensor products of univariate bases. A B-spline basis supported on $[0,1]^d$ of order r and resolution level l has dimension $J = (2^l + r - 1)^d$. The set of possible sieve dimensions J is therefore $\mathcal{T} = \{(2^l + r - 1)^d : l \in \mathbb{N} \cup \{0\}\}$.

We now review properties of B-spline bases that are used in the technical arguments below. The following Lemma summarizes Lemma E.2 of Chen and Christensen (2018).

Lemma D.1 *Let Assumption 1(i) hold. Then for $\psi^J(x)$ formed from tensor product B-splines, there are constants $C_\psi, a_\zeta > 0$ depending only on a_f such that (i) $\sup_{x \in [0,1]^d} \|\psi^J(x)\|_{\ell^1} \leq C_\psi$; (ii) $C_\psi^{-1} J^{-1} \leq \lambda_{\min}(G_{\psi,J}) \leq \lambda_{\max}(G_{\psi,J}) \leq C_\psi J^{-1}$; (iii) $\sqrt{J} \leq \zeta_{\psi,J} \leq a_\zeta \sqrt{J}$.*

[17] This basis is equivalent to a B-spline basis with interior knots at $2^{-l}, \ldots, 1-2^{-l}$. This knot placement ensures bases are nested across different l (equivalently, J). For irregularly spaced data, interior knots can be placed at the $2^{-l}, \ldots, 1-2^{-l}$ quantiles of the distribution of X.

Corollary D.1 *Let Assumption 1(ii) hold. Then for $b^{K(J)}(w)$ formed from tensor product B-splines and $J \leq K(J) \lesssim J$, there are constants $C_b, a_\zeta > 0$ depending only on a_f such that (i) $\sup_{w\in[0,1]^{d_w}} \|b^{K(J)}(w)\|_{\ell^1} \leq C_b$; (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 B-splines in the proofs. Note that $N_r(\cdot)$ is Lipschitz with $r = 2$ and $r-2$ times continuously differentiable when $r > 2$. Hence, $\|G_{\psi,J}^{-1/2}\big([\psi^J(x_1)]-[\psi^J(x_2)]\big)\|_{\ell^2} \leq CJ^\omega\|x_1-x_2\|_{\ell^2}^{\omega'}$ holds for some positive constants C, ω, ω'. The B-spline basis also satisfies a Bernstein inequality (or inverse estimate): $\|\partial^a f\|_\infty \lesssim J^{|a|/d}\|f\|_\infty$ holds for any $f \in \Psi_J$ and multi-index a with $|a| < r-1$.

D.2 CDV Wavelets

The construction of CDV wavelet bases supported on $[0,1]$ is reviewed in Appendix E.2 of Chen and Christensen (2018) and follows Cohen, Daubechies, and Vial (1993); see also chapter 4.3.5 of Giné and Nickl (2016). The basis is characterized by an *order* $N \in \mathbb{N}$. Let L denote the smallest integer for which $2^L \geq 2N$. For each *resolution level* $l \geq L$, there are a total of 2^l basis functions. In the multivariate case we generate bases supported on $[0,1]^d$ by taking tensor products of univariate bases. The set of possible J is therefore $\mathcal{T} = \{2^{ld} : l = L+1, L+2, \ldots\}$.

We say that the CDV wavelet basis is S-regular if it is S times continuously differentiable. A S-regular basis can always be chosen by choosing the order N such that $0.18(N-1) \geq S$ (Giné and Nickl, 2016, Theorem 4.2.10(e)). 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 $[\underline{p}, \overline{p}]$ 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 is analogous to choosing the order of a kernel in kernel-based nonparametric estimation.

CDV wavelet bases for the d_w-dimensional instrumental variable W are 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 $\lim_{J\to\infty} K(J)/J = c \in [1,\infty)$. 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 properties of CDV wavelet bases that are used in the proofs below. The following Lemma summarizes Lemma E.4 of Chen and Christensen (2018).

Lemma D.2 *Let Assumption 1(i) hold. Then with $\psi^J(x)$ formed from tensor product CDV wavelets, there are constants $C_\psi, a_\zeta > 0$ depending only on a_f such that (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) $\sqrt{J} \leq \zeta_{\psi,J} \leq a_\zeta\sqrt{J}$.*

Corollary D.2 *Let Assumption 1(ii) hold. Then with $b^{K(J)}(w)$ formed from tensor product CDV wavelets and $J \leq K(J) \lesssim J$, there are constants $C_b, a_\zeta > 0$ depending only on a_f such that (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 CDV wavelets in the proofs. As the Daubechies wavelet functions are S times continuously differentiable on their supports, it follows by Lemma D.2(ii) that the basis functions are Hölder continuous, in the sense that $\|G_{\psi,J}^{-1/2}\big([\psi_{x_1}^J] - [\psi_{x_2}^J]\big)\|_{\ell^2} \leq CJ^\omega\|x_1 - x_2\|_{\ell^2}^{\omega'}$ holds for some positive constants C, ω, ω'. This basis also satisfies a Bernstein inequality (or inverse estimate): $\|\partial^a f\|_\infty \lesssim J^{|a|/d}\|f\|_\infty$ holds for any $f \in \Psi_J$ and multi-index a with $|a| < S$.

D.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 $B_{\infty,\infty}^p(M) = \{h \in B_{\infty,\infty}^p : \|h\|_{B_{\infty,\infty}^p} \leq M\}$ denote the Hölder ball of smoothness p and radius M. For $p \notin \mathbb{N}$, we have $h \in B_{\infty,\infty}^p$ if and only if

$$\|h\|_{C^{\lfloor p\rfloor}} + \sum_{a:|a|=\lfloor p\rfloor} \sup_{\substack{x,y\in[0,1]^d:\\ x\neq y}} \frac{|\partial^a h(x) - \partial^a h(y)|}{|x-y|^{p-\lfloor p\rfloor}} < \infty\,,$$

where

$$\|h\|_{C^{\lfloor p\rfloor}} = \|h\|_\infty + \sum_{|a|=\lfloor p\rfloor} \|\partial^a h\|_\infty\,.$$

The space $B_{\infty,\infty}^p$ may equivalently be defined by the error in approximating a function using a linear B-spline basis (see DeVore and Popov (1988) and DeVore and Lorentz (1993)). To do so, let Ψ_J be a CDV wavelet space of regularity $S > p$ or dyadic B-spline space of degree $r-1 > p$ at resolution level L_J that generates J. Let $d(h,\Psi_J) = \inf_{g\in\Psi_J} \|h-g\|_\infty$. We then have

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

and, moreover, $\|h\|_\infty + \sup_{J:J\in\mathcal{T}} J^{p/d} d(h,\Psi_J)$ is equivalent to $\|h\|_{B^p_{\infty,\infty}}$; see, e.g., Theorem 12.3.3. of DeVore and Lorentz (1993) for the scalar case and Theorem 4.8 of DeVore and Popov (1988) for the multivariate case. By Lebesgue's lemma (DeVore and Lorentz, 1993, p. 30), we have

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

where $\|\Pi_J\|_\infty := \sup_{h:\|h\|_\infty \le 1} \|\Pi_J h\|_\infty$ is the L^∞ norm of the L^2_X projection onto Ψ_J (sometimes referred to as the Lebesgue constant). Huang (2003) and Chen and Christensen (2015) 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. Hence,

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

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

E Technical Results and Proofs of Main Results

In this Appendix we first introduce additional notation. We then present technical results and proofs of the main results from Sections 4.2 and 4.3. We finally present technical results and the proofs of main results for Section 4.4.

E.1 Notation

By the discussion in Appendix D, there are finite positive constants a_ζ and a_b such that

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

For any sequence $(Z_i)_{i=1}^n$ of random vectors and any function g, let $\mathbb{E}_n[g(Z)] = \frac{1}{n}\sum_{i=1}^n g(Z_i)$. Estimators of the matrices defined at the beginning of Appendix D and their orthogonalized versions are

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

Sieve variances and related terms are

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

where

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

with $\hat{\sigma}_{J,J_2}^2(x)$ and $\tilde{\sigma}_{J,J_2}(x)$ given in (9), and

$$\begin{aligned}
\widehat{\Omega}_{J,J_2} &= \mathbb{E}_n\left[\hat{u}_J\hat{u}_{J_2}b_W^{K(J)}b_W^{K(J_2)}\right]'\,, & \hat{u}_{i,J} &= Y_i - \hat{h}_J(X_i)\,, & \widehat{\Omega}_J &= \widehat{\Omega}_{J,J}\,,\\
\Omega_{J,J_2} &= \mathbb{E}\left[u^2 b_W^{K(J)}b_W^{K(J_2)}\right]'\,, & u_i &= Y_i - h_0(X_i)\,, & \Omega_J &= \Omega_{J,J}\,.
\end{aligned}$$

Recall that Π_J is the L_X^2 projection onto Ψ_J. We also define

$$\Delta_J h_0 = h_0 - \Pi_J h_0\,, \qquad \tilde{h}_J(x) = \hat{L}_{J,x}\mathbb{E}_n[b_W^{K(J)}h_0(X)]\,.$$

For bootstrap and related processes, we use the notation

$$\mathbb{Z}_n^*(x,J,J_2) = \frac{1}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\hat{L}_{J,x}b_{W_i}^{K(J)}\hat{u}_{i,J} - \hat{L}_{J_2,x}b_{W_i}^{K(J_2)}\hat{u}_{i,J_2}\right)\varpi_i\right)\,, \tag{34}$$

where $(\varpi_i)_{i=1}^n$ are IID $N(0,1)$ draws independent of the data, and

$$\mathbb{Z}_n^*(x,J) \equiv \frac{D_J^*(x)}{\hat{\sigma}_J(x)} = \frac{1}{\|\hat{\sigma}_{x,J}\|_{sd}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \hat{L}_{J,x}b_{W_i}^{K(J)}\hat{u}_{i,J}\varpi_i\right)\,, \tag{35}$$

$$\widehat{\mathbb{Z}}_n(x,J) = \frac{1}{\|\sigma_{x,J}\|_{sd}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n L_{J,x}b_{W_i}^{K(J)}u_i\varpi_i\right)\,, \tag{36}$$

$$\mathbb{Z}_n(x,J) = \frac{1}{\|\sigma_{x,J}\|_{sd}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n L_{J,x}b_{W_i}^{K(J)}u_i\right)\,. \tag{37}$$

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

E.2 Technical Results

Here we present several technical results that are used in the proofs of the main results in Section 4. The proofs of these technical results are presented in our earlier working paper version (Chen, Christensen, and Kankanala, 2022). The following Lemmas E.1 to E.7 are labelled as Lemmas D.1 to D.7 in Chen et al. (2022), whereas the following Theorems E.1 and E.2 are labelled as Theorems D.1 and D.2 in Chen et al. (2022).

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 any positive constant R, define

$$\bar{J}_{\max}(R) = \sup\left\{ J \in \mathcal{T} : J\sqrt{\log J}\left[(\log n)^4 \vee \tau_J\right] \le R\sqrt{n} \right\}. \tag{38}$$

For $D > 0$ and $p \in [\underline{p}, \overline{p}]$, define

$$\begin{aligned} J_0(p, D) &= \sup\left\{ J \in \mathcal{T} : \tau_J \frac{\sqrt{J}\theta^*_{1-\hat{\alpha}}}{\sqrt{n}} \le D J^{-\frac{p}{d}} \right\}, \\ J_0^+(p, D) &= \inf\{ J \in \mathcal{T} : J > J_0(p, D)\}. \end{aligned} \tag{39}$$

Lemma E.1 *Let Assumptions 1-4 hold and let $\tau_J \asymp J^{\varsigma/d}$ with $\varsigma \ge 0$. Then: with $\bar{J}_{\max}(R)$ as defined in (38) for any $R > 0$ and $J_0^+(p, D)$ as defined in (39) 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 lemma relates to resolution levels in the severely ill-posed case. For $R > 0$ and $p \in [\underline{p}, \overline{p}]$, define

$$\bar{J}^*_{\max}(R) = \sup\left\{ J \in \mathcal{T} : \tau_J J\sqrt{\log J} \le R\sqrt{n} \right\}, \tag{40}$$

$$\begin{aligned} M_0(p, R) &= \sup\{ J \in \mathcal{T} : \tau_J J^{\frac{p}{d}+\frac{1}{2}}\sqrt{\log J} \le R\sqrt{n}\}, \\ M_0^+(p, R) &= \inf\{ J \in \mathcal{T} : J > M_0(p, R)\}. \end{aligned} \tag{41}$$

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

Lemma E.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.*

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

Recall the definition of $\bar{J}_{\max}(R)$ from (38) and that $\Delta_J h_0 = h_0 - \Pi_J h_0$.

Theorem E.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 $C_{E.1} > 0$ such that*

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

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

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

Recall the definition of $\bar{J}_{\max}(R)$ from (38). 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 $\mathcal{J}_n = \{J \in \mathcal{T} : J_{\min} \leq J \leq \bar{J}_{\max}\}$,

$$\mathcal{S}_n = \{(x, J, J_2) \in \mathcal{X} \times \mathcal{J}_n \times \mathcal{J}_n : J_2 > J\} \tag{42}$$

and

$$\delta_n = \tau_{\bar{J}_{\max}} \sqrt{\frac{\bar{J}_{\max} \log \bar{J}_{\max}}{n}} + \left(\frac{\bar{J}^2_{\max} \log \bar{J}_{\max}}{n} \right)^{1/3} + J_{\min}^{-p/d}\,. \tag{43}$$

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

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

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

(ii) we have

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

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

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

In particular,

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

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

For the following lemma, recall $\hat{J}_{\max}$ from (10) and $\bar{J}_{\max}(R)$ from (38).

Lemma E.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 (10), there exists $R_1, R_2>0$ which satisfy*

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

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

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

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

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

For the following Lemma, recall the critical value $\theta^*_{1-\hat{\alpha}}$ from Section 2.3.

Lemma E.6 *Let Assumptions 1-4 hold. Then: with $\bar{J}_{\max}(R)$ as defined in (38) for any $R>0$, there exists constants $C_4, C_5>0$ for which*

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

The second is a companion result concerning the critical value $z^*_{1-\alpha}$ from Section 2.4:

Lemma E.7 *Let Assumptions 1-4 hold. Then: with $\bar{J}_{\max}(R)$ as defined in (38) for any $R>0$, there exists a constant $C_{E.7}>0$ for which*

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

E.2.5 Uniform Consistency for the Bootstrap

Recall $\bar{J}_{\max}\equiv \bar{J}_{\max}(R)$ from (38) and $\mathcal{J}_n$ and $\mathcal{S}_n$ from (42).

Theorem E.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:*

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

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

E.3 Proofs of Main Results in Sections 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 (38) sufficiently large so that by Lemma E.5 we have $\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}(\hat{J}_{\max} \le \bar{J}_{\max}(R_2)) \to 1$. Let $\bar{J}_{\max} \equiv \bar{J}_{\max}(R_2)$ for the remainder of the proof. By Theorem E.1(i), there exists $C_{E.1} > 0$ which satisfies

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\|\tilde{h}_J - \Pi_J h_0\|_\infty \le C_{E.1}\|\Pi_J h_0 - h_0\|_\infty \quad \forall\, J \in [1, \bar{J}_{\max}] \cap \mathcal{T} \right) \to 1. \tag{44}$$

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

$$\sup_{p\in[\underline{p},\overline{p}]} \sup_{h_0\in\mathcal{H}^p} J^{\frac{p}{d}} \|\Pi_J h_0 - h_0\|_\infty \le B_2 \qquad \forall\;\; J \in \mathcal{T}\,. \tag{45}$$

Let $\hat{\mathcal{S}} = \{(x,J,J_2) \in \mathcal{X} \times \hat{\mathcal{J}} \times \hat{\mathcal{J}} : J_2 > J\}$. Lemmas E.3 and E.5, Assumption 4(i), and the fact that $\delta_n \downarrow 0$ (cf. (43)) imply that there exists $C_2, C_3 > 0$ which satisfy

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J,J_2)\in\hat{\mathcal{S}}} \frac{\tau_{J_2}\sqrt{J_2}}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} \le C_3 \right) \to 1\,,\; \inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J,J_2)\in\hat{\mathcal{S}}} \frac{\|\hat{\sigma}_{x,J,J_2}\|_{sd}}{\tau_{J_2}\sqrt{J_2}} \le C_2 \right) \to 1\,. \tag{46}$$

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

$$\inf_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(C_4\sqrt{\log \bar{J}_{\max}} \le \theta^*_{1-\hat{\alpha}} \le C_5\sqrt{\log \bar{J}_{\max}} \right) \to 1. \tag{47}$$

Part (i), step 1: We verify that $\hat{J}$ achieves the optimal rate under mild ill-posedness. Note by the procedure in Appendix A this is sufficient for adaptivity of $\tilde{J}$ for nonparametric regression. Fix $\xi > 1$ ($\xi = 1.1$ in the main text). Choose $D > 0$ such that

$2B_2(C_1+1)D^{-1}C_3 < (\xi-1)$. Recall $J_0(p,D)$ and $J_0^+(p,D)$ from (39); we drop dependence of these quantities on (p,D) hereafter to simplify notation. By Lemma E.1, $\inf_{p\in[\underline{p},\overline{p}]}\inf_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}(J_0^+ < \bar{J}_{\max}) \to 1$. It then follows from Lemmas E.1 and E.5 that $\inf_{p\in[\underline{p},\overline{p}]}\inf_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}(J_0^+ < \hat{J}_{\max}) \to 1$. We therefore assume for the remainder of the proof of part (i) that $J_0^+ < \hat{J}_{\max}, \bar{J}_{\max}$.

By Lemma E.5, $\hat{\mathcal{J}} \subseteq \mathcal{J}_n := \{J \in \mathcal{T} : 0.1(\log \bar{J}_{\max})^2 \le J \le \bar{J}_{\max}\}$ wpa1 $\mathcal{H}$-uniformly. Then for all $J \in \hat{\mathcal{J}}$ with $J > J_0^+$, by the triangle inequality, displays (44) and (45), and definition of J_0, we may deduce that

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

wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p},\overline{p}]$. By (46), we 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 $J > J_0^+$,

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

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

$$
\begin{aligned}
&\sup_{p\in[\underline{p},\overline{p}]}\sup_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}\big(\hat{J} > J_0^+\big) \\
&\le \sup_{p\in[\underline{p},\overline{p}]}\sup_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}\left(\sup_{J\in\hat{\mathcal{J}}:J>J_0^+}\sup_{x\in\mathcal{X}}\frac{\sqrt{n}|\hat{h}_{J_0^+}(x) - \hat{h}_J(x)|}{\|\hat{\sigma}_{x,J_0^+,J}\|_{sd}} > \xi\theta^*_{1-\hat{\alpha}}\right) \\
&\le \sup_{h_0\in\mathcal{H}}\mathbb{P}_{h_0}\left(\sup_{(x,J,J_2)\in\hat{\mathcal{S}}}\frac{\sqrt{n}|\hat{h}_J(x) - \hat{h}_{J_2}(x) - (\tilde{h}_J(x) - \tilde{h}_{J_2}(x))|}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} > \theta^*_{1-\hat{\alpha}}\right) + o(1)\,. \qquad (48)
\end{aligned}
$$

To control the r.h.s. probability in (48), 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 $\theta^*_{1-\hat{\alpha};\tilde{J}}$ denote the $(1-0.5\wedge$

$\sqrt{(\log \tilde{J})/\tilde{J}})$ quantile of $\sup_{(x,J,J_2)\in\hat{\mathcal{S}}(\tilde{J})} |\mathbb{Z}_n^*(x,J,J_2)|$. Then by Lemma E.5 and Theorem E.2(ii), we have

$$
\begin{aligned}
&\sup_{h_0\in\mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J,J_2)\in\hat{\mathcal{S}}} \frac{\sqrt{n}|\hat{h}_J(x)-\hat{h}_{J_2}(x)-(\tilde{h}_J(x)-\tilde{h}_{J_2}(x))|}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} > \theta^*_{1-\hat{\alpha}}\right) \\
&\le \sup_{h_0\in\mathcal{H}} \sum_{\tilde{J}\in\mathcal{T}:\tilde{J}=\bar{J}_{\max}(R_1)}^{\bar{J}_{\max}(R_2)} \mathbb{P}_{h_0}\left(\sup_{(x,J,J_2)\in\hat{\mathcal{S}}(\tilde{J})} \frac{\sqrt{n}|\hat{h}_J(x)-\hat{h}_{J_2}(x)-(\tilde{h}_J(x)-\tilde{h}_{J_2}(x))|}{\|\hat{\sigma}_{x,J,J_2}\|_{sd}} > \theta^*_{1-\hat{\alpha};\tilde{J}}\right) \\
&\le \sum_{\tilde{J}\in\mathcal{T}:\tilde{J}=\bar{J}_{\max}(R_1)}^{\bar{J}_{\max}(R_2)} \left(\sqrt{(\log \tilde{J})/\tilde{J}}+\gamma_n+o(1)\right) \to 0\,, \qquad (49)
\end{aligned}
$$

where the final line holds for all n large, because $\bar{J}_{\max}(R_1)\to\infty$, $\gamma_n\downarrow 0$, and, by our choice of sieve and Remark E.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 (48), 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 (46) that wpa1 uniformly for $h_0\in\mathcal{H}^p$ and $p\in[\underline{p},\overline{p}]$, we have

$$
\begin{aligned}
\|\hat{h}_{\hat{J}}-h_0\|_\infty &\le \|\hat{h}_{\hat{J}}-\hat{h}_{J_0^+}\|_\infty+\|\hat{h}_{J_0^+}-h_0\|_\infty \\
&\le C_2\xi\theta^*_{1-\hat{\alpha}}\tau_{J_0^+}\sqrt{J_0^+/n}+\|\hat{h}_{J_0^+}-\tilde{h}_{J_0^+}\|_\infty+\|\tilde{h}_{J_0^+}-h_0\|_\infty.
\end{aligned}
$$

Then by Theorem E.1, definition of J_0^+, and the lower bound on $\theta^*_{1-\hat{\alpha}}$ in display (47), we may deduce that there exists a constant $C>0$ for which

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

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

$$
\inf_{p\in[\underline{p},\overline{p}]}\inf_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}\left(\|\hat{h}_{\hat{J}}-h_0\|_\infty\le CC'DJ_0^{-p/d}\right)\to 1. \qquad (50)
$$

By the upper bound on $\theta^*_{1-\hat{\alpha}}$ in display (47) and because $\sqrt{\log \hat{J}_{\max}}\asymp\sqrt{\log n}$ (as the model is mildly ill-posed), there exists a constant $E > 0$ such that by defining

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

we have $\inf_{p\in[\underline{p},\overline{p}]}\inf_{h_0\in\mathcal{H}^p}\left(J_n^*(p,E) \le J_0(p,D)\right) \to 1$. Hence, as $\tau_J \asymp J^{\varsigma/d}$ we have $J_n^*(p,E) \asymp (n/\log n)^{d/(2(p+\varsigma)+d)}$. The desired result now follows from (50).

Part (i), step 2: We verify that $\tilde{J}$ achieves the optimal rate under mild ill-posedness. By step 1, we have $\inf_{p\in[\underline{p},\overline{p}]}\inf_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}(\hat{J} \le J_0^+) \to 1$. 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 $\theta^*_{1-\hat{\alpha}}$ in display (47) and the fact that $\sqrt{\log \bar{J}_{\max}} \asymp \sqrt{\log n}$ (as the model is mildly ill-posed), we may deduce that there exists a constant $E' > 0$ such that $\inf_{p\in[\underline{p},\overline{p}]}\inf_{h_0\in\mathcal{H}^p}\left(J_n^\dagger(p,E') \ge J_0^+(p,D)\right) \to 1$ 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 E.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. To simplify notation we assume a CDV wavelet basis, though a similar argument applies (albeit with more complicated notation) for B-splines. 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 (40). By Theorem E.1, Lemma E.5, and Remark E.1, we may deduce that there exist constants $D, D' > 0$ for which

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

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

Recall the definition of $M_0(p,R_2)$ from (41). By Lemma E.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 $\{n_k\}_{k\ge1}$ 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 (40) 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

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

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 $\|\hat{h}_{\hat{J}} - h_0\|_\infty \le \|\hat{h}_{\hat{J}} - \hat{h}_{\hat{J}_n}\|_\infty + \|\hat{h}_{\hat{J}_n} - h_0\|_\infty$, 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 (46) and (47), we have

$$\begin{aligned}
&\sup_{p\in[\underline{p},\overline{p}]}\sup_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}\big(\|\hat{h}_{\hat{J}}-\hat{h}_{\hat{J}_n}\|_\infty > D(\log n)^{-p/\varsigma} \text{ and } \hat{J}<\hat{J}_n\big)\\
&\leq \sup_{p\in[\underline{p},\overline{p}]}\sup_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}\left(\xi C_2\theta^*_{1-\hat{\alpha}}\tau_{\hat{J}_n}\sqrt{\hat{J}_n/n}>D(\log n)^{-p/\varsigma}\right)+o(1)\\
&\leq \sup_{p\in[\underline{p},\overline{p}]}\mathbb{1}\left[\xi C_2C_5\tau_{2^{-d}\bar{J}^*_{\max}(R_2)}\sqrt{2^{-d}\bar{J}^*_{\max}(R_2)\log(2^{-d}\bar{J}^*_{\max}(R_2))/n}>D(\log n)^{-p/\varsigma}\right]+o(1)\,.
\end{aligned}$$

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 (39). 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

$$\inf_{p\in[\underline{p},\overline{p}]}\inf_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}\left(\|\partial^a\hat{h}_{\hat{J}}-\partial^a h_0\|_\infty\leq CJ_0^{(|a|-p)/d}\right)\to 1.$$

The result follows from noting, as in the proof of part (i), step 1 of the proof of Theorem 4.1, that

$$\inf_{p\in[\underline{p},\overline{p}]}\inf_{h_0\in\mathcal{H}^p}\mathbb{P}_{h_0}\big(J_n^*(p,E)\leq J_0(p,D)\big)\to 1\,,$$

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 (40) 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

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

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

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

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,

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

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 E.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 (with more complicated notation) applies for B-splines.

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 result for the band

$$C_n(x) = \left[\hat{h}_{\tilde{J}}(x) - \left(z^*_{1-\alpha} + \hat{A}\theta^*_{1-\hat{\alpha}}\right)\hat{\sigma}_{\tilde{J}}(x),\ \hat{h}_{\tilde{J}}(x) + \left(z^*_{1-\alpha} + \hat{A}\theta^*_{1-\hat{\alpha}}\right)\hat{\sigma}_{\tilde{J}}(x)\right],$$

(cf. (16)). Note by Appendix A this implies the result holds for our UCBs for nonparametric regression as well. Fix $R_2 > 0$ in the definition of $\bar{J}_{\max}(R_2)$ from (38) sufficiently large so that by Lemma E.5 we have $\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}(\hat{J}_{\max} \le \bar{J}_{\max}(R_2)) \to 1$. Let $\bar{J}_{\max} \equiv \bar{J}_{\max}(R_2)$ for the remainder of the proof. Recall the constants $C_{E.1}$ from (44), $\underline{B}$ and $\overline{B}$ from the discussion preceding the statement of this theorem, and C_4 and C_5 from (47). Also note that by Lemmas E.3 and E.5, Assumption 4(i), and the fact that $\delta_n \downarrow 0$ (cf. (43)) imply that there exists $C_2, C_3 > 0$ which satisfy

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J) \in \mathcal{X} \times \hat{\mathcal{J}}} \frac{\tau_J \sqrt{J}}{\|\hat{\sigma}_{x,J}\|_{sd}} \le C_3 \right) \to 1\,, \; \inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0}\left(\sup_{(x,J) \in \mathcal{X} \times \hat{\mathcal{J}}} \frac{\|\hat{\sigma}_{x,J}\|_{sd}}{\tau_J \sqrt{J}} \le C_2 \right) \to 1\,. \tag{51}$$

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

Define $J_0(p,E)$ as in (39). Part (i), step 1 of the proof of Theorem 4.1 implies that $J_0(p,E) \gtrsim (n/\log n)^{d/(2(p+\varsigma)+d)}$. By Lemma E.5 and mild ill-posedness, for any constant $C > 0$ we have $J_0(p,E)/(\log \hat{J}_{\max})^2 \ge C$ wpa1 uniformly for $h_0 \in \mathcal{H}^p$ and $p \in [\underline{p}, \overline{p}]$. Hence, $\inf\{J \in \mathcal{T} : J \ge \beta J_0(p,E)\} > \log \hat{J}_{\max}$ 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))

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

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 $\inf\{J : J \in \hat{\mathcal{J}}\} \ge \underline{J}$ holds wpa1 $\mathcal{H}$-uniformly by Lemma E.5. Recalling the Lebesgue constant $\|\Pi_J\|_\infty$ from Appendix D.3, we may then deduce

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

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 (44) and the discussion preceding the statement of this theorem that

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

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 (51) that

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

where the final line is by (49).

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

$$
\inf_{p\in[\underline{p},\overline{p}]}\inf_{h_0\in\mathcal{G}^p}\mathbb{P}_{h_0}\big(\beta J_0(p,E)\le\hat{J}\le J_0^+(p,D)\big)\to 1\,. \tag{52}
$$

Therefore, by (44), (51), (52), and definition of $\overline{B}$, for every $h_0\in\mathcal{G}^p$ and $x\in\mathcal{X}$ we have

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

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 (39) that whenever the preceding inequality holds, we have

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

where the final inequality holds uniformly for $h_0\in\mathcal{G}^p$ and $p\in[\underline{p},\overline{p}]$ for a constant $A_0>0$

because $\sup_{J\in\mathcal{T}} \tau_{2^d J}/\tau_{\lceil \beta J\rceil} < \infty$ by virtue of mild ill-posedness. Hence for any $A \geq A_0$,

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

where the final line is because $\hat{J} \in \underline{\mathcal{J}}_n := \{J \in \mathcal{T} : 0.1(\log \bar{J}_{\max}(R_2))^2 \leq J \leq \bar{J}^-_{\max}(R_1)\}$ with $\bar{J}^-_{\max}(R_1) = \sup\{J \in \mathcal{T} : J < \bar{J}_{\max}(R_1)\}$ 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 (52) and Lemma E.1 and the latter holds by Lemma E.5 and the fact that $\hat{J} = \tilde{J}$ wpa1 $\mathcal{H}$-uniformly. 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 E.2(i) and the definition of $\underline{z}^*_{1-\alpha}$.

Part (ii): By Lemmas E.4, E.6, and E.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 (52) with $J_0 = J_0(p,D)$ and $\bar{A} = 1 + A$, we have that

$$
\sup_{x\in\mathcal{X}} |C_n(x,A)| \lesssim \bar{A}\tau_{J_0^+}\sqrt{(J_0^+\log \bar{J}_{\max})/n} \lesssim \bar{A}\tau_{J_0}\sqrt{(J_0\log \bar{J}_{\max})/n} \lesssim \bar{A}\frac{\sqrt{\log \bar{J}_{\max}}}{\theta^*_{1-\hat{\alpha}}} J_0^{-p/d}
$$

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 (39). It follows by Lemma E.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 $J^+=2^dJ$ for $J\in\mathcal{T}$. A similar argument (with more complicated notation) applies for B-spline bases.

Part (ii): First note by Lemma E.5 and the fact that $\bar{J}_{\max}(R)=\bar{J}^*_{\max}(R)$ (see (40)) 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 (41). By Lemma E.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 E.4, E.6, and E.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 (39). By similar arguments to Lemma E.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 E.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 E.1 and Lemma E.5, there exists a constant $A_0 > 0$ for which

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

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

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

Suppose that $J^*_{\max}(R_2) \ge 2^{2d} J^*_{\max}(R_1) \in \mathcal{T}$. Then by definition of $\bar{J}^*_{\max}(R)$ and Remark E.1, we have

$$\frac{\tau_{J^*_{\max}(R_2)}}{\tau_{2^{2d} J^*_{\max}(R_1)}} \asymp \frac{\tau_{J^*_{\max}(R_2)}}{\tau_{2^{2d} J^*_{\max}(R_1)}} \frac{J^*_{\max}(R_2)\sqrt{\log J^*_{\max}(R_2)}}{2^{2d} J^*_{\max}(R_1)\sqrt{\log J^*_{\max}(R_1)}} \le \frac{R_2}{R_1}\,. \tag{53}$$

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

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

which contradicts (53). 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 E.5 that $\hat{J}_{\max} \in \{2^{-d}\bar{J}^*_{\max}(R_2), \bar{J}^*_{\max}(R_2)\}$ wpa1 $\mathcal{H}$-uniformly. Therefore, $\tilde{J} \le 2^{-d}\bar{J}^*_{\max}(R_2)$ holds wpa1 $\mathcal{H}$-uniformly. But by part (ii) we also have that $\tilde{J} \ge c\bar{J}^*_{\max}(R_2)$ holds for a sufficiently small $c > 0$ wpa1 uniformly $h_0 \in \mathcal{G}^p$ and $p \in [\underline{p}, \overline{p}]$. Therefore, $\tilde{J} \in \underline{\mathcal{J}}_n := \{J \in \mathcal{T} : c\bar{J}^*_{\max}(R_2) \le J \le 2^{-d}\bar{J}_{\max}(R_2)\}$ 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 $\underline{z}^*_{1-\alpha} \le z^*_{1-\alpha}$ must hold whenever $\hat{\mathcal{J}} \supseteq \underline{\mathcal{J}}_n$, we therefore have

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

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

E.4 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 (43). We introduce the bootstrap process for the derivatives:

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

where $\|\hat{\sigma}_{x,J}^a\|_{sd}^2 \equiv n\hat{\sigma}_J^{a2}(x) = \hat{L}_{J,x}^a \widehat{\Omega}_{J,J} (\hat{L}_{J,x}^a)'$ 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: $\partial^a \psi_x^J = (\partial^a \psi_{J1}(x), \ldots, \partial^a \psi_{JJ}(x))'$. Proofs of these supplemental results are presented in our earlier working paper version Chen et al. (2022), where they are labelled as Lemmas E.12, E.13, and E.14, respectively.

Lemma E.8 *Let Assumptions 1-3 hold. Then: there is a universal constant $C_{E.8} > 0$ such that*

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

Lemma E.9 *Let Assumptions 1-4 hold. For a given $\alpha \in (0,1)$, let $z_{1-\alpha}^{a*}$ denote the $1-\alpha$ quantile of $\sup_{(x,J) \in \mathcal{X} \times \hat{\mathcal{J}}} |\mathbb{Z}_n^{a*}(x,J)|$. Then: with $\bar{J}_{\max}(R)$ as defined in (38) for any $R > 0$, there exists a constant $C_{E.9} > 0$ for which*

$$\inf_{h_0 \in \mathcal{H}} \mathbb{P}_{h_0} \left(z_{1-\alpha}^{a*} \leq C_{E.9} \sqrt{\log \bar{J}_{\max}(R)} \right) \to 1\,.$$

Lemma E.10 *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*

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

holds wpa1 $\mathcal{H}$-uniformly.

E.5 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 E.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 E.8 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 (52) 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 E.10 in place of Theorem E.2(i).

Part (ii): By Lemma E.6, Lemmas E.8 and E.9 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 (52), with $J_0 = J_0(p, D)$ we have that

$$\begin{aligned}\sup_{x \in \mathcal{X}} |C^a_n(x, A)| &\lesssim (1 + A)\tau_{J_0^+} (J_0^+)^{1/2+|a|/d} \sqrt{(\log \bar{J}_{\max})/n} \\ &\lesssim (1 + A)\tau_{J_0} J_0^{|a|/d} \sqrt{(J_0 \log \bar{J}_{\max})/n} \lesssim (1 + A) \frac{\sqrt{\log \bar{J}_{\max}}}{\theta^*_{1-\hat{\alpha}}} J_0^{(|a|-p)/d}\end{aligned}$$

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 (39). 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 E.5, Theorem E.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

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

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

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

The remainder of the proof now follows similarly to the proof of Theorem 4.3, using Lemma E.10 in place of Theorem E.2(i).

Part (ii): By Lemmas E.2, E.6, E.8, and E.9 and Assumption 4(iii), there exist constants $C, C' > 0$ for which

$$\begin{aligned}
\sup_{x \in \mathcal{X}} |C_n^a(x, A)| &\leq C(1 + A)\tau_{\tilde{J}} \tilde{J}^{1/2 + |a|/d} \sqrt{\log(\bar{J}^*_{\max}(R_2))/n} + A\tilde{J}^{(|a|-\underline{p})/d} \\
&\leq C'(1 + A)(J^*_{\max}(R_2))^{(|a|-p)/d} + A\tilde{J}^{(|a|-\underline{p})/d}
\end{aligned}$$

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

Beare, B. K. (2010). Copulas and temporal dependence. *Econometrica 78*(1), 395–410.

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

Chen, X., T. Christensen, and S. Kankanala (2022). Adaptive estimation and uniform confidence bands for nonparametric structural functions and elasticities. arxiv:2107.11869v2 [econ.em], (dated October 11, 2022).

Chen, X. and T. M. Christensen (2015). 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.

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

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.

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

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