# Adaptive Inference in Multivariate Nonparametric Regression Models Under Monotonicity*

Koohyun Kwon † Soonwoo Kwon ‡

November 26, 2020

**Abstract**

We consider the problem of adaptive inference on a regression function at a point under a multivariate nonparametric regression setting. The regression function belongs to a Hölder class and is assumed to be monotone with respect to some or all of the arguments. We derive the minimax rate of convergence for confidence intervals (CIs) that adapt to the underlying smoothness, and provide an adaptive inference procedure that obtains this minimax rate. The procedure differs from that of Cai and Low (2004), intended to yield shorter CIs under practically relevant specifications. The proposed method applies to general linear functionals of the regression function, and is shown to have favorable performance compared to existing inference procedures.

*We thank our advisors Donald Andrews and Timothy Armstrong for continuous support. Xiaohong Chen, Yuichi Kitamura, and participants at the Yale Econometrics Prospectus Lunch provided valuable feedback. We thank David Tomas Jacho-Chavez for providing us with the dataset used in Section 6.

†Department of Economics, Yale University, koohyun.kwon@yale.edu

‡Department of Economics, Yale University, soonwoo.kwon@yale.edu

# 1 Introduction

We consider the problem of inference on a regression function at a point under the nonparametric regression model

$$y_i = f(x_i) + u_i, \quad u_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2(x_i)),$$

where $f$ is assumed to lie in a Hölder class with exponent $\gamma \in (0, 1]$. Procedures based on $\gamma$ is conservative (or suboptimal) when the true regression function in fact lies in a smoother Hölder class with $\gamma' > \gamma$. Adaptive procedures try to overcome this issue by automatically adjusting to the (unknown) underlying smoothness class. However, unlike in the case of estimation, where adaptation to the unknown smoothness class is in general possible with an additional logarithmic term (Lepskii, 1991), adaptation is impossible in the case of inference without further restrictions on the function class (Low, 1997) .

Two shape restrictions that can be used to overcome this impossibility have been discussed in the literature, convexity and monotonicity. In this paper, we impose monotonicity on the regression function to construct a CI that adapts to the underlying smoothness of the regression function. The main difference with other papers that consider adaptation under a monotonicity condition (Cai et al., 2013; Armstrong (2015)) is our general treatment of the dimension of $x_i$. To our knowledge, this is the first paper to construct adaptive CIs, under a multivariate nonparametric regression setting.

We consider coordinate-wise monotonicity with respect to all or some of the coordinates. A function $f$ is *coordinate-wise monotone* with respect to $\mathcal{V} \subseteq \{1, \ldots, k\}$ if $x_j \geq z_j$ for all $j \in \mathcal{V}$ and $x_j = z_j$ for all $j \notin \mathcal{V}$ imply $f(z) \geq f(z)$. The minimax expected length of a CI over the Hölder class with exponent $\gamma$ converges to 0 at the well-known rate of $n^{-1/(2+k/\gamma)}$. When the regression function is monotone in all variables, i.e., $\mathcal{V} = \{1, \ldots, k\}$, we can construct a CI that achieves this minimax rate over all $\gamma \in (0, 1]$ just as in the univariate case. Also, again as in the univariate case, if the regression is not monotone to any of the variables so that $\mathcal{V} = \emptyset$, there is no scope for adaptation.

An interesting case is when the function is monotone with respect to only some of the variables so that $k_+ := |\mathcal{V}| < k$, which can arise due to the multivariate nature of the problem. In this case, we show that for a CI that maintains coverage over the

Hölder class with exponent $\gamma$, the minimax expected length over a smoother class $\gamma' > \gamma$ converges to 0 at the rate $n^{-1/(2+k_+/\gamma'+(k-k_+)/\gamma)}$. The denominator of the exponent can be written as $2 + k/\gamma - k_+(1/\gamma - 1/\gamma')$. This is the sum of a term that comes from the minimax rate over $\gamma$, $2 + k/\gamma$, and $-k_+(1/\gamma - 1/\gamma')$. In this sense, $k_+(1/\gamma - 1/\gamma')$ exactly quantifies the possible gain from monotonicity, indicating larger gains if the regression function is monotone in more variables and/or smoother.

We propose a CI that obtains this minimax rate (of adaptation) for a sequence of Hölder exponents $\{\gamma_j\}_{j=1}^{J} \subset (0, 1]$. While the method provided by Cai and Low (2004) can be used to construct such a CI, we provide an alternative method that builds upon the one-sided CI proposed by Armstrong and Kolesár (2018). Their one-sided CI "directs power" to a smoother class while maintaining coverage over a larger class of functions. Our CI is constructed by combining the lower and upper versions of their one-sided CI to create a two-sided CI, and then taking the intersection of a sequence of such two-sided CIs that direct power to each $\gamma_j$. An appropriate Bonferroni correction is used to obtain correct coverage. This CI can be used in more general nonparametric regression settings, as long as the parameter of interest is a linear functional of the regression function and the regression functions lies in a convex function class.

While the proposed CI obtains the minimax length over $\gamma_j$ for each $j$ up to a constant factor that does not depend on the sample size, this constant does depend on the number of parameter spaces $J$ the CI adapts to. This is in contrast with the CI of Cai and Low (2004), which gives a multiplicative constant that does not depend on $J$. However, the multiplicative constant of our CI grows slowly with $J$ at a $(\log J)^{1/2}$ rate, and is smaller than the constant given by Cai and Low (2004) for any reasonable specification of $J$. Even if one wishes to adapt to $J = 10^3$ parameter spaces, our CI obtains the minimax expected length of each parameter space within a multiplicative constant of 4.14, whereas this constant is 16 for the CI by Cai and Low (2004). A simulation study confirms that our CI can be significantly shorter in practice as well. Nonetheless, the uniform constant that Cai and Low (2004) obtain is theoretically attractive and allows one to adapt to the continuum of Hölder exponents $(0, 1]$ in this context.

**Related literature.** An adaptation theory for CIs in a nonparametric regression setting was developed by Cai and Low (2004). Cai et al. (2013) provide a procedure for constructing adaptive CIs that adapt to each individual function under monotonicity and convexity. Armstrong (2015) provides an inference method for the regression function at a point, possibly on the boundary of the support, that adapts to the underlying Hölder classes under a monotonicity assumption. As noted earlier, the main difference of our paper is that we consider a multivariate regression setting where there is no restriction on the dimension of the independent variable as long as it is fixed and finite. The adaptation theory for CIs builds upon the more classical minimax theory for CIs, which has been developed in Donoho (1994) and Low (1997). Cai (2012) provides an excellent review on the theory of minimax and adaptive CIs, along with the minimax and adaptive estimation problems.

While the focus of this paper is on adaptive CIs, there are other forms of confidence sets that are of interest in the context of nonparametric regression setting. Adaptive confidence balls have been considered in Genovese and Wasserman (2005), Cai and Low (2006) and Robins and van der Vaart (2006). An adaptation theory for confidence bands has been considered in, for example, Dümbgen (1998), Genovese and Wasserman (2008), and Cai et al. (2014). In the context of density estimation, adaptive confidence bands have also been considered in Hengartner and Stark (1995), Giné and Nickl (2010), and Hoffmann and Nickl (2011).

Recently, there has been interest in isotonic regression in general dimensions. The monotonicity condition imposed in such models is the same as the one we impose here with $\mathcal{V} = \{1, \ldots, k\}$. Han et al. (2019) derive minimax rates for the least squares estimation problem. Deng et al. (2020) provide a method for constructing CIs at a point based on block max-min and min-max estimators.

**Outline.** Section 2 describes the nonparametric regression model and the function class we consider. Section 3 introduces the notion of adaptivity in more detail and describes our procedure for constructing adaptive CIs. Section 4 presents the main result of the paper, the minimax rate of adaptation, and an adaptive CI that obtains this rate by solving the corresponding modulus problem. Section 5 provides a simulation study, and Section 6 illustrates our method in the context of production function estimation.

Any proof omitted in the main text can be found in the appendix. Appendix A collects the proofs for lemmas and corollaries. Appendix B contains the proof for our main theoretical result, Theorem 4.1.

# 2 Nonparametric Regression Under Monotonicity

We observe $\{(y_i, x_i)\}_{i=1}^n$ and consider a nonparametric regression model,

$$y_i = f(x_i) + u_i, \tag{1}$$

where $x_i \in \mathcal{X} \subset \mathbb{R}^k$ is a (fixed) regressor, $f : \mathbb{R}^k \to \mathbb{R}$ is the unknown regression function that lies in some function class $\mathcal{F}$, and $u_i$'s are independent with $u_i \sim N(0, \sigma^2(x_i))$ and $\sigma^2(\cdot)$ known. The parameter of interest is $f(x_0)$. For the rate results provided in Section 4.2, we require that $x_0 \in \operatorname{Int} \mathcal{X}$. However, we note that the solution to the modulus problem given in Section 4.1 does not depend on whether $x_0$ is on the boundary or not. Without loss of generality, we normalize $x_0$ to be 0.

We take the $\mathcal{F}$ to be the class of functions that are Hölder continuous and nondecreasing in all or some of the variables. Let $\Lambda(\gamma, C)$ denote the set of functions from $\mathbb{R}^k$ to $\mathbb{R}$ that are Hölder continuous with Hölder constants $(\gamma, C)$,

$$\Lambda(\gamma, C) := \left\{ f \in \mathcal{F}(\mathbb{R}^k, \mathbb{R}) : |f(x) - f(z)| \leq C \|x - z\|^\gamma \text{ for all } x, z \in \mathcal{X} \right\},$$

where $\mathcal{F}\left(\mathbb{R}^k, \mathbb{R}\right)$ is the set of functions from $\mathbb{R}^k$ to $\mathbb{R}$, $\gamma \in [0, 1]$, $C \geq 0$ and $\|\cdot\|$ is a norm on $\mathbb{R}^k$. For notational simplicity, we omit the dependence of the function class on the choice of the norm $\|\cdot\|$. We impose the following restriction that $\|\cdot\|$ is monotone in the magnitude of each element, which is satisfied by most norms used in practice. such as the $\ell_p$ norm or a weighted version of it. We discuss the relationship between this assumption and the monotonicity of the regression function in Remark 4.1.

**Assumption 2.1.** *$\|\cdot\|$ is a norm on $\mathbb{R}^k$ such that $\|z\|$ is nondecreasing in $|z_j|$ for each $j \in \{1, ..., k\}$.*

We now define the (coordinate-wise) monotone Hölder class. For a subset of the covariate indices $\mathcal{V} \subset \{1, \ldots, k\}$, write

$$\Lambda_{+,\mathcal{V}}(\gamma, C) := \{ f \in \Lambda(\gamma, C) : f(x) \geq f(z) \text{ if } x_j \geq z_j \ \forall j \in \mathcal{V} \text{ and } x_j = z_j \ \forall j \notin \mathcal{V} \}.$$

This is the set of Hölder continuous functions that are nondecreasing, coordinate-wise, with respect to the $j$th element for $j \in \mathcal{V}$. Define $k_+ := |\mathcal{V}|$. By a relabeling argument, it is without loss of generality to write $\mathcal{V} := \{1, \ldots, k_+\}$. If $k_+ = k$, then

$\Lambda_{+,\mathcal{V}}(\gamma, C)$ is the set of nondecreasing and Hölder continuous functions where the monotonicity is with respect to the coordinate-wise partial ordering on $\mathbb{R}^k$.

# 3 Adaptive Confidence Intervals

## 3.1 Notion of Adaptivity

In this section, we discuss the problem of inference for a general linear functional of the regression function, $Lf$. Consider a sequence of convex parameter spaces $\mathcal{F}_1$, $\ldots$, $\mathcal{F}_J$, with the requirement that $\mathcal{F}_j \subset \mathcal{F}_J$ for all $j \leq J$. Note that the parameter spaces are not necessarily nested, but there is a largest convex parameter space that nests all the other parameter spaces. Here, $\mathcal{F}_J$ reflects a conservative choice of the parameter space where the researcher believes the true regression function to lie in. Hence, the CI we construct will be required to maintain correct coverage over this space. An adaptive CI maintains this correct coverage over the largest parameter space $\mathcal{F}_J$ while having good performance (e.g. shorter expected length) when the true function happens to lie in the smaller parameter space $\mathcal{F}_j$, simultaneously for all $j \leq J$.

Then, a natural question is how well a CI that maintains coverage over $\mathcal{F}_J$ can perform over $\mathcal{F}_j$, which is one of the main questions that Cai and Low (2004) raise and address in detail in the context of two-sided CIs. The case of one-sided CIs has been considered by Armstrong and Kolesár (2018), along with other questions.

### 3.1.1 Two-sided Adaptive CIs

Let $\mathcal{I}^J_{\alpha,2}$ denote the set of all two-sided CIs that have coverage at least $1-\alpha$ over $\mathcal{F}_J$. Following Cai and Low (2004), the performance criterion we consider for two-sided CIs is the worst-case expected length. That is, the performance of a CI, $CI$, over the parameter space $\mathcal{F}_j$ is measured by $\sup_{f \in \mathcal{F}_j} \mathbf{E}_f\, \mu(CI)$ with smaller values of this quantity meaning better performance. Here, $\mathbf{E}_f$ denotes the expectation when the true regression function is $f$ and $\mu$ is the Lebesgue measure on the real line. Then, the shortest possible worst-case expected length a CI can achieve over $\mathcal{F}_j$ (while maintaining correct coverage over $\mathcal{F}_J$) is characterized by the quantity

$$L^*_{j,J} := \inf_{CI \in \mathcal{I}^J_{\alpha,2}} \sup_{f \in \mathcal{F}_j} \mathbf{E}_f\, \mu(CI).$$

Following Cai and Low (2004), we say a CI is *adaptive* if it achieves $L^*_{j,J}$ for all $j \le J$ up to a multiplicative constant that does not depend on the sample size. Let $z_q$ denote the $q$–quantile of the standard normal distribution. Cai and Low (2004) show that $L^*_{j,J} \asymp \omega_+(z_{1-\alpha}, \mathcal{F}_j, \mathcal{F}_J)$, with $\asymp$ denoting asymptotic equivalence[1] and $\omega_+(\delta, \mathcal{F}_j, \mathcal{F}_J)$ is the *between class modulus of continuity* defined as

$$
\begin{aligned}
&\omega_+(\delta, \mathcal{F}_j, \mathcal{F}_J) \\
:=& \sup\left\{|Lf_J - Lf_j| : \textstyle\sum_{i=1}^n \left((f_J(x_i) - f_j(x_i))/\sigma(x_i)\right)^2 \le \delta^2, f_j \in \mathcal{F}_j, f_J \in \mathcal{F}_J\right\},
\end{aligned}
$$

for $\delta \ge 0$.[2] In general, $\omega_+(z_{1-\alpha}, \mathcal{F}_j, \mathcal{F}_J)$ is more tractable than $L^*_{j,J}$, and thus the strategy is to construct a CI that has worst case length over $\mathcal{F}_j$ bounded by $\omega_+(z_{1-\alpha}, \mathcal{F}_j, \mathcal{F}_J)$, up to a multiplicative constant. We refer to the rate at which $\omega_+(z_{1-\alpha}, \mathcal{F}_j, \mathcal{F}_J)$ converges to 0 as the *minimax rate of adaptation* (of $\mathcal{F}_j$ over $\mathcal{F}_J$). If $\mathcal{F}_j = \mathcal{F}_J$, this is the minimax rate over $\mathcal{F}_J$, which is the fastest rate at which the worst-case expected length over $\mathcal{F}_J$ of a CI that maintains correct coverage over the same space $\mathcal{F}_J$ can achieve.

#### 3.1.2 One-sided CIs

While our main focus is on adaptive two-sided CIs, the construction of our adaptive CI relies heavily on the one-sided CI proposed by Armstrong and Kolesár (2018). Hence, we briefly describe the notion of adaptivity in the context of one-sided CIs. For one-sided CIs, we follow Armstrong and Kolesár (2018) and consider the $\beta$th quantile of excess length as the performance criterion. More specifically, for a one sided lower CI, $[\hat{c}, \infty)$, we denote the $\beta$th quantile of the excess length at $f$ as $q_{\beta,f}(Lf - \hat{c})$, where $q_{\beta,f}(\cdot)$ denotes the $\beta$th quantile function when the true regression function is $f$. Under this criterion, the best possible performance over $\mathcal{F}_j$ is quantified by

$$
\ell^*_{j,J} := \inf_{\hat{c}:[\hat{c},\infty)\in\mathcal{I}^J_{\alpha,\ell}} \sup_{f\in\mathcal{F}_j} q_{\beta,f}(Lf - \hat{c}),
$$

[1] We write $a_n \asymp b_n$ if $0 < \liminf_{n\to\infty}(a_n/b_n) \le \limsup_{n\to\infty}(a_n/b_n) < \infty$.

[2] Note that the definition is slightly different with Cai and Low (2004) due to the $\sigma(x_i)$ term that appears in the denominator of the summand. This is because we divide both sides of (1) by the (known) $\sigma(x_i)$ to convert the model into the same form as that of Cai and Low (2004).

where $\mathcal{I}_{\alpha,\ell}^{J}$ denotes the set of all one-sided lower CIs that have coverage at least $1-\alpha$ over $\mathcal{F}_J$. Armstrong and Kolesár (2018) showed that $\ell_{j,J}^{*}=\omega(z_{1-\alpha}+z_\beta,\mathcal{F}_J,\mathcal{F}_j)$, where $\omega(z_{1-\alpha}+z_\beta,\mathcal{F}_J,\mathcal{F}_j)$ is the *ordered class modulus of continuity* defined as

$$
\begin{aligned}
&\omega(\delta,\mathcal{F}_j,\mathcal{F}_k)\\
:=&\sup\left\{Lf_k-Lf_j:\textstyle\sum_{i=1}^{n}\left((f_k(x_i)-f_j(x_i))/\sigma(x_i)\right)^2\leq\delta^2, f\in\mathcal{F}_j, f_J\in\mathcal{F}_J\right\},
\end{aligned}
$$

for any $\delta\geq 0$ and $j,k\leq J$. We refer to the optimization problem in the definition as the ordered modulus problem. Naturally, an analogous result holds for upper one-sided CIs so that $u_{j,J}^{*}=\omega(z_{1-\alpha}+z_\beta,\mathcal{F}_j,\mathcal{F}_J)$, where

$$
u_{j,J}^{*}:=\inf_{\hat{c}:(-\infty,\hat{c}]\in\mathcal{I}_{\alpha,u}^{J}}\sup_{f\in\mathcal{F}_j}q_{\beta,f}(\hat{c}-Lf),
$$

with $\mathcal{I}_{\alpha,u}^{J}$ denoting the set of all one-sided upper CIs that have coverage at least $1-\alpha$ over $\mathcal{F}_J$.

We say a one-sided lower CI, $[\hat{c}^*,\infty)$, is adaptive if there exists some $c>0$ that does not depend on $n$ such that

$$
\sup_{f\in\mathcal{F}_j}q_{\beta,f}(Lf-\hat{c}^*)\leq c\,\omega(z_{1-\alpha}+z_\beta,\mathcal{F}_j,\mathcal{F}_J)
$$

for all $j\leq J$, and similarly for one-sided upper CIs.

### 3.1.3 Modes of Adaptation

Note that it must be the case that $\omega_+(z_{1-\alpha},\mathcal{F}_j,\mathcal{F}_j)\leq\omega_+(z_{1-\alpha},\mathcal{F}_j,\mathcal{F}_J)$ (and similarly for the ordered moduli) because $\omega_+(z_{1-\alpha},\mathcal{F}_j,\mathcal{F}_j)$ takes the supremum over a smaller set. However, if it happens to be the case that $\omega_+(z_{1-\alpha},\mathcal{F}_j,\mathcal{F}_j)\asymp\omega_+(z_{1-\alpha},\mathcal{F}_j,\mathcal{F}_J)$, an adaptive CI, $CI^*$, satisfies $\sup_{f\in\mathcal{F}_j}\mathbf{E}_f\,\mu(CI^*)\leq\bar{c}\,L_{j,j}^{*}$ for all $j\leq J$. Cai and Low (2004) define such CI to be *strongly adaptive*. This is an ideal case because we obtain $L_{j,j}^{*}$, up to a multiplicative constant, which is the minimax length we could have achieved if we "knew" that our true regression function lied in the smaller class $\mathcal{F}_j$ (i.e., if we made a stronger assumption that the true regression function lies in this smaller class). While adaptive CIs exist in general, strong adaptation is possible only when $\omega_+(z_{1-\alpha},\mathcal{F}_j,\mathcal{F}_J)\asymp\omega_+(z_{1-\alpha},\mathcal{F}_j,\mathcal{F}_j)$ for all $j\leq J$. This is not a property of a given procedure, but of the given statistical model.

The least desirable case is when $\omega_+(z_{1-\alpha}, \mathcal{F}_j, \mathcal{F}_J) \asymp \omega_+(z_{1-\alpha}, \mathcal{F}_J, \mathcal{F}_J)$, because this leaves no scope of adaptation. An intermediate case is when

$$\omega_+(z_{1-\alpha}, \mathcal{F}_j, \mathcal{F}_j) \prec \omega_+(z_{1-\alpha}, \mathcal{F}_j, \mathcal{F}_J) \prec \omega_+(z_{1-\alpha}, \mathcal{F}_J, \mathcal{F}_J),$$

so that the minimax rate of adaptation is better than the worst-case minimax rate over $\mathcal{F}_J$ but not as good as the minimax rate over $\mathcal{F}_j$.[3] That is, one can do better than simply taking the most conservative parameter space as the true space but not quite as good as knowing that the true function actually lies in the smaller parameter space. Hence, the minimax adaptation rate plays an important role in determining whether sharp adaptation is possible. In Section 4.2, we derive the minimax rates of adaptation under the model given in Section 2.

### 3.2 Construction of Adaptive CIs

Cai and Low (2004) provide a general method of constructing adaptive CIs of $Lf$ under the general model (1). Here, we provide an alternative method that is intuitive and gives smaller constants in the case of non-nested parameter spaces.[4] For the nested case, the CI of Cai and Low (2004) has a bounded constant even as $J \to \infty$, which is an attractive theoretical property. For the CI we propose, the constant will grow with $J$ in general. In practice, however, one can only adapt to finitely many parameter spaces due to computational constraints. The proposed procedure gives a smaller constant than that of Cai and Low (2004) even for unrealistically large values of $J$ (e.g., $J = 10^{10}$).

The main building block for our adaptive CI is the minimax one-sided CI proposed by Armstrong and Kolesár (2018), which relies on the ordered modulus. We say that $(f_j, f_k) \in \mathcal{F}_j \times \mathcal{F}_k$ is a solution to $\omega(\delta, \mathcal{F}_j, \mathcal{F}_k)$ if $(f_j, f_k)$ solves the optimization problem corresponding to $\omega(\delta, \mathcal{F}_j, \mathcal{F}_k)$. Let $(f^{*,Jj}_{J,\delta}, g^{*,Jj}_{j,\delta}) \in \mathcal{F}_J \times \mathcal{F}_j$ be a solution to

[3] For positive sequences $\{a_n\}$ and $\{b_n\}$, we write $a_n \prec b_n$ if $\liminf_{n\to\infty}(b_n/a_n) = \infty$.

[4] For a given adaptive CI, $CI^*$, we refer to the positive number $c$ (that does not depend on $n$) such that $\sup_{f\in\mathcal{F}_j} E\mu(CI^*) \leq c\,\omega_+\left(z_\alpha, \mathcal{F}_j, \mathcal{F}_J\right)$, as the "constant" of $CI^*$.

the ordered modulus $\omega\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right)$, and define the estimator

$$\begin{aligned}\hat{L}_\delta^{\ell,j} =& \frac{1}{2} L(g_{j,\delta}^{*,Jj} + f_{J,\delta}^{*,Jj}) \\ &+ \frac{\omega'\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right)}{\delta} \sum_{i=1}^{n} (g_{j,\delta}^{*,Jj}(x_i) - f_{J,\delta}^{*,Jj}(x_i)) \left( \frac{y_i}{\sigma(x_i)} - \frac{g_{j,\delta}^{*,Jj}(x_i) + f_{J,\delta}^{*,Jj}(x_i)}{2} \right),\end{aligned} \tag{2}$$

where $\omega'(\cdot, \mathcal{F}_J, \mathcal{F}_j)$ is the derivative of $\omega(\cdot, \mathcal{F}_J, \mathcal{F}_j)$. Based on this estimator, define a lower one-sided CI by subtracting the maximum bias and an appropriately scaled normal quantile:

$$\hat{c}_{\alpha,\delta}^{\ell,j} := \hat{L}_\delta^{\ell,j} - \frac{1}{2}\omega\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right) + \frac{1}{2}\delta\omega'\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right) - z_{1-\alpha}\omega'\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right). \tag{3}$$

The following theorem from Armstrong and Kolesár (2018) shows that for a specific choice of $\delta$, this CI is optimal in the sense that it achieves $\ell_{j,J}^*$.

**Lemma 3.1** (Theorem 3.1 of Armstrong and Kolesár (2018))**.** *Let $\underline{\delta} = z_\beta + z_{1-\alpha}$. Then,*

$$\sup_{f \in \mathcal{F}_j} q_{f,\beta}(Lf - \hat{c}_{\alpha,\underline{\delta}}^{\ell,j}) = \ell_{j,J}^* = \omega(\underline{\delta}, \mathcal{F}_J, \mathcal{F}_j).$$

The excess length $Lf - \hat{c}_{\alpha,\underline{\delta}}^{\ell,j}$ follows a Gaussian distribution because it is a affine transformation of the data, which follows a Gaussian distribution by assumption. Hence, the median and mean of the excess length are the same. Taking $\beta = 1/2$, we can replace $q_{f,\beta}$ with the expectation under $f$, which gives

$$\sup_{f \in \mathcal{F}_j} \mathbf{E}_f \left( Lf - \hat{c}_\alpha^{\ell,j} \right) = \omega\left(z_{1-\alpha}, \mathcal{F}_J, \mathcal{F}_j\right), \tag{4}$$

where we define $\hat{c}_\alpha^{\ell,j} := \hat{c}_{\alpha,z_{1-\alpha}}^{\ell,j}$. Likewise, we can define an optimal upper one-sided CI $(-\infty, \hat{c}_{\alpha,\underline{\delta}}^{\ell,j}]$ such that

$$\sup q_{f,\beta}(\hat{c}_{\alpha,\underline{\delta}}^{u,j} - Lf) = u_{j,J}^* = \omega(\underline{\delta}, \mathcal{F}_j, \mathcal{F}_J), \tag{5}$$

where the precise definition of $\hat{c}_{\alpha,\underline{\delta}}^{\ell,j}$ is given in Appendix E. Similarly, let $\hat{c}_\alpha^{u,j}$ denote the upper counterpart of $\hat{c}_\alpha^{\ell,j}$.

Using the optimal one-sided CIs, we first show how a naive Bonferroni procedure leads to a two-sided adaptive CI. We then provide a method that improves upon this naive Bonferroni CI by taking into account the correlation among the CIs. The naive Bonferroni CI is defined as

$$CI_{\alpha}^{Bon,J} := \cap_{j=1}^{J}[\hat{c}_{\alpha/2J}^{\ell,j}, \hat{c}_{\alpha/2J}^{u,j}]. \tag{6}$$

This has coverage at least $1-\alpha$ over $\mathcal{F}_J$ because each $[\hat{c}_{\alpha/2J}^{\ell,j}, \hat{c}_{\alpha/2J}^{u,j}]$ has coverage $1-\alpha/J$ over $\mathcal{F}_J$ and $CI_{\alpha}^{Bon,J}$ is simply the intersection of such CIs. The following theorem shows that this CI is indeed adaptive.

**Theorem 3.1.** *For any $j = 1, \ldots, J$, we have*

$$\sup_{f\in\mathcal{F}_j} \mathbf{E}\,\mu(CI_{\alpha}^{Bon,\mathcal{J}}) \leq \frac{2z_{1-\frac{\alpha}{2J}}}{z_{1-\frac{\alpha}{2}}}\,\omega_{+}(z_{1-\frac{\alpha}{2}}, \mathcal{F}_j, \mathcal{F}_J). \tag{7}$$

The constant $2z_{1-\frac{\alpha}{2J}}/z_{1-\frac{\alpha}{2}}$ increases with the number of parameter spaces $J$.[5] On the other hand, the constant given in Cai and Low (2004) is 16 and thus does not depend on the number of parameter spaces. However, we note that $2z_{1-\frac{\alpha}{2J}}/z_{1-\frac{\alpha}{2}}$ is not too large, in fact smaller than 16, for reasonable specifications of $J$. For example, when $\alpha = 0.05$ and $J = 50$, we get $2z_{1-\frac{\alpha}{2J}}/z_{1-\frac{\alpha}{2}} \approx 3.36$, which is considerably smaller than the constant given in Cai and Low (2004). Even for unrealistically large $J$ such as $J = 10^{10}$, we have $2z_{1-\frac{\alpha}{2J}}/z_{1-\frac{\alpha}{2}} < 8$, which is still less than half of the constant given by Cai and Low (2004). Simulation results given in Section 5 confirm that not only the upper bound, but also the actual length itself is often much shorter for our CI.

**Remark 3.1.** Suppose one is interested in constructing the one-sided CI in an adaptive way. Note that Lemma 3.1 implies that any one-sided CI $[\hat{c}_{\alpha}^{\ell,J}, \infty)$ with coverage probability $1-\alpha$ should satisfy

$$\sup_{f\in\mathcal{F}_j} \mathbf{E}(Lf - \hat{c}_{\alpha}^{\ell,J}) \geq \omega\left(z_{1-\alpha}, \mathcal{F}_J, \mathcal{F}_j\right).$$

[5]The constant, $z_{1-\frac{\alpha}{2J}}/z_{1-\frac{\alpha}{2}}$, grows with $J$ at the rate $(\log J)^{1/2}$. This is the same rate that Cai and Low (2004) find in their analysis of the case with non-nested parameter spaces. Their constant is at least eight times greater than what we provide here, but does not require that the largest space in consideration is convex.

Define $\hat{c}_{\alpha}^{\ell,J} = \max_j \hat{c}_{\alpha/J}^{\ell,j}$. Then, by an analogous argument to Theorem 3.1, we have

$$\sup_{f\in\mathcal{F}_j} \mathbf{E}(Lf - \hat{c}_{\alpha}^{\ell,J}) \leq \frac{z_{1-\frac{\alpha}{J}}}{z_{1-\alpha}} \omega\left(z_{1-\alpha}, \mathcal{F}_J, \mathcal{F}_j\right).$$

Therefore, $\left[\hat{c}_{\alpha}^{\ell,J}, \infty\right)$ is an adaptive one-sided CI in a similar sense with the two-sided case.

The naive CI given in (6) does not take into account the possible correlation among the CIs that we take the intersection of. However, if parameter spaces are "close" to each other, the corresponding CIs will be correlated, implying that there is room for improvement over the Bonferroni procedure. Consider the CIs of the form $CI^{\tau,\mathcal{J}} = \cap_{j=1}^{J}[\hat{c}_{\tau}^{\ell,j}, \hat{c}_{\tau}^{u,j}]$. If we take $\tau = \alpha/(2J)$, this is precisely the CI given in (6). The CI that gives the smallest constant among CIs of such forms is $CI^{\tau^*,\mathcal{J}}$, where $\tau^*$ is the largest possible $\tau$ such that $CI^{\tau,\mathcal{J}}$ has correct coverage over $\mathcal{F}_J$:

$$\tau^* := \sup_{\tau} \tau \quad \text{s.t.} \quad \inf_{f\in\mathcal{F}_J} \mathbf{P}_f(Lf \in CI^{\tau,\mathcal{J}}) \geq 1-\alpha.$$

We know that $\tau = \alpha/(2J)$ satisfies the constraint, and also that any $\tau > \alpha$ does not because then $[\hat{c}_{\tau}^{\ell,j}, \infty)$ will have coverage probability $1-\tau < 1-\alpha$. Hence, we can restrict $\tau$ to lie in $[\alpha/(2J), \alpha]$.

However, the coverage probability $\inf_{f\in\mathcal{F}_J} \mathbf{P}_f(Lf \in CI^{\tau,\mathcal{J}})$ is unknown in general, rendering $CI^{\tau^*,\mathcal{J}}$ infeasible. Instead, we replace this coverage probability with a lower bound that we can calculate either analytically or via simulation. Then, we take $\tau^*$ as the largest value that makes this lower bound at least $1-\alpha$. As we show later, using $\tau^*$ rather than $\alpha/(2J)$ can only make the resulting CI shorter.

Let $(V(\tau)', W(\tau)')'$ be a centered Gaussian random vector with unit variance. The covariance terms for $V(\tau) = (V_1(\tau), ..., V_J(\tau))'$ is given by

$$\operatorname{Cov}\left(V_j(\tau), V_\ell(\tau)\right) = \frac{1}{z_{1-\tau}^2} \sum_{i=1}^{n} \left(g_{j,z_{1-\tau}}^{*,Jj}(x_i) - f_{J,z_{1-\tau}}^{*,Jj}(x_i)\right)\left(g_{\ell,z_{1-\tau}}^{*,J\ell}(x_i) - f_{J,z_{1-\tau}}^{*,J\ell}(x_i)\right).$$

Likewise, the covariance terms for $W(\tau) = \left(W_1(\tau), ..., W_J(\tau)\right)'$ is given by

$$\operatorname{Cov}\left(W_j(\tau), W_\ell(\tau)\right) = \frac{1}{z_{1-\tau}^2} \sum_{i=1}^{n} \left(g_{j,z_{1-\tau}}^{*,jJ}(x_i) - f_{J,z_{1-\tau}}^{*,jJ}(x_i)\right)\left(g_{\ell,z_{1-\tau}}^{*,\ell J}(x_i) - f_{J,z_{1-\tau}}^{*,\ell J}(x_i)\right).$$

Finally, the covariance terms across $V(\tau)$ are $W(\tau)$ given as

$$\mathrm{Cov}\big(V_j(\tau), W_\ell(\tau)\big) = \frac{1}{z_{1-\tau}^2} \sum_{i=1}^{n} \big(g^{*,Jj}_{j,z_{1-\tau}}(x_i) - f^{*,Jj}_{J,z_{1-\tau}}(x_i)\big)\big(g^{*,\ell J}_{\ell,z_{1-\tau}}(x_i) - f^{*,\ell J}_{J,z_{1-\tau}}(x_i)\big).$$

This Gaussian random vector can be used to tune the critical value, as the following lemma implies.

**Lemma 3.2.** *Let $\tau^* \in \left[\frac{\alpha}{2J}, \alpha\right]$ to be the largest value of $\tau$ such that*

$$\mathbf{P}\left(\max\left\{V(\tau)', W(\tau)'\right\} > z_{1-\tau}\right) \leq \alpha. \tag{8}$$

*Then, we have $\sup_{f\in\mathcal{F}_J} \mathbf{P}\left(f(0) \notin CI^{\mathcal{J}}_{\tau^*}\right) \leq \alpha$.*

Such a $\tau^*$ always exists because the inequality (8) holds with $\tau = \alpha/(2J)$ due to the union bound. A solution $\tau^*$ can be found via numerical simulation. By construction, its length will be also bounded by (7). In Section 4.3, we show that as $n \to \infty$ the distribution of $(V(\tau)', W(\tau)')'$ does not depend on $\tau$, under our setting of $Lf = f(0)$ with $f$ belonging to a Hölder class. Hence, finding $\tau^*$ boils down to simply finding the $1-\alpha$ quantile of the maximum of a Gaussian vector in this case.

# 4 Adaptive Inference for $f(0)$

In this section, we provide an adaptive inference procedure for $f(0)$. To construct the adaptive CI introduced in Section 3, we first solve the corresponding modulus problem. By using this solution to the modulus problem, we derive the minimax rate of adaptation. Finally, we provide a CI that obtain this rate, using the method described in Section 3.

## 4.1 Solution to the Modulus Problem

Let $\Lambda_{+,\mathcal{V}}(\gamma_j, C_j) \subset \Lambda_{+,\mathcal{V}}(\gamma_J, C_J)$ with $\gamma_j \geq \gamma_J$ and $C_j \leq C_J$. To construct the adaptive CI, we first calculate the ordered moduli, $\omega\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_j, C_j\right), \Lambda_{+,\mathcal{V}}\left(\gamma_J, C_J\right)\right)$ and $\omega\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_J, C_J\right), \Lambda_{+,\mathcal{V}}\left(\gamma_j, C_j\right)\right)$, for each $j = 1, \ldots, J$. For notational simplicity, we consider the case with $J = 2$ and solve $\omega_+\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right)$, from which the general solution follows immediately.

Recall the definition of the ordered modulus of continuity

$$\begin{aligned}&\sup f_2(0) - f_1(0)\\ \text{s.t.}\ &\sum_{i=1}^{n} \left(\left(f_2(x_i) - f_1(x_i)\right)/\sigma(x_i)\right)^2 \leq \delta^2,\ f_j \in \Lambda_{+,\mathcal{V}}\left(\gamma_j, C_j\right)\ \text{for } j = 1, 2,\end{aligned}$$

with the maximized value denoted by $\omega\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right).$ It is convenient to solve the inverse modulus problem instead, which is defined as

$$\begin{aligned}&\inf \sum_{i=1}^{n} \left(\left(f_2(x_i) - f_1(x_i)\right)/\sigma(x_i)\right)^2\\ \text{s.t.}\ & f_2(0) - f_1(0) = b,\ f_j \in \Lambda_{+,\mathcal{V}}\left(\gamma_j, C_j\right)\ \text{for } j = 1, 2,\end{aligned} \tag{9}$$

for $b > 0$, with the square root of the maximized value denoted by the inverse (ordered) modulus $\omega^{-1}\left(b, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right).$ We provide a closed form solution for the this inverse problem, from which we can recover the solution to the original problem by finding $b$ such that $\omega^{-1}\left(b, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right) = \delta.$ Note that this is simply a search problem on the positive real line.

To characterize the solution to (9), we show two simple lemmas about the properties of the class $\Lambda_{+,\mathcal{V}}\left(\gamma, C\right)$. For $z = (z_1, \ldots, z_k) \in \mathbb{R}^k$, define

$$(z)_{\mathcal{V}+} = \begin{cases} \max\left\{z_i, 0\right\} & i \in \mathcal{V} \\ z_i & i \notin \mathcal{V} \end{cases}$$

and $(z)_{\mathcal{V}-} = (-z)_{\mathcal{V}+}$.

**Lemma 4.1.** *Suppose Assumption 2.1 holds, and let $\gamma \in [0, 1]$ and $C > 0$. Define*

$$h_+(x) = C\left\|(x)_{\mathcal{V}+}\right\|^{\gamma} \quad \textit{and} \quad h_-(x) = -C\left\|(x)_{\mathcal{V}-}\right\|^{\gamma}.$$

*Then, $h_+,\ h_- \in \Lambda_{+,\mathcal{V}}\left(\gamma, C\right)$.*

The following lemma asserts that the class of functions we consider is closed under the maximum operator.

**Lemma 4.2.** *Suppose $h_1, h_2 \in \Lambda_{+,\mathcal{V}}\left(\gamma, C\right)$. Then, $\max\left\{h_1, h_2\right\} \in \Lambda_{+,\mathcal{V}}\left(\gamma, C\right)$.*

The next lemma can be used to establish the solutions to the problem (9). This is a generalization of Proposition 4.1 of Beliakov (2005), which gives the same result for the special case of $\gamma = 1$.

**Lemma 4.3.** *Given* $f_0 \in \mathbb{R}$ *and* $0 < \gamma \leq 1$, *define*

$$\Lambda^{f_0}_{+,\mathcal{V}}(\gamma, C) = \{f \in \Lambda_{+,\mathcal{V}}(\gamma, C) : \; f(0) = f_0\}.$$

*Then, for any* $x \in \mathbb{R}^k$, *we have*

$$\begin{aligned} \max_{f \in \Lambda^{f_0}_{+,\mathcal{V}}(\gamma,C)} f(x) &= f_0 + C \left\|(x)_{\mathcal{V}+}\right\|^{\gamma} \\ \min_{f \in \Lambda^{f_0}_{+,\mathcal{V}}(\gamma,C)} f(x) &= f_0 - C \left\|(x)_{\mathcal{V}-}\right\|^{\gamma}. \end{aligned}$$

We are now ready to characterize the solution to the inverse modulus problem (9). For $r \in \mathbb{R}$, define $(r)_+ := \max\{r, 0\}$.

**Proposition 4.1.** *Suppose Assumption 2.1 holds, and define*

$$\begin{aligned} f_1^*(x) &= C_1 \left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1} \\ f_2^*(x) &= \max\left\{b - C_2 \left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2}, \; C_1 \left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1}\right\}. \end{aligned}$$

*Then,* $(f_1^*, f_2^*)$ *solves the inverse modulus problem (9), and the inverse modulus is given by*

$$\begin{aligned} &\omega^{-1}\left(b, \Lambda_{+,\mathcal{V}}(\gamma_1, C_1), \Lambda_{+,\mathcal{V}}(\gamma_2, C_2)\right) \\ =&\left(\sum_{i=1}^{n} \left(\left(b - C_1 \left\|(x_i)_{\mathcal{V}+}\right\|^{\gamma_1} - C_2 \left\|(x_i)_{\mathcal{V}-}\right\|^{\gamma_2}\right)/\sigma(x_i)\right)_+^2\right)^{1/2}. \end{aligned} \tag{10}$$

*Proof.* To solve (9), note that it is without loss of generality to restrict attention to the functions with $f_1(0) = 0$ and $f_2(0) = b$, which is satisfied by $f_1^*$ and $f_2^*$. To simplify notation, write $\mathcal{F}_1^0 = \Lambda^0_{+,\mathcal{V}}(\gamma_1, C_1)$ and $\mathcal{F}_2^b = \Lambda^b_{+,\mathcal{V}}(\gamma_2, C_2)$. Since $f_2(0) > f_1(0)$, we want $f_1(x) = \max_{f \in \mathcal{F}_1^0} f(x)$ and $f_2(x) = \min_{f \in \mathcal{F}_2^b} f(x)$ as long as $x \in \mathcal{X}$ satisfies $\min_{f \in \mathcal{F}_2} f(x) \geq \max_{f \in \mathcal{F}_1} f(x)$, and $f_1(x) = f_2(x)$ otherwise. Note that $f_1^*$ and $f_2^*$ are designed exactly to achieve this goal, which follows by Lemma 4.3.

It remains to check whether $f_1^* \in \mathcal{F}_1$ and $f_2^* \in \mathcal{F}_2$. The former case is trivial. For the latter case, note that $f_1^* \in \Lambda_{+,\mathcal{V}}(\gamma_1, C_1) \subseteq \Lambda_{+,\mathcal{V}}(\gamma_2, C_2)$. Now, by Lemma 4.2, we have $f_2^* \in \mathcal{F}_2$. □

The following corollary states an analogous result regarding the inverse modulus $\omega^{-1}\left(b, \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right), \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right)\right)$.

**Corollary 4.1.** *Define*

$$\begin{aligned} g_1^*(x) &= b - C_1 \left\|(x)_{\mathcal{V}-}\right\|^{\gamma_1} \\ g_2^*(x) &= \min\left\{b - C_1 \left\|(x)_{\mathcal{V}-}\right\|^{\gamma_1},\ C_2 \left\|(x)_{\mathcal{V}+}\right\|^{\gamma_2}\right\}. \end{aligned}$$

*Then, $(g_1^*, g_2^*)$ solves the inverse modulus $\omega^{-1}\left(b, \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right), \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right)\right)$.*

**Remark 4.1** (Role of Assumption 2.1)**.** Proposition 4.1 requires Assumption 2.1 due to the specific form of monotonicity we consider. By considering coordinate-wise monotonicity, we must take a norm that is "aligned" with this direction of monotonicity. The assumption precisely imposes this. This is a unique feature that arises in the multivariate setting. To allow for more general norms, let $\mathcal{B}$ be an orthonormal basis of $\mathbb{R}^k$, and denote by $z^{\mathcal{B}}$ the coordinate vector of $z \in \mathbb{R}^k$ with respect to $\mathcal{B}$ and $z_j^{\mathcal{B}}$ its $j$th component. Suppose the regression function is monotone in the coefficients with respect to this basis $\mathcal{B}$, so that the monotone Hölder class is given as

$$\Lambda_{+,\mathcal{V}}(\gamma, C) := \left\{f \in \Lambda(\gamma, C) : f(x) \geq f(z) \text{ if } x_j^{\mathcal{B}} \geq z_j^{\mathcal{B}}\ \forall j \in \mathcal{V} \text{ and } x_j^{\mathcal{B}} \geq z_j^{\mathcal{B}}\ \forall j \notin \mathcal{V}\right\}.$$

Then, the condition we want to impose on the norm $\|\cdot\|$ is monotonicity with respect to the magnitude of $z_j^{\mathcal{B}}$. A special case is the Mahalanobis distance.

## 4.2 Minimax Rate of Adaptation

Using this solution to the inverse modulus, we derive the rate of convergence of the between class of modulus, which characterizes how fast the worst-case expected length of the adaptive CIs can go to 0 as $n \to \infty$. We derive the rates under the assumption that the sequence of design points $\{x_i\}_{i=1}^{\infty}$ is a realization of a sequence of independent and identically distributed random vectors $\{X_i\}_{i=1}^{\infty}$ drawn from a distribution that satisfies some mild regularity conditions. This gives an intuitive restriction on the design points, and also shows that the result applies under random design points as well.[6] Define $r(\gamma_1, \gamma_2) = (2 + k_+/\gamma_1 + (k - k_+)/\gamma_2)^{-1}$. The following theorem fully characterizes the minimax rate of adaptation.

[6]Consider the model $y_i = f(X_i) + \varepsilon_i$, for $i = 1, \ldots, n$, with the $X_i \overset{\text{i.i.d.}}{\sim} p_X$ with $\varepsilon_i | X_i \sim N(0, \sigma^2(X_i))$. Then, conditional on $\{X_i\}_{i=1}^n = \{x_i\}_{i=1}^n$, this model is equivalent with our model.

**Theorem 4.1.** *Let $\{X_i\}_{i=1}^{\infty}$ be an i.i.d. sequence of random vectors with support $\mathcal{X}$. Suppose $X_i$ admits a probability density function $p_X(\cdot)$ that is continuous at 0 with $p_X(0) > 0$, and assume $\sigma(\cdot) = 1$. Then, for almost all realizations $\{x_i\}_{i=1}^{\infty}$ of $\{X_i\}_{i=1}^{\infty}$ and for all $\delta > 0$, we have*

$$\lim_{n\to\infty} n^{r(\gamma_1,\gamma_2)}\omega\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right) = \left(\delta^2/c_{1,2}^*\right)^{r(\gamma_1,\gamma_2)}, \text{ and}$$
$$\lim_{n\to\infty} n^{r(\gamma_1,\gamma_2)}\omega\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right), \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right)\right) = \left(\delta^2/c_{2,1}^*\right)^{r(\gamma_1,\gamma_2)},$$

*where $c_{1,1}^*$ and $c_{2,1}^*$ are constants that depend only on the function spaces.*

**Remark 4.2.** The result immediately implies the rate of convergence for the between class modulus

$$\begin{aligned}&\lim_{n\to\infty} n^{r(\gamma_1,\gamma_2)}\omega_+\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right)\\ =&\max\left\{\delta^2/c_{1,2}^*, \delta^2/c_{2,1}^*\right\}^{r(\gamma_1,\gamma_2)}.\end{aligned}$$

Hence, if a CI maintains coverage over $\Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)$, the best possible worst-case length of this CI over $\Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)$ goes to 0 at the same rate as $n^{-r(\gamma_1,\gamma_2)}$.

**Remark 4.3** (Heteroskedasticity)**.** For simplicity, the theorem imposes a homoskedasticity condition (i.e., $\sigma(\cdot) = 1$). However, allowing for general $\sigma(\cdot)$ is straightforward and requires only weak regularity conditions on $\sigma(\cdot)$. See Appendix C for details.

Theorem 4.1 shows how the monotonicity restriction plays a role in determining the minimax rates of adaptation to Hölder coefficients under the multivariate nonparametric regression setting. When $k_+ = k$, the minimax rate of adaptation is $n^{-\frac{1}{2+k/\gamma_1}}$, which equals the minimax convergence rate over $\omega\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right)\right)$. This shows that strong adaptation is possible if the regression function is monotone with respect to all the variables, just like in the univariate case. On the other hand, when $k_+ = 0$, the rate becomes $n^{-\frac{1}{2+k/\gamma_2}}$, consistent with the previous findings that there is no scope of adaptation for general Hölder classes without any shape constraint. Importantly, Theorem 4.1 characterizes the convergence rate for the case where $0 < k_+ < k$, where it gives an intuitive intermediate rate between the two extreme.

## 4.3 Construction of the Adaptive CI

Here, we give the explicit formula of the CIs for our parameters of interest, now that we have derived the form of the moduli of continuity and the solutions to the modulus problems in the previous section. We first consider $L_0 f$. Before stating the result, it is convenient to define the following functions

$$\begin{aligned} D_{Jj,\delta}(x_i) :=& (\omega(\delta, \mathcal{F}_J, \mathcal{F}_j) - C_j \|(x_i)_{\mathcal{V}-}\|^{\gamma_j} - C_J \|(x_i)_{\mathcal{V}+}\|^{\gamma_J})_+, \text{ and} \\ D_{jJ,\delta}(x_i) :=& (\omega(\delta, \mathcal{F}_j, \mathcal{F}_J) - C_J \|(x_i)_{\mathcal{V}-}\|^{\gamma_J} - C_j \|(x_i)_{\mathcal{V}+}\|^{\gamma_j})_+. \end{aligned}$$

**Corollary 4.2.** *For $Lf = f(0)$ and $\beta = 1/2$, the lower CI defined in (3) is given by*

$$\hat{c}_{\delta}^{\ell,j} = \hat{L}_{\delta}^{\ell,j} - \frac{1}{2}\left(\omega\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right) + \frac{\delta^2}{\sum_{i=1}^n D_{Jj,\delta}\left(x_i\right)}\right),$$

*where*

$$\begin{aligned} \hat{L}_{\delta}^{\ell,j} =& \frac{\sum_{i=1}^n D_{Jj,\delta}\left(x_i\right) y_i}{\sum_{i=1}^n D_{Jj,\delta}\left(x_i\right)} + \frac{\omega\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right)}{2} \\ &- \frac{\sum_{i=1}^n D_{Jj,\delta}\left(x_i\right)\left(\omega\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right) - C_j \left\|(x_i)_{\mathcal{V}-}\right\|^{\gamma_j} + C_J \left\|(x_i)_{\mathcal{V}+}\right\|^{\gamma_J}\right)}{2\sum_{i=1}^n D_{Jj,\delta}\left(x_i\right)}. \end{aligned}$$

*Likewise, the upper bound of the CI is given by*

$$\hat{c}_{\delta}^{u,j} = \hat{L}_{\delta}^{u,j} + \frac{1}{2}\left(\omega\left(\delta, \mathcal{F}_j, \mathcal{F}_J\right) + \frac{\delta^2}{\sum_{i=1}^n D_{jJ,\delta}\left(x_i\right)}\right),$$

*where*

$$\begin{aligned} \hat{L}_{\delta}^{u,j} =& \frac{\sum_{i=1}^n D_{jJ,\delta}\left(x_i\right) y_i}{\sum_{i=1}^n D_{jJ,\delta}\left(x_i\right)} + \frac{\omega\left(\delta, \mathcal{F}_j, \mathcal{F}_J\right)}{2} \\ &- \frac{\sum_{i=1}^n D_{jJ,\delta}\left(x_i\right)\left(\omega\left(\delta, \mathcal{F}_j, \mathcal{F}_J\right) - C_J \left\|(x_i)_{\mathcal{V}-}\right\|^{\gamma_J} + C_j \left\|(x_i)_{\mathcal{V}+}\right\|^{\gamma_j}\right)}{2\sum_{i=1}^n D_{jJ,\delta}\left(x_i\right)}. \end{aligned}$$

The first terms in the formula of $\hat{L}_{\delta}^{\ell,j}$ and $\hat{L}_{\delta}^{u,j}$ are the random terms linear in $y_i$ while the remaining terms are non-random fixed terms. If $\mathcal{V} = \{1, ..., k\}$ (so the function is monotone in every coordinate), the random terms can be viewed as a kernel estimator with a data-dependent bandwidth. Too see this, if we define

$$k(x) = \left[1 - C_j \left\|(x)_{\mathcal{V}-}\right\|^{\gamma_j} - C_J \left\|(x)_{\mathcal{V}+}\right\|^{\gamma_J}\right]_+,$$

and

$$h_{mn}(x) = \begin{cases} \omega(\delta, \mathcal{F}_J, \mathcal{F}_j)^{1/\gamma_J} & \text{if the } m\text{th coordinate of } x \geq 0 \\ \omega(\delta, \mathcal{F}_J, \mathcal{F}_j)^{1/\gamma_j} & \text{otherwise,} \end{cases}$$

we have

$$\frac{\sum_{i=1}^n D_{Jj,\delta}(x_i)\, y_i}{\sum_{i=1}^n D_{Jj,\delta}(x_i)} = \frac{\sum_{i=1}^n k\left(x_{1i}/h_{1n}(x_i), ..., x_{ki}/h_{kn}(x_i)\right) y_i}{\sum_{i=1}^n k\left(x_{1i}/h_{1n}(x_i), ..., x_{ki}/h_{kn}(x_i)\right)}.$$

Hence, the CI can be considered to be based on a Nadaraya-Watson type estimator, correcting for the bias.

As described in Section 3.2, the proposed CI is given by $\cap_{j=1}^J [\hat{c}^{\ell,j}_{z_{1-\tau^*}}, \hat{c}^{\ell,j}_{z_{1-\tau^*}}]$, where $\tau^*$ is defined in Lemma 3.2. Here, we show that the distribution of $(V(\tau)', W(\tau)')'$ does not depend on $\tau$ as $n \to \infty$. The implication of this invariance with respect to $\tau$, is that calculating $\tau^*$ boils down to calculating the quantile of the maximum of Gaussian vectors. The variance matrix of this limiting Gaussian random vector is known, and thus the said quantile can be easily simulated. Moreover, when $\gamma_1 = \cdots = \gamma_J$ so that the parameters spaces differs only in $C_j$, $\tau^*$ can be shown to be bounded away from zero by a constant that does not depend on $J$, for large $n$. Hence, the constant of the CI does not grow to infinity as $J \to \infty$ in this case.[7]

**Lemma 4.4.** *Under the same set of conditions of Theorem 4.1, $(V(\tau)', W(\tau)')' \xrightarrow{d} (V_\infty', W_\infty')'$ as $n \to \infty$, where $(V_\infty', W_\infty')'$ is a Gaussian random vector that does not depend on $\tau$. Moreover, if $\gamma_1 = \cdots = \gamma_J$, then, for large $n$, $\tau^* > \eta$ for some $\eta > 0$ that does not depend on $J$.*

**Remark 4.4** (Dependence on $J$)**.** The proof reveals that when all $J$ parameter spaces correspond to different Hölder exponents (i.e., $\gamma_1 > \cdots > \gamma_J$), the dependence of $\tau^*$ on $J$ does not vanish and in fact results in CIs whose constants grow at the same rate as the naive Bonferroni CI, $(\log J)^{1/2}$. However, some finite sample improvement in terms of the length of the resulting CI compared to the naive Bonferroni CI is

[7] This is especially useful when one wishes to adapt to $C$ while keeping $\gamma$ fixed. For example, Kwon and Kwon (2020) take $\gamma_j = 1$ and consider the problem of adapting to the Lipschitz constant in a regression discontinuity setting.

shown in the empirical exercise. When some of the parameter spaces have the same Hölder exponent, the improvement can be significant. As an extreme case, when $\gamma_1 = \cdots = \gamma_J$, $\tau^*$ can is bounded away from 0 by a constant that does not depend on $J$, which is exactly what the second part of the lemma asserts.

# 5 Simulation Results

In this section, we compare the performances of the adaptive CI of Cai and Low (2004) and the adaptive CI constructed using the naive Bonferroni procedure described in Section 3.2. As a benchmark, we also provide the lengths of the shortest fixed length confidence intervals of Donoho (1994), referred to as minimax CIs. We consider inference for $f(0)$, given some regression function $f$. We consider the case where the researcher is uncertain about the value of the Hölder exponent $\gamma$, and thus tries to adapt to its value.

First, we construct adaptive CIs with respect to two smoothness parameters $(\gamma_1, \gamma_2) = (1, 10^{-3})$ while fixing $C = 1$, which gives $J = 2$. We vary $n$ over $\{10^2,\ 5 \times 10^2,\ 10^3,\ 5 \times 10^3,\ 10^4\}$ to investigate the rate of adaptation as the sample size grows. The true regression function is over $\mathbb{R}^2$ and given by either $f_1$ or $f_2$, defined as

$$f_1(x_1, x_2) = 0, \quad f_2(x_1, x_2) = \left\|(x_1, x_2)_{\mathcal{V}+}\right\|_2^{\gamma_2}, \quad \mathcal{V} = \{1, 2\}.$$

By construction, we have $f_j \in \Lambda_{+,\mathcal{V}}(\gamma_j, 1)$. The covariates are drawn from a uniform distribution over $[-1/(2\sqrt{2}), 1/(2\sqrt{2})]^2$, and the noise terms, $\{u_i\}_{i=1}^n$, are drawn from a standard normal distribution. The outcome variable is given as $y_i = f(x_i) + u_i$, for $f \in \{f_1, f_2\}$. We fix the draw of $\{x_i\}_{i=1}^n$ within each simulation iteration. We run 500 iterations to calculate the average lengths and coverage probabilities of CIs. The nominal coverage probability is .95 for all CIs.

Table 1 shows the results for the case where $f = f_1$. Each column corresponds to 1) our proposed (naive) Bonferroni adaptive procedure (AdaptBonf), 2) the adaptive CI of Cai and Low (2004) (CL, henceforth), 3) the minimax CI with respect to $\Lambda_{+,\mathcal{V}}(\gamma_2, 1)$, and 4) the minimax CI with respect to $\Lambda_{+,\mathcal{V}}(\gamma_1, 1)$. Regarding the last two minimax procedures, we refer to them as the "conservative minimax CI" and the "oracle minimax CI", respectively. Note that the oracle minimax CI is an optimal

Table 1: Lengths of CIs when $f = f_1$ with $J = 2$

| | AdaptBonf | CL | Minimax ($\gamma = \gamma_2$) | Minimax ($\gamma = \gamma_1$) |
|---|---|---|---|---|
| n = 100 | 0.925 | 5.092 | 1.459 | 0.617 |
| n = 500 | 0.478 | 1.857 | 1.212 | 0.391 |
| n = 1,000 | 0.382 | 1.530 | 1.150 | 0.329 |
| n = 5,000 | 0.256 | 1.037 | 1.064 | 0.222 |
| n = 10,000 | 0.212 | 0.868 | 1.045 | 0.187 |

Table 2: Lengths of CIs when $f = f_2$ with $J = 2$

| | AdaptBonf | CL | Minimax ($\gamma = \gamma_2$) |
|---|---|---|---|
| n = 100 | 1.599 | 5.092 | 1.459 |
| n = 500 | 1.279 | 3.504 | 1.212 |
| n = 1,000 | 1.197 | 3.222 | 1.150 |
| n = 5,000 | 1.086 | 2.860 | 1.064 |
| n = 1,0000 | 1.060 | 2.703 | 1.045 |

Table 3: Coverage probabilities of adaptive CIs ($J = 2$)

| | $f = f_1$ | | $f = f_2$ | |
|---|---|---|---|---|
| | AdaptBonf | CL | AdaptBonf | CL |
| n = 100 | 0.988 | 1.000 | 0.992 | 1.000 |
| n = 500 | 0.972 | 1.000 | 0.976 | 0.998 |
| n = 1,000 | 0.970 | 1.000 | 0.982 | 1.000 |
| n = 5,000 | 0.972 | 1.000 | 0.970 | 1.000 |
| n = 10,000 | 0.974 | 1.000 | 0.976 | 1.000 |

benchmark, which is only feasible when we actually know the true regression function is in the smaller parameter space $\Lambda_{+,\mathcal{V}}(\gamma_1, 1)$. In Table 1, the average lengths of both adaptive confidence intervals decrease considerably as $n$ increases from 100 to 10,000. In comparison, the length of the conservative minimax CI (column 3) decreases only about 28% for the same change in the sample size. This shows the lengths of the adaptive confidence intervals decrease more sharply when the true function is smooth, as predicted by the theory.

To compare the performances of different adaptive inference procedures, note that the average lengths of the CI of CL adapting to the Hölder exponents (column 2) are often wider than the conservative minimax CI (column 3). When $n = 100$, the former is more than three times wider than the latter, and the adaptive procedure starts to dominate the minimax procedure only when $n$ is greater than 5,000. In comparison, our proposed Bonferroni adaptive procedure (column 1) yields shorter CIs than those by CL, as predicted in Section 3.2. To compare the Bonferroni adaptive CI with the conservative minimax CI, the lengths of the former are always exceeded by those of the minimax CI, even for the relatively small sample size of $n = 100$. Moreover, the length of the adaptive CI becomes only 20% of the length of the conservative minimax CI for the sample size of $n = 10^4$. The Bonferroni procedure also performs well even when compared to the infeasible oracle minimax CI (column 4), with the length of the former only 13% wider than the latter when $n = 10^4$. This demonstrates the strong adaptivity property of the adaptive procedure when the regression function is monotone with respect to all variables, as shown in Section 4.2.

Table 2 demonstrates the analogous simulation results when $f = f_2$. In this case, the minimax CI with respect to $\Lambda_{+,\mathcal{V}}(\gamma_2, 1)$ (column 3) is referred to as the oracle minimax CI. While the lengths of the oracle minimax procedure are considerably shorter than the CIs of CL for various values of $n$, the performance of the Bonferroni CIs almost matches that of the oracle minimax procedure. Especially, the performance of the Bonferroni adaptive procedure becomes extremely close to the oracle minimax procedure when $n$ is greater than 500.

Table 3 shows the coverage probabilities of adaptive CIs for both of the cases when $f = f_1$ and $f = f_2$. While all the CIs achieve the correct coverage, none of those CIs exactly achieves the nominal coverage of .95, reflecting the conservative nature of the adaptive CIs. We can see that the adaptive procedure of CL is particularly conservative, almost always yielding 100% coverage probabilities.

Table 4: Lengths of CIs when $f = f_3$ with $J = 6$

| | AdaptBonf | CL | Minimax ($\gamma = \gamma_6$) | Minimax ($\gamma = 0.5$) |
|---|---|---|---|---|
| n = 100 | 1.495 | 5.092 | 1.459 | 0.908 |
| n = 500 | 0.972 | 2.417 | 1.212 | 0.649 |
| n = 1,000 | 0.833 | 2.441 | 1.150 | 0.580 |
| n = 5,000 | 0.615 | 1.816 | 1.064 | 0.447 |
| n = 10,000 | 0.543 | 1.145 | 1.045 | 0.400 |

So far we considered adapting to the smoothness parameters at two extremes, $\gamma \in (0.001, 1)$. Since the multiplicative constant for the Bonferroni procedure increases with $J$, a concern is that the performance of the Bonferroni procedure relative to the CL procedure might get worse when $J$ is larger. To investigate the possibility, we consider adapting to a wider set of parameters, $\{\gamma_j\}_{j=1}^6$, where $\gamma_j = 1 - (j-1)/5$ for $j = 1, ..., 5$ and $\gamma_6 = 10^{-3}$. Moreover, rather than taking the extreme value of $\gamma$ as the true parameter, we consider the case where $\gamma$ takes an intermediate value, $\gamma = 1/2$. The true regression function is given by

$$f_3(x_1, x_2) = \left\|(x_1, x_2)_{\mathcal{V}+}\right\|_2^{1/2}, \quad \mathcal{V} = \{1, 2\},$$

so that $f_3 \in \Lambda_{+,\mathcal{V}}(1/2, 1)$.

Table 4 displays the simulation results corresponding to this specification. Each column corresponds to 1) our proposed Bonferroni adaptive procedure, 2) the adaptive CI of CL, 3) the minimax CI with respect to $\Lambda_{+,\mathcal{V}}(\gamma_6, 1)$, and 4) the minimax CI with respect to $\Lambda_{+,\mathcal{V}}(1/2, 1)$. As before, we refer to the last two CIs as the conservative minimax CI and the oracle minimax CI, respectively. We observe the same pattern as in the case of adapting to two parameters—adaptive CIs shrink faster than the conservative minimax CI as the sample size increases, and the Bonferroni adaptive CIs are shorter than the ones of CL. While the ratio of the length of the Bonferroni CI to that of the CI of CL is larger in this case compared to the case where $J = 2$, especially when $n$ is large, the Bonferroni CI is still more than 50 % narrower than the CI of CL, and not much wider than the oracle minimax CI.

Table 5: Summary statistics - Chinese chemical industry dataset for the year 2001

| | Min | Mean | Median | Max |
|---|---|---|---|---|
| Log output | 6.472 | 9.952 | 9.937 | 13.233 |
| Log fixed asset | 7.463 | 10.818 | 10.759 | 14.226 |
| Log labor | 3.664 | 6.352 | 6.386 | 9.142 |

# 6 Empirical Illustration

In this section, we apply our procedure to the production function estimation problem for the Chinese chemical industry. Specifically, we use the firm-level data of Jacho-Chávez et al. (2010) for the year 2001, which was also used by Horowitz and Lee (2017) to illustrate their method of constructing the uniform confidence band for the production function under shape restrictions.

In the dataset, the dependent variable is the logarithm of value-added real output ($y$), and the explanatory variables are the logarithms of the net value of the real fixed asset ($k$) and the number of employees ($\ell$). After removing the outliers for $y, k$ and $\ell$, the remaining sample size was $n = 1,636$.[8] Table 5 shows the brief summary of the variables used in our analysis. We are interested in construction of the confidence interval for $f(k_0, \ell_0) := \mathbf{E}[y|k = k_0, \ell = \ell_0]$. We take $(k_0, \ell_0)$ to be medians of each variable.

The first step is to estimate the variance of the error term. We assume homoskedastic errors for simplicity. The variance estimator is defined as

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (y_i - \hat{r}(k_i, \ell_i))}{n - 2\nu_1 + \nu_2},$$

where $\hat{r}(k_i, \ell_i)$ is the estimator for the conditional mean using kernel regression, $\nu_1 = \mathrm{tr}(L)$, $\nu_2 = \mathrm{tr}(L'L)$, where $L$ is the weight matrix for the kernel estimator. Refer to Wassermann (2006) for a justification for this variance estimator. We used the Gaussian kernel with the bandwidth chosen by expected Kullback-Leibler cross validation as in Hurvich et al. (1998).

For the function space, we consider adapting to a sequence of parameter spaces

[8] We used the conventional way of outlier detection, removing the observations that are greater than the third quantile plus IQR times 1.5, or less than the first quantile minus IQR times 1.5. Our resulting sample size is close to Horowitz and Lee (2017), who have $n = 1,638$.

Table 6: 95% confidence intervals for $f(k_0, \ell_0)$

| | CI | Length |
|---|---|---|
| Minimax ($\gamma = \gamma_6$) | [7.501, 13.005] | 5.504 |
| Minimax ($\gamma = \gamma_1$) | [9.922, 10.484] | 0.562 |
| AdaptBonf (Naive) | **[9.766, 11.264]** | **1.498** |
| AdaptBonf (Calibrated) | **[9.864, 11.188]** | **1.324** |
| Cai and Low | [7.134, 12.049] | 4.915 |

$\{\Lambda_{+,\mathcal{V}}(\gamma_j, C)\}_{j=1}^{6}$ with $\gamma_j = 1 - (j-1)/5$ for $j = 1, ..., 5$ and $\gamma_6 = 10^{-3}$. We take $\mathcal{V} = \{1, 2\}$, assuming that the production function is nondecreasing in both fixed assets and labor, which is consistent with economic theory. To make $\Lambda_{+,\mathcal{V}}(\gamma_j, C) \subset \Lambda_{+,\mathcal{V}}(\gamma_6, C)$ hold for all $j = 1, ..., 5$, we only use observations in a restricted support, and the effective sample size is given by $n_{\text{eff}} = 272$.

For the norm, we use the Euclidean norm weighted by the inverse of the standard deviation of each input, $\|(k, \ell)\| = ((k/s_k)^2 + (\ell/s_\ell)^2)^{1/2}$ where $s_k$ and $s_\ell$ are standard deviations of $k$ and $\ell$, respectively. We take conservative values of $C$ by setting

$$C = 2 \times \max_{(i,j) \in \{1,...,n_{\text{eff}}\}^2} \frac{|y_j - y_i|}{\|(k_j, \ell_j) - (k_i, \ell_i)\|^{\gamma_6}}.$$

We compare different procedures to construct CIs. The methods in comparison are the minimax CI with respect to the largest space $\Lambda_{+,\mathcal{V}}(\gamma_6, C)$ (row 1), the restricted minimax CI with respect to the smallest space $\Lambda_{+,\mathcal{V}}(\gamma_1, C)$ (row 2), the adaptive Bonferroni CI adapting to $\{\gamma_j\}_{j=1}^{6}$ (row 3), the same adaptive CI, but taking into account the correlations between different CIs (fourth row), and the adaptive CI of Cai and Low (2004) (henceforth CL) adapting to $\{\gamma_j\}_{j=1}^{6}$ (fifth row). Note that all the CIs maintain correct coverage over the largest space $\Lambda_{+,\mathcal{V}}(\gamma_6, C)$, except for the second one, which is valid only over the smallest space $\Lambda_{+,\mathcal{V}}(\gamma_1, C)$. We refer to the first minimax CI as the conservative minimax CI.

Table 6 demonstrates the 95% confidence intervals for $f(k_0, \ell_0)$ produced by different inference methods. First of all, the lengths of the adaptive Bonferroni CIs are much shorter than the conservative minimax CI, while the procedure of CL yields a wider CI, almost as long as the conservative minimax CI. We can also observe that the adaptive Bonferroni CI using the calibrated value of $\tau^*$ (fourth row) is relatively narrower than its naive version taking $\tau = 0.05/2J$ (third row). Lastly, while the length of the second minimax procedure (second row) is the shortest, it is only valid when we are confident that the true regression function is in the smallest function space we consider, $\Lambda_{+,\mathcal{V}}(\gamma_1, C)$. Together with the simulation results in the previous section, our empirical analysis demonstrates the advantage of using an adaptive procedure when the monotonicity restriction is plausible as well as good finite sample performance of our proposed Bonferroni adaptive procedure.

# References

Armstrong, T. (2015): "Adaptive Testing on a Regression Function at a Point," *The Annals of Statistics*, 43, 2086–2101.

Armstrong, T. B. and M. Kolesár (2016): "Optimal inference in a class of regression models," *working paper*.

——— (2018): "Optimal Inference in a Class of Regression Models," *Econometrica*, 86, 655–683.

Beliakov, G. (2005): "Monotonicity Preserving Approximation of Multivariate Scattered Data," *BIT Numerical Mathematics*, 45, 653–677.

Cai, T. T. (2012): "Minimax and Adaptive Inference in Nonparametric Function Estimation," *Statistical Science*, 27, 31–50.

Cai, T. T., M. Low, and Z. Ma (2014): "Adaptive Confidence Bands for Nonparametric Regression Functions," *Journal of the American Statistical Association*, 109, 1054–1070.

Cai, T. T. and M. G. Low (2004): "An Adaptation Theory for Nonparametric Confidence Intervals," *The Annals of statistics*, 32, 1805–1840.

——— (2006): "Adaptive Confidence Balls," *Annals of Statistics*, 34, 202–228.

Cai, T. T., M. G. Low, and Y. Xia (2013): "Adaptive Confidence Intervals for Regression Functions under Shape Constraints," *Annals of Statistics*, 41, 722–750.

Deng, H., Q. Han, and C.-H. Zhang (2020): "Confidence Intervals for Multiple Isotonic Regression and Other Monotone Models," *arXiv:2001.07064 [math, stat]*.

Donoho, D. L. (1994): "Statistical Estimation and Optimal Recovery," *Annals of Statistics*, 22, 238–270.

Dümbgen, L. (1998): "New Goodness-of-Fit Tests and Their Application to Nonparametric Confidence Sets," *The Annals of Statistics*, 26, 288–314.

Genovese, C. and L. Wasserman (2008): "Adaptive Confidence Bands," *The Annals of Statistics*, 36, 875–905.

Genovese, C. R. and L. Wasserman (2005): "Confidence Sets for Nonparametric Wavelet Regression," *Annals of Statistics*, 33, 698–729.

Giné, E. and R. Nickl (2010): "Confidence Bands in Density Estimation," *Annals of Statistics*, 38, 1122–1170.

Han, Q., T. Wang, S. Chatterjee, and R. J. Samworth (2019): "Isotonic Regression in General Dimensions," *The Annals of Statistics*, 47, 2440–2471.

Hengartner, N. W. and P. B. Stark (1995): "Finite-Sample Confidence Envelopes for Shape-Restricted Densities," *Annals of Statistics*, 23, 525–550.

Hoffmann, M. and R. Nickl (2011): "On Adaptive Inference and Confidence Bands," *Annals of Statistics*, 39, 2383–2409.

Horowitz, J. L. and S. Lee (2017): "Nonparametric Estimation and Inference Under Shape Restrictions," *Journal of Econometrics*, 201, 108–126.

Hurvich, C. M., J. S. Simonoff, and C.-L. Tsai (1998): "Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60, 271–293.

Jacho-Chávez, D., A. Lewbel, and O. Linton (2010): "Identification and Nonparametric Estimation of a Transformed Additively Separable Model," *Journal of Econometrics*, 156, 392–407.

Kwon, K. and S. Kwon (2020): "Inference in Regression Discontinuity Designs under Monotonicity," *working paper*.

Lepskii, O. V. (1991): "On a Problem of Adaptive Estimation in Gaussian White Noise," *Theory of Probability & Its Applications*, 35, 454–466.

Low, M. G. (1997): "On Nonparametric Confidence Intervals," *The Annals of Statistics*, 25, 2547–2554.

Robins, J. and A. van der Vaart (2006): "Adaptive Nonparametric Confidence Sets," *Annals of Statistics*, 34, 229–253.

van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer Science & Business Media.

Wassermann, L. (2006): *All of Nonparametric Statistics*, Springer-Verlag New York.

# Appendix A Proofs of Lemmas

## A.1 Proof of Corollary E.1

*Proof.* Suppose $\hat{c}^{\ell,j}_{\alpha,-}$ solves

$$\min_{\hat{c}:[\hat{c},\infty)\in\mathcal{I}^J_{\alpha,1,-}} \sup_{f\in\mathcal{F}_j} q_{\beta,f}(-Lf-\hat{c}), \tag{11}$$

where $\mathcal{I}^J_{\alpha,1,-}$ denotes the set of one-sided CIs that covers $-Lf$ with probability at least $1-\alpha$ over $\mathcal{F}_J$. Then, taking $\hat{c}^{u,J}_{\alpha} = -\hat{c}^{\ell,J}_{\alpha,-}$, we have $\left(-\infty, \hat{c}^{u,J}_{\alpha}\right] \in \mathcal{I}^J_{\alpha,1}$ and $\hat{c}^{u,J}_{\alpha}$ solves (5). Applying Theorem 3.1 of Armstrong and Kolesár (2018) with $\widetilde{L}f = -Lf$, we get the desired result. $\square$

## A.2 Proof of Theorem 3.1

*Proof.* Consider the CI $[\hat{c}^{u,j}_{\alpha/2J}, \hat{c}^{\ell,j}_{\alpha/2J}]$, and observe that

$$\mathbf{E}[\hat{c}^{u,j}_{\alpha/2J} - \hat{c}^{\ell,j}_{\alpha/2J}] = \mathbf{E}[\hat{c}^{u,j}_{\alpha/2J} - Lf] + \mathbf{E}[Lf - \hat{c}^{\ell,j}_{\alpha/2J}],$$

for any $f\in\mathcal{F}_j$. Then, by (4) and (E.1), we have

$$\sup_{f\in\mathcal{F}_j} \mathbf{E}[\hat{c}^{u,j}_{\alpha/2J} - \hat{c}^{\ell,j}_{\alpha/2J}] \le \omega(z_{1-\frac{\alpha}{2J}}, \mathcal{F}_J, \mathcal{F}_j) + \omega(z_{1-\frac{\alpha}{2J}}, \mathcal{F}_j, \mathcal{F}_J) \le 2\omega_+(z_{1-\frac{\alpha}{2J}}, \mathcal{F}_j, \mathcal{F}_J).$$

It follows that

$$\begin{aligned}
\sup_{f\in\mathcal{F}_j} \mathbf{E}\,\mu(CI^{Bon,J}_{\alpha}) &= \sup_{f\in\mathcal{F}_j} \mathbf{E}[\min_j \hat{c}^{u,j}_{\alpha/2J} - \max_j \hat{c}^{\ell,j}_{\alpha/2J}] \\
&\le \sup_{f\in\mathcal{F}_j} \mathbf{E}[\hat{c}^{u,j}_{\alpha/2J} - \hat{c}^{\ell,j}_{\alpha/2J}] \\
&\le 2\omega_+(z_{1-\frac{\alpha}{2J}}, \mathcal{F}_j, \mathcal{F}_J)
\end{aligned}$$

for any $j = 1, ..., J$. Noting that

$$2\omega_+(z_{1-\frac{\alpha}{2J}}, \mathcal{F}_j, \mathcal{F}_J) \le \frac{2z_{1-\frac{\alpha}{2J}}}{z_{1-\frac{\alpha}{2}}}\omega_+(z_{1-\frac{\alpha}{2}}, \mathcal{F}_j, \mathcal{F}_J),$$

which follows from the concavity of the ordered modulus of continuity, we obtain the desired result. $\square$

### A.3 Proof of Lemma 3.2

*Proof.* First, note that we can write

$$\begin{aligned}
&\mathbf{P}\left(Lf < \hat{c}_\tau^{L,j}\right)\\
=&\,\mathbf{P}\left(\hat{c}_\tau^{L,j} - Lf > 0\right)\\
=&\,\mathbf{P}\left(\frac{\hat{c}_\tau^{L,j} - Lf}{\omega'\left(z_{1-\tau}, \Lambda_{+,\mathcal{V}}\left(\gamma, C_J\right), \Lambda_{+,\mathcal{V}}\left(\gamma, C_j\right)\right)} > 0\right)\\
=&\,\mathbf{P}\left(\frac{\hat{c}_\tau^{L,j} - Lf}{\omega'\left(z_{1-\tau}, \Lambda_{+,\mathcal{V}}\left(\gamma, C_J\right), \Lambda_{+,\mathcal{V}}\left(\gamma, C_j\right)\right)} + z_{1-\tau} > z_{1-\tau}\right)\\
\equiv&\,\mathbf{P}(\widetilde{V}_j(\tau) > z_{1-\tau}).
\end{aligned}$$

Likewise, we can write

$$\begin{aligned}
&\mathbf{P}\left(Lf > \hat{c}_\tau^{U,j}\right)\\
=&\,\mathbf{P}\left(\frac{Lf - \hat{c}_\tau^{U,j}}{\omega'\left(z_{1-\tau}, \Lambda_{+,\mathcal{V}}\left(\gamma, C_j\right), \Lambda_{+,\mathcal{V}}\left(\gamma, C_J\right)\right)} + z_{1-\tau} \geq z_{1-\tau}\right)\\
\equiv&\,\mathbf{P}(\widetilde{W}_j(\tau) > z_{1-\tau}).
\end{aligned}$$

Therefore, writing $\widetilde{V}(\tau) = \left(\widetilde{V}_1(\tau), ..., \widetilde{V}_J(\tau)\right)'$ and similarly for $\widetilde{W}(\tau)$, we have

$$\mathbf{P}(Lf \notin CI_\tau^{\mathcal{J}}) = \mathbf{P}(\max\{\widetilde{V}(\tau)', \widetilde{W}(\tau)'\} > z_{1-\tau}).$$

Now, we want to find an upper bound on

$$\sup_{f \in \mathcal{F}_J} \mathbf{P}(\max\{\widetilde{V}(\tau)', \widetilde{W}(\tau)'\} > z_{1-\tau}).$$

Note that the quantile of $\max\{\widetilde{V}(\tau)', \widetilde{W}(\tau)'\}$ is increasing in the mean of each $\widetilde{V}_j(\tau)$'s and $\widetilde{W}_j(\tau)$'s. Moreover, the variances and covariances of $(\widetilde{V}(\tau)', \widetilde{W}(\tau)')'$ do not depend on the true regression function $f$, by the construction of $\hat{c}_\tau^{L,j}$ and $\hat{c}_\tau^{U,j}$. Therefore, it is useful to consider $\sup_{f \in \mathcal{F}_J} \mathbf{E}\,\widetilde{V}_j(\tau)$ and $\sup_{f \in \mathcal{F}_J} \mathbf{E}\,\widetilde{W}_j(\tau)$. Actually, Lemma A.1

in AK can be used to show

$$\sup_{f\in\mathcal{F}_J} \mathbf{E}\,\widetilde{V}_j(\tau) = \sup_{f\in\mathcal{F}_J} \mathbf{E}\,\widetilde{W}_j(\tau) = 0.$$

Moreover, it is straightforward to show that the variance matrix of $(\widetilde{V}(\tau)', \widetilde{W}(\tau)')'$ is given by the formula in the statement of Lemma 3.2. Therefore, we have

$$\sup_{f\in\mathcal{F}_J} \mathbf{P}(\max\{\widetilde{V}(\tau)', \widetilde{W}(\tau)'\} > z_{1-\tau}) \leq \mathbf{P}(\max\{V(\tau)', W(\tau)'\} > z_{1-\tau}),$$

and by setting $\tau^*$ so that the latter term becomes $\alpha$, we get the desired result. □

## A.4 Proof of Lemma 4.1

*Proof.* First, we note that $h(x) \equiv C\,\|x\|^\gamma$ satisfies the Hölder continuity condition. This is because for any $x, z \in \mathbb{R}^k$, such that (without loss of generality) $\|x\| \geq \|z\|$, we have

$$|h(x) - h(z)| = C\left(\|x\|^\gamma - \|z\|^\gamma\right) \leq C\,\|x - z\|^\gamma .$$

The inequality holds because we have

$$\|x\|^\gamma \leq \left(\|x - z\| + \|z\|\right)^\gamma ,$$

by the triangle inequality, and thus

$$\|x - z\| + \|z\| \leq \left(\|x - z\|^\gamma + \|z\|^\gamma\right)^{1/\gamma} ,$$

using the fact that $\gamma \in (0, 1]$

Next, we show that $h_+(x) \equiv C\|(x)_{\mathcal{V}+}\|^\gamma$ also satisfies Hölder continuity. For $x, z \in \mathbb{R}^k$, define $\widetilde{x} = (x)_{\mathcal{V}+}$ and $\widetilde{z} = (z)_{\mathcal{V}_+}$. Then, we can see that $\|x - z\| \geq \|\widetilde{x} - \widetilde{z}\|$ , since $|x_m - z_m| \geq |\widetilde{x}_m - \widetilde{z}_m|$ for $m \in \mathcal{V}$ and $|x_m - z_m| = |\widetilde{x}_m - \widetilde{z}_m|$ otherwise. Therefore, for any $x, z \in \mathbb{R}^k$ with $x \neq z$, we have

$$\frac{|h_+(x) - h_+(z)|}{\|x - z\|^\gamma} = \frac{|h(\widetilde{x}) - h(\widetilde{z})|}{\|x - z\|^\gamma} \leq \frac{|h(\widetilde{x}) - h(\widetilde{z})|}{\|\widetilde{x} - \widetilde{z}\|^\gamma} \leq C,$$

where the last inequality follows from the Hölder continuity of $h$.

Lastly, for monotonicity, note that for any $x, z \in \mathbb{R}^k$ such that $z_i \geq x_i$ for some

$i \in \mathcal{V}$ and $z_j = x_j$ for all $j \neq i$, we have $|\widetilde{z}_i| \geq |\widetilde{x}_i|$. Therefore, we have $h_+(z) \geq h_+(x)$.

For $h_-(x)$, note that $h_-(x) = -h_+(-x)$. So Hölder continuity and monotonicity follows. $\square$

### A.5 Proof of Lemma 4.2

*Proof.* First of all, monotonicity easily follows from the monotonicity of each $h_1$ and $h_2$. For the Hölder continuity, fix some $x, z \in \mathbb{R}^k$, and suppose $h_1(x) \geq h_2(x)$ without loss of generality. Then, we have

$$|\max\{h_1(x), h_2(x)\} - \max\{h_1(z), h_2(z)\}| = \begin{cases} |h_1(x) - h_1(z)| & \text{if } h_1(z) \geq h_2(z) \\ |h_1(x) - h_2(z)| & \text{if } h_1(z) < h_2(z). \end{cases}$$

For the former case, $|h_1(x) - h_1(z)| \leq C\|x - z\|^\gamma$. For the latter case, note that if $h_1(x) \geq h_2(x)$

$$|h_1(x) - h_2(z)| < |h_1(x) - h_1(z)| \leq C\|x - z\|^\gamma.$$

Moreover, if $h_1(x) < h_2(x)$, we have

$$|h_1(x) - h_2(z)| < |h_2(x) - h_2(z)| \leq C\|x - z\|^\gamma,$$

which proves our claim. $\square$

### A.6 Proof of Lemma 4.3

*Proof.* We only prove the claim about the maximum, since the proof for the minimum is analogous. First, note that due to Lemma 4.1, $f^*(x) = f_0 + C\|(x)_{\mathcal{V}+}\|^\gamma$ is in $\Lambda^{f_0}_{+,\mathcal{V}}(\gamma, C)$. Now, for some $x \in \mathbb{R}^k$, suppose there exists some $f^\dagger \in \Lambda^{c}_{+,\mathcal{V}}(\gamma, C)$ such that $f^\dagger(x) > f^*(x)$. Then, we have

$$f^\dagger(x) - f^\dagger(0) > f^*(x) - f^*(0) = C\|(x)_{\mathcal{V}+}\|^\gamma.$$

Define $z = (z_1, ..., z_k)$ such that

$$z_i = \begin{cases} \max\{0, x_i\} & \text{if } i \in \mathcal{V} \\ x_i & \text{otherwise.} \end{cases}$$

Then, we have $z_i \geq x_i$ for all $i \in \mathcal{V}$, so we must have $f^\dagger(z) \geq f^\dagger(x)$. Similarly, we also have $f^\dagger(z) \geq f^\dagger(0) = f_0$. Moreover, by definition of $z$, we have

$$\|(x)_{\mathcal{V}+}\| = \|(z)_{\mathcal{V}+}\| = \|z\|.$$

Then, we can see that

$$\begin{aligned} \left|f^\dagger(z) - f^\dagger(0)\right| =& f^\dagger(z) - f_0 \\ \geq& f^\dagger(x) - f_0 \\ >& C \left\|(x)_{\mathcal{V}+}\right\|^\gamma \\ =& C\|z\|^\gamma, \end{aligned}$$

which violates Hölder continuity. Therefore, $f^*(x)$ attains the maximum. □

## A.7 Proof of Corollary 4.2

*Proof.* We first note that the function classes $\Lambda_{+,\mathcal{V}}(\gamma_j, C_j)$'s are *translation invariant* as defined in Armstrong and Kolesár (2016).

**Definition 1.** For some linear functional $L$ on $\mathcal{F}$, the function class $\mathcal{F}$ is translation invariant if there exists a function $\iota \in \mathcal{F}$ such that $L\iota = 1$ and $f + c\iota \in \mathcal{F}$ for all $c \in \mathbb{R}$ and $f \in \mathcal{F}$.

In our case, by taking $\iota = 1$, we can easily see that the function class $\mathcal{F}_j = \Lambda_{+,\mathcal{V}}(\gamma_j, C_j)$ satisfies translation invariance for our linear function $Lf = f(0)$ for all $j = 1, ..., J$. Let $f^*_{j,\delta} \in \mathcal{F}_j$ and $f^*_{J,\delta} \in \mathcal{F}_J$ solve the the modulus of continuity problem with respect to $\omega\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right)$. Then, by Lemma B.3 in Armstrong and Kolesár (2016), we have

$$\omega'\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right) = \frac{\delta}{\sum_{i=1}^n \left(f^*_{j,\delta}(x_i) - f^*_{J,\delta}(x_i)\right)}. \tag{12}$$

Therefore, we can rewrite $\hat{L}^{\ell,j}_\delta$ in (2) as

$$\hat{L}^{\ell,j}_\delta = \frac{f^*_{j,\delta}(0) + f^*_{J,\delta}(0)}{2} + \frac{\sum_{i=1}^n \left(f^*_{j,\delta}(x_i) - f^*_{J,\delta}(x_i)\right)\left(y_i - \frac{f^*_{j,\delta}(x_i) + f^*_{J,\delta}(x_i)}{2}\right)}{\sum_{i=1}^n \left(f^*_{j,\delta}(x_i) - f^*_{J,\delta}(x_i)\right)}.$$

Next, using Corollary 4.1, we have

$$
\begin{aligned}
\hat{L}_{\delta}^{\ell,j} =& \frac{\sum_{i=1}^{n} D_{Jj,\delta}\left(x_i\right) y_i}{\sum_{i=1}^{n} D_{Jj,\delta}\left(x_i\right)} + \frac{\omega\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right)}{2} \\
& - \frac{\sum_{i=1}^{n} D_{Jj,\delta}\left(x_i\right)\left[\omega\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right) - C_j \left\|(x_i)_{\mathcal{V}-}\right\|^{\gamma_j}\right]}{2\sum_{i=1}^{n} D_{Jj,\delta}\left(x_i\right)} \\
& - \frac{\sum_{i=1}^{n} D_{Jj,\delta}\left(x_i\right) \min\left\{\omega\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right) - C_j \left\|(x_i)_{\mathcal{V}-}\right\|^{\gamma_j}, C_J \left\|(x_i)_{\mathcal{V}+}\right\|^{\gamma_J}\right\}}{2\sum_{i=1}^{n} D_{Jj,\delta}\left(x_i\right)}.
\end{aligned}
$$

Noting that

$$
\begin{aligned}
& D_{Jj,\delta}(x_i) \min\{\omega(\delta, \mathcal{F}_J, \mathcal{F}_j) - C_j\|(x_i)_{\mathcal{V}-}\|^{\gamma_j}, C_J\|(x_i)_{\mathcal{V}+}\|^{\gamma_J}\} \\
=& D_{Jj,\delta}(x_i) \min\{\omega(\delta, \mathcal{F}_J, \mathcal{F}_j) - C_j\|(x_i)_{\mathcal{V}-}\|^{\gamma_j} - C_J\|(x_i)_{\mathcal{V}+}\|^{\gamma_J}, 0\} + C_J\|(x_i)_{\mathcal{V}+}\|^{\gamma_J} \\
=& C_J\|(x_i)_{\mathcal{V}+}\|^{\gamma_J},
\end{aligned}
$$

by the definition of $D_{Jj,\delta}\left(x_i\right)$, we can rewrite the formula for $\hat{L}_{\delta}^{\ell,j}$ as in the statement of the corollary. To get the lower end of the CI, we subtract from $\hat{L}_{\delta}^{\ell,j}$

$$
\begin{aligned}
& \frac{1}{2}(\omega(\delta, \mathcal{F}_J, \mathcal{F}_j) + \delta\omega'(\delta, \mathcal{F}_J, \mathcal{F}_j)) \\
=& \frac{1}{2}\left(\omega(\delta, \mathcal{F}_J, \mathcal{F}_j) + \frac{\delta^2}{\sum_{i=1}^{n} D_{Jj,\delta}(x_i)}\right),
\end{aligned}
$$

where the equality is from the equation (12). The upper end of the CI can be derived in an analogous way, this time using Proposition 4.1. □

## A.8 Proof of Lemma 4.4

*Proof.* We first show that the limiting distribution is invariant with respect to $\tau$. For notational simplicity, we write $\delta = z_{1-\tau}$ and show invariance with respect to $\delta$. With some abuse of notation, we write $(V(\delta)', W(\delta)')'$ as this reparametrized version whose value is the same with $(V(\tau)', W(\tau)')'$ if $\delta = z_{1-\tau}$. Because $(V(\delta)', W(\delta)')'$ is centered and has unit variance, it suffices to show that the covariance terms converge to a limit that does not depend on $\delta$. We show that this is the case for the covariance terms of $V(\delta) := (V_1(\delta), \ldots, V_J(\delta))'$. The same invariance for other covariance terms (covariance between elements of $W(\delta)$ and the covariance between an element of $V(\delta)$ and of $W(\delta)$) follows by an analogous calculation.

Again, we consider the case where $\sigma(\cdot) = 1$. However, this can be relaxed (with more notation) under mild regularity conditions given in Appendix C. Define $b_{Jj,\delta} := \omega\left(\delta, \mathcal{F}_J, \mathcal{F}_j\right)$ for $j \leq J$. Note that

$$\begin{aligned}\operatorname{cov}\left(V_j(\delta), V_\ell(\delta)\right) =& \frac{\sum_{i=1}^n D_{Jj,\delta}(x_i) D_{J\ell}(x_i)}{\delta^2} \\ =& \frac{\sum_{i=1}^n (D_{Jj,\delta}(x_i)/b_{Jj,\delta})(D_{J\ell}(x_i)/b_{Jj,\delta})}{\delta^2/(b_{Jj,\delta} b_{J\ell,\delta})}.\end{aligned} \tag{13}$$

The numerator of the right-hand side is

$$\sum_{i=1}^n \left(1 - \frac{C_j}{b_{Jj,\delta}} \|(x_i)_{\mathcal{V}-}\|^{\gamma_j} - \frac{C_J}{b_{Jj,\delta}} \|(x_i)_{\mathcal{V}+}\|^{\gamma_J}\right)_+ \left(1 - \frac{C_\ell}{b_{J\ell,\delta}} \|(x_i)_{\mathcal{V}-}\|^{\gamma_\ell} - \frac{C_J}{b_{J\ell,\delta}} \|(x_i)_{\mathcal{V}+}\|^{\gamma_J}\right)_+ .$$

We investigate the term

$$\int \left(1 - \frac{C_j}{b_{Jj,\delta}} \|(x)_{\mathcal{V}-}\|^{\gamma_j} - \frac{C_J}{b_{Jj,\delta}} \|(x)_{\mathcal{V}+}\|^{\gamma_J}\right)_+ \left(1 - \frac{C_\ell}{b_{J\ell,\delta}} \|(x)_{\mathcal{V}-}\|^{\gamma_\ell} - \frac{C_J}{b_{J\ell,\delta}} \|(x)_{\mathcal{V}+}\|^{\gamma_J}\right)_+ dx.$$

We consider the case $\gamma_j > \gamma_\ell$, but the case where $\gamma_j = \gamma_\ell$ can be dealt with by taking analogous steps. By a similar argument made in the proof of Theorem 4.1, showing that this integral term and $b_{J\ell,\delta}^{1+k_+/\gamma_\ell+(k-k_+)/\gamma_j} b_{Jj,\delta} n$ are both $o(1)$ will establish $\operatorname{cov}\left(V_j(\delta), V_\ell(\delta)\right) \to 0$. By Theorem 4.1, we have $b_{Jj,\delta} << b_{J\ell,\delta}$, with both going to 0 as $n \to \infty$. By applying a change of variable

$$(x_{[1,m]}/b_{J\ell,\delta}^{1/\gamma_\ell}, x_{[m+1,k]}/b_{J\ell,\delta}^{1/\gamma_J}) = z,$$

we have for a given orthant $O \in \mathcal{O}$

$$\begin{aligned}&\int_O \left(1 - \frac{C_j}{b_{Jj,\delta}} \|(x)_{\mathcal{V}-}\|^{\gamma_j} - \frac{C_J}{b_{Jj,\delta}} \|(x)_{\mathcal{V}+}\|^{\gamma_J}\right)_+ \left(1 - \frac{C_\ell}{b_{J\ell,\delta}} \|(x)_{\mathcal{V}-}\|^{\gamma_\ell} - \frac{C_J}{b_{J\ell,\delta}} \|(x)_{\mathcal{V}+}\|^{\gamma_J}\right)_+ dx \\ =& b_{J\ell,\delta}^{m/\gamma_\ell+(k-m)/\gamma_J} \int_O I_{j\ell J,\delta}(z)\, dz.\end{aligned}$$

Here, $I_{j\ell J,\delta}(z)$ is defined as

$$
\begin{aligned}
& I_{j\ell J,\delta}(z) \\
&= \Big(1 - C_j \|(z_{[1,m]} b_{J\ell,\delta}^{1/\gamma_\ell} b_{Jj,\delta}^{-1/\gamma_j}, 0, z_{[k_+ +1,k]} b_{J\ell,\delta}^{1/\gamma_J} b_{Jj,\delta}^{-1/\gamma_j})\|^{\gamma_j} - C_J \|(0, -z_{[m+1,k]} b_{J\ell,\delta}^{1/\gamma_J} b_{Jj,\delta}^{-1/\gamma_J})\|^{\gamma_J}\Big)_+ \\
&\quad \cdot \Big(1 - C_\ell \|(z_{[1,m]}, 0, z_{[k_+ +1,k]} b_{J\ell,\delta}^{1/\gamma_J - 1/\gamma_\ell})\|^{\gamma_\ell} - C_J \|(0, -z_{[m+1,k]})\|^{\gamma_J}\Big)_+
\end{aligned}
$$

The limit behavior of this term depends on the limit of the following four quantities:

$$1)\ b_{J\ell,\delta}^{1/\gamma_\ell} b_{Jj,\delta}^{-1/\gamma_j},\ 2)\ b_{J\ell,\delta}^{1/\gamma_J} b_{Jj,\delta}^{-1/\gamma_j},\ 3)\ b_{J\ell,\delta}^{1/\gamma_J} b_{Jj,\delta}^{-1/\gamma_J}, \text{ and } 4)\ b_{J\ell,\delta}^{1/\gamma_J - 1/\gamma_\ell}.$$

Since $\gamma > \gamma_\ell \geq \gamma$ and $b_{J\ell,\delta} \asymp n^{-1/2 + k_+/\gamma_\ell + (k-k_+)/\gamma}$, we have $b_{Jj,\delta} << b_{J\ell,\delta}$ and $b_{J\ell,\delta}^{1/\gamma} << b_{J\ell,\delta}^{1/\gamma_\ell} << b_{Jj,\delta}^{1/\gamma}$. This gives

$$1)\ b_{J\ell,\delta}^{1/\gamma_\ell} b_{Jj,\delta}^{-1/\gamma_j} \to 0,\ 2)\ b_{J\ell,\delta}^{1/\gamma_J} b_{Jj,\delta}^{-1/\gamma_j} \to 0,\ 3)\ b_{J\ell,\delta}^{1/\gamma_J} b_{Jj,\delta}^{-1/\gamma_J} \to \infty, \text{ and } 4)\ b_{J\ell,\delta}^{1/\gamma_J - 1/\gamma_\ell} \to 0.$$

Hence, we have

$$\int_O I_{j\ell J,\delta}(z)\, dz = o(1),$$

by a dominated convergence argument, and the convergence rate is the slowest on the orthant where $m = k_+$.

Now, it remains to show that

$$b_{J\ell,\delta}^{k_+/\gamma_\ell + (k-k_+)/\gamma_J} b_{J\ell,\delta} b_{Jj,\delta} n = o(1).$$

Note that the order of the expression on the left-hand side is $n^r$ where $r$ is

$$\frac{1}{2 + k_+/\gamma_\ell + (k-k_+)/\gamma_J} - \frac{1}{2 + k_+/\gamma_j + (k-k_+)/\gamma_J} < 0.$$

This establishes that

$$\operatorname{cov}\left(V_j(\delta), V_\ell(\delta)\right) \to 0$$

for any $j \neq \ell$ and for any $\delta > 0$.

Now, to establish the second half of the lemma, consider the case when $\gamma_j = \gamma$ for all $j$. In such case, we have

$$\int_O \left(1 - \frac{C_j}{b_{Jj,\delta}}\|(x)_{\mathcal{V}-}\|^\gamma - \frac{C_J}{b_{Jj,\delta}}\|(x)_{\mathcal{V}+}\|^\gamma\right)_+ \left(1 - \frac{C_\ell}{b_{J\ell,\delta}}\|(x)_{\mathcal{V}-}\|^\gamma - \frac{C_J}{b_{J\ell,\delta}}\|(x)_{\mathcal{V}+}\|^\gamma\right)_+ dx$$

$$=b_{J\ell,\delta}^{\frac{m}{\gamma}+\frac{k-m}{\gamma}} \int_O \Bigg( \left(1 - C_\ell\|(z_{[1,m]}, 0, z_{[k_++1,k]} b_{J\ell,\delta}^{\frac{1}{\gamma}-\frac{1}{\gamma}})\|^\gamma - C_J\|(0, -z_{[m+1,k]})\|^\gamma\right)_+ \cdot$$

$$\left(1 - C_j\|(z_{[1,m]} b_{J\ell,\delta}^{\frac{1}{\gamma}} b_{Jj,\delta}^{-\frac{1}{\gamma}}, 0, z_{[k_++1,k]} b_{J\ell,\delta}^{\frac{1}{\gamma}} b_{Jj,\delta}^{-\frac{1}{\gamma}})\|^\gamma - C_J\|(0, -z_{[m+1,k]} b_{J\ell,\delta}^{\frac{1}{\gamma}} b_{Jj,\delta}^{-\frac{1}{\gamma}})\|^\gamma\right)_+ \Bigg) dx$$

$$=b_{J\ell,\delta}^{\frac{k}{\gamma}} \int_O \Bigg( \left(1 - C_\ell\|(z_{[1,m]}, 0, z_{[k_++1,k]})\|^\gamma - C_J\|(0, -z_{[m+1,k]})\|^\gamma\right)_+ \cdot$$

$$\left(1 - C_j\|(z_{[1,m]} b_{J\ell,\delta}^{\frac{1}{\gamma}} b_{Jj,\delta}^{-\frac{1}{\gamma}}, 0, z_{[k_++1,k]} b_{J\ell,\delta}^{\frac{1}{\gamma}} b_{Jj,\delta}^{-\frac{1}{\gamma}})\|^\gamma - C_J\|(0, -z_{[m+1,k]} b_{J\ell,\delta}^{\frac{1}{\gamma}} b_{Jj,\delta}^{-\frac{1}{\gamma}})\|^\gamma\right)_+ \Bigg) dx$$

We know that $b_{J\ell,\delta} \asymp n^{-\frac{1}{2+k/\gamma}} (\delta^2/c^*_{J\ell})^{\frac{1}{2+k/\gamma}}$ for some constant $c^*_{J\ell}$, by Theorem 4.1. It follows that

$$(b_{J\ell,\delta}/b_{Jj,\delta})^{1/\gamma} \asymp (c^*_{Jj}/c^*_{J\ell})^{\frac{1/\gamma}{2+k/\gamma}},$$

which then implies

$$\int_O \left(1 - C_1\|(z_{[1,m]} b_{J\ell,\delta}^{\frac{1}{\gamma_\ell}} b_{Jj,\delta}^{-\frac{1}{\gamma}}, 0, z_{[k_++1,k]} b_{J\ell,\delta}^{\frac{1}{\gamma}} b_{Jj,\delta}^{-\frac{1}{\gamma}})\|^{\gamma_1} - C_2\|(0, -z_{[m+1,k]} b_{J\ell,\delta}^{\frac{1}{\gamma}} b_{Jj,\delta}^{-\frac{1}{\gamma}})\|^{\gamma_2}\right)_+ \cdot$$

$$\left(1 - C_\ell\|(z_{[1,m]}, 0, z_{[k_++1,k]})\|^{\gamma_1} - C_J\|(0, -z_{[m+1,k]})\|^{\gamma_2}\right)_+ dx$$

$$= \int_O \Bigg( \left(1 - C_\ell\|(z_{[1,m]}, 0, z_{[k_++1,k]})\|^\gamma - C_J\|(0, -z_{[m+1,k]})\|^\gamma\right)_+ \cdot$$

$$\left(1 - C_j\left(\frac{c^*_{Jj}}{c^*_{J\ell}}\right)^{\frac{1}{2+k/\gamma}}\|(z_{[1,m]}, 0, z_{[k_++1,k]})\|^\gamma - C_J\left(\frac{c^*_{Jj}}{c^*_{J\ell}}\right)^{\frac{1}{2+k/\gamma}}\|(0, -z_{[m+1,k]})\|^\gamma\right)_+ \Bigg) dx + o(1).$$

Denote the integral term following the last equality as $B_{j\ell,O}$, and $B_{j\ell} = \sum_{O\in\mathcal{O}} B_{j\ell,O}$. Plugging this result back into (13), we have that

$$\text{cov}\left(V_j(\delta), V_\ell(\delta)\right) = b_{J\ell,\delta}^{k/\gamma} b_{Jj,\delta} b_{J\ell,\delta} n B_{j\ell}(1 + o(1))/\delta^2.$$

Now, note that

$$\begin{aligned} b_{J\ell,\delta}^{k/\gamma} b_{Jj,\delta} b_{J\ell,\delta} n B_{j\ell}/\delta^2 =& \left(\delta^2/c^*_{J\ell}\right)^{\frac{k/\gamma}{2+k/\gamma}} \left(\delta^2/c^*_{J\ell}\right)^{\frac{1}{2+k/\gamma}} \left(\delta^2/c^*_{Jj}\right)^{\frac{1}{2+k/\gamma}} B_{j\ell}/\delta^2 + o(1) \\ =& {c^*_{J\ell}}^{-\frac{k/\gamma}{2+k/\gamma}} {c^*_{J\ell}}^{-\frac{1}{2+k/\gamma}} {c^*_{Jj}}^{-\frac{1}{2+k/\gamma}} B_{j\ell} + o(1). \end{aligned}$$

While this calculation is sufficient to show the invariance of the limiting covariance with respect to $\delta$, we further simplify the term by some additional calculations.

By changing the role of $j$ and $\ell$ in the above change of variables, we know that

$$\begin{aligned} & b_{J\ell,\delta}^{\frac{k}{\gamma}} \int_O \left(1 - C_\ell \|(z_{[1,m]}, 0, z_{[k_+ +1,k]})\|^\gamma - C_J \|(0, -z_{[m+1,k]})\|^\gamma\right)_+ \cdot \\ & \qquad \left(1 - C_j (c^*_{Jj}/c^*_{J\ell})^{\frac{1}{2+k/\gamma}} \|(z_{[1,m]}, 0, z_{[k_+ +1,k]})\|^\gamma - C_J (c^*_{Jj}/c^*_{J\ell})^{\frac{1}{2+k/\gamma}} \|(0, -z_{[m+1,k]})\|^\gamma\right)_+ dx \\ =& b_{Jj,\delta}^{\frac{k}{\gamma}} \int_O \left(1 - C_j \|(z_{[1,m]}, 0, z_{[k_+ +1,k]})\|^\gamma - C_J \|(0, -z_{[m+1,k]})\|^\gamma\right)_+ \cdot \\ & \qquad \left(1 - C_\ell (c^*_{J\ell}/c^*_{Jj})^{\frac{1}{2+k/\gamma}} \|(z_{[1,m]}, 0, z_{[k_+ +1,k]})\|^\gamma - C_J (c^*_{J\ell}/c^*_{Jj})^{\frac{1}{2+k/\gamma}} \|(0, -z_{[m+1,k]})\|^\gamma\right)_+ dx. \end{aligned}$$

Now, consider the change of variables given by

$$(x_{[1,m]}/(b_{J\ell,\delta}^{1/(2\gamma)} b_{Jj,\delta}^{1/(2\gamma)}), x_{[m+1,k]}/(b_{J\ell,\delta}^{1/(2\gamma)} b_{Jj,\delta}^{1/(2\gamma)})) = z.$$

We have

$$\begin{aligned} & \int_O \left(1 - \frac{C_j}{b_{Jj,\delta}} \|(x)_{\mathcal{V}-}\|^\gamma - \frac{C_J}{b_{Jj,\delta}} \|(x)_{\mathcal{V}+}\|^\gamma\right)_+ \left(1 - \frac{C_\ell}{b_{J\ell,\delta}} \|(x)_{\mathcal{V}-}\|^\gamma - \frac{C_J}{b_{J\ell,\delta}} \|(x)_{\mathcal{V}+}\|^\gamma\right)_+ dx \\ =& b_{Jj,\delta}^{\frac{k}{2\gamma}} b_{J\ell,\delta}^{\frac{k}{2\gamma}} \cdot \\ & \int_O \left(1 - C_j (c^*_{Jj}/c^*_{J\ell})^{\frac{1/(2\gamma)}{2+k/\gamma}} \|(z_{[1,m]}, 0, z_{[k_+ +1,k]})\|^\gamma - C_J (c^*_{Jj}/c^*_{J\ell})^{\frac{1/(2\gamma)}{2+k/\gamma}} \|(0, -z_{[m+1,k]})\|^\gamma\right)_+ \cdot \\ & \qquad \left(1 - C_\ell (c^*_{J\ell}/c^*_{Jj})^{\frac{1/(2\gamma)}{2+k/\gamma}} \|(z_{[1,m]}, 0, z_{[k_+ +1,k]})\|^\gamma - C_J (c^*_{J\ell}/c^*_{Jj})^{\frac{1/(2\gamma)}{2+k/\gamma}} \|(0, -z_{[m+1,k]})\|^\gamma\right)_+ dx \\ & + o\left(b_{Jj,\delta}^{\frac{k}{2\gamma}} b_{J\ell,\delta}^{\frac{k}{2\gamma}}\right) \end{aligned}$$

Here, we used

$$(b_{J\ell,\delta}/b_{Jj,\delta})^{1/(2\gamma)} = (c^*_{Jj}/c^*_{J\ell})^{\frac{1/(2\gamma)}{2+k/\gamma}} + o(1)$$

Now, write the integral in the last term as $B^*_{j\ell,O}$, and $B^*_{j\ell} = \sum_{O \in \mathcal{O}} B^*_{j\ell,O}$ Finally, by similar calculations as above

$$
\begin{aligned}
\operatorname{cov}(V_j(\delta), V_\ell(\delta)) &= n b_{Jj,\delta}^{\frac{k}{2\gamma}} b_{J\ell,\delta}^{\frac{k}{2\gamma}} b_{Jj,\delta} b_{J\ell,\delta} B_{j\ell}^* / \delta^2 + o(1) \\
&= {c_{Jj}^*}^{-\frac{k/(2\gamma)}{2+k/\gamma}} {c_{J\ell}^*}^{-\frac{k/(2\gamma)}{2+k/\gamma}} {c_{Jj}^*}^{-\frac{1}{2+k/\gamma}} {c_{J\ell}^*}^{-\frac{1}{2+k/\gamma}} B_{j\ell}^* + o(1) \\
&= {c_{Jj}^*}^{-1/2} {c_{J\ell}^*}^{-1/2} B_{j\ell}^* + o(1).
\end{aligned}
$$

This shows that $\operatorname{cov}(V_j(\delta), V_\ell(\delta)) \to {c_{Jj}^*}^{-1/2} {c_{J\ell}^*}^{-1/2} B_{j\ell}^*$ as $n \to \infty$. Note that the limiting covariance term does not depend on $\delta$.

Note that we have $V_j(\delta) \overset{d}{=} \sum_{i=1}^n D_{Jj,\delta}(x_i) Z_i / \delta$ where $Z_i$'s are i.i.d standard normal random variables. Furthermore, this equivalence holds jointly for $V_j(\delta)$, $j = 1, \ldots, J$. Let $\{x_i\}_{i=1}^\infty$ be a sequence where the under which where Theorem 4.1 holds. Define for $C \in [C_1, C_J]$

$$
Z_{ni}(C) = (\omega(\delta, \mathcal{F}_J, \Lambda_{+,\mathcal{V}}(\gamma, C)) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma)_+ Z_i/\delta,
$$

and consider the stochastic process $\sum_{i=1}^n Z_{ni}(C)$ indexed by $C \in [C_1, C_J]$. We show that this process weakly converges to a tight Gaussian process, from which the fact that the quantile of the maximum of $V(\delta)$ does not depend on $J$ follows.

We use Theorem 2.11.1 of van der Vaart and Wellner (1996) to establish this convergence. Specifically, we use the result given by Example 2.11.13. Given the results we already have, it suffices to show that

$$
\sum_{i=1}^n \left| \frac{\partial}{\partial C} (\omega(\delta, \mathcal{F}_J, \Lambda_{+,\mathcal{V}}(\gamma, C)) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|)_+ \right|^2 = O(1),
$$

and that a Lindeberg condition is satisfied.

With some abuse of notation, we write

$$
D_{C,n,\delta}(x_i) := (\omega(\delta, \mathcal{F}_J, \Lambda_{+,\mathcal{V}}(\gamma, C)) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|)_+,
$$

and $\omega(\delta, C_J, C) = \omega(\delta, \mathcal{F}_J, \Lambda_{+,\mathcal{V}}(\gamma, C))$ with $\omega^{-1}(b, C_J, C)$ defined similarly. Recall that

$$
\omega^{-1}(b, C_J, C) = \left( \sum_{i=1}^n (b - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma)_+^2 \right)^{\frac{1}{2}}.
$$

From the identity $\delta = \omega^{-1}(\omega(\delta, C_J, C), C_J, C)$, we have

$$0 = \frac{\partial}{\partial b}\omega^{-1}(\omega(\delta, C_J, C), C_J, C)\frac{\partial}{\partial C}\omega(\delta, C_J, C) + \frac{\partial}{\partial C}\omega^{-1}(\omega(\delta, C_J, C), C_J, C)$$

so that

$$\begin{aligned}\frac{\partial}{\partial C}\omega(\delta, C_J, C) &= -\frac{\frac{\partial}{\partial C}\omega^{-1}(\omega(\delta, C_J, C), C_J, C)}{\frac{\partial}{\partial b}\omega^{-1}(\omega(\delta, C_J, C), C_J, C)} \\ &= \frac{\sum_{i=1}^n \|(x_i)_{\mathcal{V}-}\|^\gamma \left[\omega(\delta, C_J, C) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma\right]_+}{\sum_{i=1}^n \left[\omega(\delta, C_J, C) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma\right]_+}.\end{aligned}$$

This gives

$$\begin{aligned}\frac{\partial}{\partial C}D_{C,n,\delta}(x_i) =& \left(\frac{\sum_{i=1}^n \|(x_i)_{\mathcal{V}-}\|^\gamma \left[\omega(\delta, C_J, C) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma\right]_+}{\sum_{i=1}^n \left[\omega(\delta, C_J, C) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma\right]_+} - \|(x_i)_{\mathcal{V}-}\|^\gamma\right) \\ &\cdot \mathbb{1}\left(\omega(\delta, C_J, C) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma \geq 0\right) \\ =& \left(\frac{\sum_{k=1}^n (\|(x_k)_{\mathcal{V}-}\|^\gamma - \|(x_i)_{\mathcal{V}-}\|^\gamma) \left[\omega(\delta, C_J, C) - C\|(x_k)_{\mathcal{V}-}\|^\gamma - C_J\|(x_k)_{\mathcal{V}+}\|^\gamma\right]_+}{\sum_{k=1}^n \left[\omega(\delta, C_J, C) - C\|(x_k)_{\mathcal{V}-}\|^\gamma - C_J\|(x_k)_{\mathcal{V}+}\|^\gamma\right]_+}\right) \\ &\cdot \mathbb{1}\left(\omega(\delta, C_J, C) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma \geq 0\right),\end{aligned}$$

with the understanding that the fraction equals 0 if the denominator is 0. We have

$$\begin{aligned}&\left|\frac{\partial}{\partial C}D_{C,n,\delta}(x_i)(x_i)\right|^2 \\ \leq& \left(\frac{\omega(\delta, C_J, C)\sum_{k=1}^n \left[\omega(\delta, C_J, C) - C\|(x_k)_{\mathcal{V}-}\|^\gamma - C_J\|(x_k)_{\mathcal{V}+}\|^\gamma\right]_+}{C\sum_{k=1}^n \left[\omega(\delta, C_J, C) - C\|(x_k)_{\mathcal{V}-}\|^\gamma - C_J\|(x_k)_{\mathcal{V}+}\|^\gamma\right]_+}\right)^2 \\ &\cdot \mathbb{1}\left(\omega(\delta, C_J, C) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma \geq 0\right) \\ \leq& (\omega(\delta, C_J, C)/C)^2 \mathbb{1}\left(\omega(\delta, C_J, C) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma \geq 0\right),\end{aligned}$$

so that

$$\begin{aligned}&\sum_{i=1}^n \left|\frac{\partial}{\partial C}D_{C,n,\delta}(x_i)(x_i)\right|^2 \\ \leq& (\omega(\delta, C_J, C)/C)^2 \sum_{i=1}^n \mathbb{1}\left(\omega(\delta, C_J, C) - C\|(x_i)_{\mathcal{V}-}\|^\gamma - C_J\|(x_i)_{\mathcal{V}+}\|^\gamma \geq 0\right) \\ \asymp& n^{-\frac{2}{2+k/\gamma}} \cdot n^{1-\frac{k/\gamma}{2+\gamma/k}} = O(1).\end{aligned}$$

To check the Lindeberg condition, note that

$$\|Z_{ni}\| := \sup_{C\in[C_1,C_J]} |Z_{ni}(C)| \leq \frac{\omega(\delta, C_J, C_J)}{\delta}|Z_i|$$

so that

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbf{E}\|Z_{ni}\|\mathbb{1}(\|Z_{ni}\| > \eta)\\
\leq&\frac{\omega_n(\delta, C_J, C_J)}{\delta}\sum_{i=1}^{n}\mathbf{E}|Z_i|\mathbb{1}(\frac{\omega_n(\delta, C_J, C_J)}{\delta}|Z_i| > \eta)\\
\leq&\frac{n\omega_n(\delta, C_J, C_J)}{\delta}\mathbf{E}|Z_i|\mathbb{1}(\frac{\omega_n(\delta, C_J, C_J)}{\delta}|Z_i| > \eta)\\
=&2\frac{n\omega_n(\delta, C_J, C_J)}{\delta}\mathbf{E}Z_i\mathbb{1}(\frac{\omega_n(\delta, C_J, C_J)}{\delta}Z_i > \eta)\\
=&2\frac{n\omega_n(\delta, C_J, C_J)}{\delta}\phi(\frac{\eta\delta}{\omega_n(\delta, C_J, C_J)}) \to 0.
\end{aligned}$$

We have already shown that the covariance function converges pointwise. Hence, we conclude that $\sum_{i=1}^{n} Z_{ni}$ converges in distribution in $\ell^{\infty}([C_1, C_J])$ to a tight Gaussian process. Moreoever, this limiting distribution does not depend on $\delta$. $\square$

# Appendix B Proof of Theorem 4.1

*Proof.* For simplicity, write $b_n = bn^{-r(\gamma_1,\gamma_2)}$, where $b > 0$ is arbitrary, and define

$$W_{i,n}(\gamma_1, C_1, \gamma_2, C_2) := \left(1 - \frac{C_1}{b_n}\left\|(X_i)_{\mathcal{V}+}\right\|^{\gamma_1} - \frac{C_2}{b_n}\left\|(X_i)_{\mathcal{V}-}\right\|^{\gamma_2}\right)_+^2 .$$

Note that $b_n \to 0$ and $n^{1-\eta}b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2} \to \infty$ for some $\eta > 0$. First, we show that, for constants $c^*_{2,1}$ and $c^*_{2,1}$ that do not depend on $b$,

$$\begin{aligned}
\text{(a)}\quad & \lim_{n\to\infty} \frac{1}{nb_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}} \sum_{i=1}^{n} W_{i,n}(\gamma_1, C_1, \gamma_2, C_2) = c^*_{1,2} > 0, \\
& \lim_{n\to\infty} \frac{1}{nb_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}} \sum_{i=1}^{n} W_{i,n}(\gamma_2, C_2, \gamma_1, C_1) = c^*_{2,1} > 0 \text{ , and} \\
\text{(b)}\quad & \lim_{n\to\infty} b_n^{-1} \min_{i\le n} \left\{C_1\left\|(X_i)_{\mathcal{V}+}\right\|^{\gamma_1} + C_2\left\|(X_i)_{\mathcal{V}c-}\right\|^{\gamma_2}\right\} = 0 \\
& \lim_{n\to\infty} b_n^{-1} \min_{i\le n} \left\{C_2\left\|(X_i)_{\mathcal{V}+}\right\|^{\gamma_2} + C_2\left\|(X_i)_{\mathcal{V}-}\right\|^{\gamma_1}\right\} = 0,
\end{aligned}$$

where all equalities hold in an almost sure sense.

To show (a), take an arbitrary $\varepsilon > 0$. Due to the regularity conditions on $p_X(\cdot)$ and $\sigma(\cdot)$, there exists a neighborhood $\mathcal{N}_\varepsilon$ of 0 such that $|p_X(x) - p_X(0)| \le \varepsilon$ for all $x \in \mathcal{N}_\varepsilon$. Writing $B_n := \left\{x \in \mathbb{R}^k : b_n - C_1\left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1} - C_2\left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2} > 0\right\}$, there exists $N_\varepsilon$ such that for all $n \ge N_\varepsilon$ we have $B_n \subset \mathcal{N}_\varepsilon \cap \mathcal{X}$ because $b_n \to 0$ and the interior of $\mathcal{X}$ contains 0. Hence, for $n \ge N_\varepsilon$, we have

$$\begin{aligned}
& (p_X(0) - \varepsilon) \int_{B_n} \left(1 - \frac{C_1}{b_n}\left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1} - \frac{C_2}{b_n}\left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2}\right)^2 dx \\
\le\ & \mathbf{E}\, W_{i,n} \\
\le\ & (p_X(0) + \varepsilon) \int_{B_n} \left(1 - \frac{C_1}{b_n}\left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1} - \frac{C_2}{b_n}\left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2}\right)^2 dx.
\end{aligned} \tag{14}$$

Let $\mathcal{O}$ be the collection of the $2^k$ orthants on $\mathbb{R}^k$. Then, we can write

$$\begin{aligned}&\int_{B_n}\left(1-\frac{C_1}{b_n}\left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1}-\frac{C_2}{b_n}\left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2}\right)^2 dx\\ =&\sum_{O\in\mathcal{O}}\int_{B_n\cap O}\left(1-\frac{C_1}{b_n}\left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1}-\frac{C_2}{b_n}\left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2}\right)^2 dx.\end{aligned} \tag{15}$$

Now, consider an orthant $O$ and let $O_+\subset\{1,\ldots,k\}$ be the index set for those elements that take positive values on $O$. Without loss of generality, suppose $O_+\cap\mathcal{V}=\{1,\ldots,m\}$[9] for $m=0,\ldots,k$, where we take $O_+\cap\mathcal{V}=\emptyset$ if $m=0$. For $k_1\le k_2$, define the subvector $z_{[k_1,k_2]}=(z_{k_1},z_{k_1+1},\ldots,z_{k_2})$ for any $z:=(z_1,\ldots,z_k)\in\mathbb{R}^k$. It follows that

$$\begin{aligned}&\int_{B_n\cap O}\left(1-\frac{C_1}{b_n}\left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1}-\frac{C_2}{b_n}\left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2}\right)^2 dx\\ =&\int_{B_n\cap O}\left(1-\frac{C_1}{b_n}\left\|(x_{[1,m]},0,x_{[k_++1,k]})\right\|^{\gamma_1}-\frac{C_2}{b_n}\left\|(0,-x_{[m+1,k_+]},-x_{[k_++1,k]})\right\|^{\gamma_2}\right)^2 dx.\end{aligned}$$

By applying a changes of variables with $(x_{[1,m]}/b_n^{1/\gamma_1},x_{[m+1,k]}/b^{1/\gamma_2})=z$, the last equation becomes

$$\begin{aligned}&\int_{B_n\cap O}\left(1-\frac{C_1}{b_n}\left\|(x_{[1,m]},0,x_{[k_++1,k]})\right\|^{\gamma_1}-\frac{C_2}{b_n}\left\|(0,-x_{[m+1,k]})\right\|^{\gamma_2}\right)^2 dx\\ =&b_n^{\frac{m}{\gamma_1}+\frac{k-m}{\gamma_2}}\int_O\left(1-C_1\left\|\left(z_{[1,m]},0,z_{[k_++1,k]}b_n^{\frac{1}{\gamma_2}-\frac{1}{\gamma_1}}\right)\right\|^{\gamma_1}-C_2\left\|(0,-z_{[m+1,k]})\right\|^{\gamma_2}\right)_+^2 dx.\end{aligned}$$

Note that by Lebesgue's dominated convergence theorem the integral in the last expression can be written as $c_O(C_1,C_2)+o(1)$ where

$$\begin{aligned}&c_O(C_1,C_2)\\ :=&\begin{cases}\int_O\left(1-C_1\left\|(z_{[1,m]},0)\right\|^{\gamma_1}-C_2\left\|(0,-z_{[m+1,k]})\right\|^{\gamma_2}\right)_+^2 dz & \text{if } \gamma_1>\gamma_2\\ \int_O\left(1-C_1\left\|(z_{[1,m]},0,z_{[k_++1,k]})\right\|^{\gamma_1}-C_2\left\|(0,-z_{[m+1,k]})\right\|^{\gamma_2}\right)_+^2 dz & \text{if } \gamma_1=\gamma_2.\end{cases}\end{aligned}$$

[9]Here, we are implicitly assuming that we modify the definition of the norm in a way that corresponds to the relabeling. More formally, we could write the modified norm as $\|\cdot\|_O$, which we do not do for succinctness. Note that this modification is unnecessary when $\|z\|$ is invariant with respect to permutations of $z$, which is the case for (unweighted) $\ell_p$ norms.

Hence, we have

$$\int_O \left(1 - \frac{C_1}{b_n} \left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1} - \frac{C_2}{b_n} \left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2}\right)_+^2 dx = b_n^{\frac{m}{\gamma_1}+\frac{k-m}{\gamma_2}} \left(c_O(C_1, C_2) + o(1)\right).$$

Moreover, note that $c_O(C_1, C_2) > 0$. If $\gamma_1 > \gamma_2$, the integrals that correspond to the orthants where $m = k_+$ determine the rate at which the entire integral goes to 0. If $\gamma_1 = \gamma_2$ note that the exponent of $b_n$ is always $k/\gamma_1$ and thus the integral is of the same order (in terms of $b_n$) on all the orthants. Let $\mathcal{O}_+$ denote the collection of those orthants with $m = k_+$, and write $c_+(C_1, C_2) = \sum_{O\in\mathcal{O}_+} c_O(C_1, C_2)$ if $\gamma_1 > \gamma_2$ and $c_+(C_1, C_2) = \sum_{O\in\mathcal{O}} c_O(C_1, C_2)$ if $\gamma_1 = \gamma_2$. Then, it follows that

$$\int \left(1 - \frac{C_1}{b_n} \left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1} - \frac{C_2}{b_n} \left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2}\right)_+^2 dx = b_n^{\frac{k_+}{\gamma_1}+\frac{k-k_+}{\gamma_2}} \left(c_+(C_1, C_2) + o(1)\right).$$

Combining this with (14), it follows that

$$\begin{aligned} &\left(c_+(C_1, C_2) + o(1)\right)(p_X(0) - \varepsilon) b_n^{\frac{k_+}{\gamma_1}+\frac{k-k_+}{\gamma_2}} \\ \leq\, &\mathbf{E}\, W_{i,n} \\ \leq\, &\left(c_+(C_1, C_2) + o(1)\right)(p_X(0) + \varepsilon) b_n^{\frac{k_+}{\gamma_1}+\frac{k-k_+}{\gamma_2}} \end{aligned}$$

for large $n$. Dividing all sides by $b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}$, taking $n \to \infty$, and then taking $\varepsilon \to 0$, we have

$$\lim_{n\to\infty} \mathbf{E}\, \frac{W_{i,n}}{b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}} = c_+(C_1, C_2) p_X(0). \tag{16}$$

Now, consider the term $EW_{i,n}^2$. We have

$$\begin{aligned} &(p_X(0) - \varepsilon) \int_{B_n} \left(1 - \frac{C_1}{b_n} \left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1} - \frac{C_2}{b_n} \left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2}\right)^4 dx \\ \leq\ &\mathbf{E}\, W_{i,n}^2 \\ \leq\ &(p_X(0) + \varepsilon) \int_{B_n} \left(1 - \frac{C_1}{b_n} \left\|(x)_{\mathcal{V}+}\right\|^{\gamma_1} - \frac{C_2}{b_n} \left\|(x)_{\mathcal{V}-}\right\|^{\gamma_2}\right)^4 dx. \end{aligned} \tag{17}$$

Hence, repeating the exact same steps that we went through for $\mathbf{E}\, W_{i,n}$, we have

$$\lim_{n\to\infty} \mathbf{E}\, \frac{W_{i,n}^2}{b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}} = c^\dagger p_X(0),$$

for some $c^\dagger > 0$, which shows that $(\mathbf{E}\, W_{i,n}^2)^{1/2} \asymp b_n^{(k_+/\gamma_1+(k-k_+)/\gamma_2)/2}$.

Now, define $\widetilde{W}_n := \frac{1}{n}\sum_{i=1}^n (W_{i,n} - \mathbf{E}\, W_{i,n})$ and $\varepsilon_n = \varepsilon b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}$. By Bernstein's inequality, we have

$$\mathbf{P}(|\widetilde{W}_n| > \varepsilon_n) \leq 2\exp\left(-\frac{1}{2}\frac{n\varepsilon_n^2}{\mathbf{E}\, W_{i,n}^2 + \varepsilon_n/3}\right) \leq 2\exp\left(-\frac{1}{2}\frac{n\varepsilon_n}{K+1/3}\right)$$

where the last inequality holds for large enough $n$ and some constant $K > 0$. It follows that, for large $n$,

$$\exp\left(-\frac{1}{2}\frac{n\varepsilon_n}{K+1/3}\right) = \exp\left(-n^\eta \frac{1}{2}\frac{n^{1-\eta} b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}\varepsilon}{K+1/3}\right) \leq \exp\left(-n^\eta\right),$$

where the inequality follows from the fact that $n^{1-\eta} b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2} \to \infty$. This shows that $\sum_{n=1}^\infty \mathbf{P}(|\widetilde{W}_n| > \varepsilon_n) < \infty$. By the Borel-Cantelli lemma, we have

$$\frac{1}{n b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}} \sum_{i=1}^n (W_{i,n} - \mathbf{E}\, W_{i,n}) \overset{a.s.}{\to} 0. \tag{18}$$

Combining (16) and (18), we have

$$\lim_{n\to\infty} \frac{1}{n b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}} \sum_{i=1}^n W_{i,n} = c_+(C_1, C_2) p_X(0)$$

almost surely, which establishes the desired result with $c_{1,2}^* = c_+(C_1, C_2)p_X(0)$. Note that $c_{1,2}^*$ does not depend on $b$.

The proof for

$$\lim_{n\to\infty} \frac{1}{n b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}} \sum_{i=1}^n W_{i,n}(\gamma_2, C_2, \gamma_1, C_1) = c_{2,1}^* > 0$$

is essentially the same, with some minor modifications. The change of variables we previously used should be modified to

$$(x_{[1,m]}/b_n^{1/\gamma_2}, x_{[m+1,k_+]}/b_n^{1/\gamma_1}, x_{[k_++1,k]}/b_n^{1/\gamma_2}) = z,$$

and, the constant $c_+(C_1, C_2)$ should be changed to $c_-(C_1, C_2) := \sum_{O\in\mathcal{O}_-} c_O(C_2, C_1)$

where $\mathcal{O}_-$ is the collection of orthants with $m = 0$.[10] Hence, here we get the desired result with $c^*_{2,1} = c_-(C_1, C_2)p_X(0)$, which again does not depend on $b$.

Now, we prove (b). We only give the proof for

$$\lim_{n\to\infty} b_n^{-1} \min_{i\le n} \left\{C_1 \left\|(X_i)_{\mathcal{V}+}\right\|^{\gamma_1} + C_2 \left\|(X_i)_{\mathcal{V}-}\right\|^{\gamma_2}\right\} = 0 \ a.s.,$$

since the other half of the statement can be proved analogously. Let $\varepsilon > 0$ be an arbitrary constant, and denote the event

$$A_{n,\varepsilon} := \left\{b_n^{-1} \min_{i\le n} \left\{C_1 \left\|(X_i)_{\mathcal{V}+}\right\|^{\gamma_1} + C_2 \left\|(X_i)_{\mathcal{V}-}\right\|^{\gamma_2}\right\} \ge \varepsilon\right\}.$$

Note that it is enough to show $\Sigma_{n=1}^{\infty} P\left(A_{n,\varepsilon}\right) < \infty$, since then the result follows from the Borel-Cantelli lemma. We have

$$\begin{aligned}
\mathbf{P}\left(A_{n,\varepsilon}\right) &= \mathbf{P}\left(\min_{i\le n} \left\{C_1 \left\|(X_i)_{\mathcal{V}+}\right\|^{\gamma_1} + C_2 \left\|(X_i)_{\mathcal{V}-}\right\|^{\gamma_2}\right\} \ge b_n\varepsilon\right) \\
&= \mathbf{P}(C_1\|(X_i)_{\mathcal{V}+}\|^{\gamma_1} + C_2\|(X_i)_{\mathcal{V}-}\|^{\gamma_2} \ge b_n\varepsilon)^n \\
&= (1 - \mathbf{P}(C_1\|(X_i)_{\mathcal{V}+}\|^{\gamma_1} + C_2\|(X_i)_{\mathcal{V}-}\|^{\gamma_2} < b_n\varepsilon))^n.
\end{aligned}$$

By an analogous calculation as in (a), we can show

$$\begin{aligned}
&\mathbf{P}(C_1 \left\|(X_1)_{\mathcal{V}+}\right\|^{\gamma_1} + C_2 \left\|(X_2)_{\mathcal{V}-}\right\|^{\gamma_2} < b_n\varepsilon) \\
=&b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}\left(c + o(1)\right),
\end{aligned}$$

where $c > 0$ and the $o(1)$ term is also positive. This gives, for large $n$ and from some positive constant $K > 0$,

$$\begin{aligned}
\mathbf{P}(A_{n,\varepsilon}) \le& \left(1 - cb_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}\right)^n \\
\le& \exp\left(-cnb_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}\right) \\
=& \exp\left(-cn^{\eta}n^{1-\eta}b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}\right) \\
\le& \exp\left(-cn^{\eta}K\right)
\end{aligned}$$

This shows that $\sum_{n=1}^{\infty} \mathbf{P}(A_{n,\varepsilon}) \le \sum_{n=1}^{\infty} \exp(-cnb_n^{-(k_+/\gamma_1+(k-k_+)/\gamma_2)}) < \infty$, which establishes (b).

[10] Again, the norms must be redefined to be consistent with the "relabeling".

Now, using (a) and (b), we prove the given rate result. Let $\{x_i\}_{i=1}^{\infty}$ be a realization of $\{X_i\}_{i=1}^{\infty}$ such that (a) and (b) hold, which is the case for almost all realizations. We prove the result for only $\omega\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right)$ because the proof for $\omega\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right), \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right)\right)$ is essentially the same. Throughout the proof, we write $w_{i,n} := w_{i,n}(\gamma_1, C_1, \gamma_2, C_2)$ for simplicity. Define

$$\widetilde{\omega}_n(\delta) := n^{r(\gamma_1,\gamma_2)}\omega\left(\delta, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right),$$

and $\widetilde{\omega}_\infty(\delta) = \left(\delta^2/c^*\right)^{\frac{1}{2+k_+/\gamma_1+(k-k_+)/\gamma_2}}$. We want to show $\widetilde{\omega}_n(\delta) \to \widetilde{\omega}_\infty(\delta)$ for all $\delta > 0$. On the range of $\widetilde{\omega}_n(\cdot)$, define its inverse $\widetilde{\omega}_n^{-1}(b)$ for $b > 0$:

$$\widetilde{\omega}_n^{-1}(b) = \omega^{-1}\left(n^{-r(\gamma_1,\gamma_2)}b, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right),$$

and let $b_n = n^{-r(\gamma_1,\gamma_2)}b$. It follows that

$$\begin{aligned}
\widetilde{\omega}_n^{-1}(b) &= \left(b_n^2 \sum_{i=1}^n w_{i,n}\right)^{1/2} \\
&= \left(n b_n^{2+k_+/\gamma_1+(k-k_+)/\gamma_2} \frac{1}{n b_n^{k_+/\gamma_1+(k-k_+)/\gamma_2}} \sum_{i=1}^n w_{i,n}\right)^{1/2} \\
&\to \left(b^{2+k_+/\gamma_1+(k-k_+)/\gamma_2} c_{1,2}^*\right)^{1/2},
\end{aligned}$$

where the last line follows by (a). Defining $\widetilde{\omega}_\infty^{-1}(b) = \left(b^{2+k_+/\gamma_1+(k-k_+)/\gamma_2} c_{1,2}^*\right)^{1/2}$, which is the precisely the inverse function of $\widetilde{\omega}_\infty(\cdot)$, on an appropriately defined domain. Now, if we can show that any $b > 0$ is in the range of $\widetilde{\omega}_n(\cdot)$ for large enough $n$, we can apply Lemma F.1 of Armstrong and Kolesár (2016) to establish that $\widetilde{\omega}_n(\delta) \to \widetilde{\omega}_\infty(\delta)$ for all $\delta > 0$. To this end, it is enough to show

$$\lim_{n\to\infty} n^{r(\gamma_1,\gamma_2)}\omega\left(0, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right) \to 0.$$

Following the derivation of the solution to the inverse modulus problem, it is easy to check that

$$\omega\left(0, \Lambda_{+,\mathcal{V}}\left(\gamma_1, C_1\right), \Lambda_{+,\mathcal{V}}\left(\gamma_2, C_2\right)\right) = \min_{i\le n}\left\{C_1\left\|(x_i)_{\mathcal{V}+}\right\|^{\gamma_1} + C_2\left\|(x_i)_{\mathcal{V}-}\right\|^{\gamma_2}\right\}.$$

It remains only to show

$$\lim_{n\to\infty} n^{r(\gamma_1,\gamma_2)} \min_{i\le n} \left\{ C_1 \left\|(x_i)_{\mathcal{V}+}\right\|^{\gamma_1} + C_2 \left\|(x_i)_{\mathcal{V}-}\right\|^{\gamma_2} \right\} = 0,$$

which is immediate from (b). □

# Appendix C Heteroskedasticity

In Theorem 4.1, we assume $\sigma(\cdot) = 1$. However, allowing for general heteroskedasticity do not change the result as long as $\sigma(\cdot)$ is continuous at 0 and $\sigma(0) > 0$. All proofs follow with minor changes. The solution to the inverse modulus problem remain unchanged. For Theorem 4.1, we can take $\varepsilon \in (0, \sigma(0))$ and replace the terms $p_X(0) - \varepsilon$ and $p_X(0) + \varepsilon$ by $(p_X(0) - \varepsilon)/(\sigma(0) + \varepsilon)$ and $(p_X(0) + \varepsilon)/(\sigma(0) - \varepsilon)$ in (14). Accordingly, we replace the right-hand side of (16) by $cp_X(0)/\sigma(0)$, and the result of the theorem remains the same with a slightly modified definition of the constant terms.

# Appendix D Adaptation Under Only Monotonicity

Define the $\Lambda_{+,\mathcal{V}}(0,\infty)$ the space of monotone functions with respect to those variables whose indices lie in $\mathcal{V}$. Specifically,

$$\Lambda_{+,\mathcal{V}}(0,\infty) := \left\{ f \in \mathcal{F}(\mathbb{R}^k) : f(x) \ge f(z) \text{ if } x_i \ge z_i \ \forall i \in \mathcal{V} \text{ and } x_i = z_i \ \forall i \notin \mathcal{V} \right\}.$$

Here, we consider the problem of adapting to $\Lambda_{+,\mathcal{V}}(\gamma, C)$ while maintaining coverage over $\Lambda_{+,\mathcal{V}}(0,\infty)$. The corresponding inverse (ordered) modulus problem

$$\inf_{f_1,f_2} \left( \sum_{i=1}^{n} \left(f_2(x_i) - f_1(x_i)\right)^2 \right)^{1/2}$$
$$\text{s.t. } f_2(0) - f_1(0) = b, \ f_1 \in \Lambda_{+,\mathcal{V}}(\gamma, C), f_2 \in \Lambda_{+,\mathcal{V}}(0,\infty).$$

Let $f_1^*(x) = \min\left\{C\left\|(x)_{\mathcal{V}+}\right\|^\gamma, b\right\}$, and

$$f_2^*(x) = \begin{cases} b & \text{if } x_j = 0\ \forall j \notin \mathcal{V} \text{ and } x_j \geq 0\ \forall j \in \mathcal{V} \\ \min\left\{C\left\|(x)_{\mathcal{V}+}\right\|^\gamma, b\right\} & \text{otherwise.} \end{cases}$$

First, we argue that $f_2^* \in \Lambda_{+,\mathcal{V}}(0,\infty)$. To show this, we must show that for any $x, z \in \mathbb{R}^k$,

$$f_2^*(x) \geq f_2^*(z) \text{ if } x_j \geq z_j\ \forall j \in \mathcal{V} \text{ and } x_j = z_j\ \forall j \notin \mathcal{V}.$$

Note that this clearly holds if both $x$ and $z$ fall into the first case or second case, respectively, in the definition of $f_2^*$. Now, suppose $x$ falls into the first case and $z$ into the second. Then, it must be the case that $z_j \neq 0$ for some $j \notin \mathcal{V}$ or $z_j < 0$ for some $j \in \mathcal{V}$. If $z_j \neq 0$ for some $j \notin \mathcal{V}$, then the monotonicity condition holds vacuously. Suppose $z_j = 0$ for all $j \notin \mathcal{V}$ and $z_j < 0$ for some $j \in \mathcal{V}$. If $x_j < z_j$ for some $j \in \mathcal{V}$,then again the monotonicity condition holds vacuously. If $x_j \geq z_j$ for all $j \in \mathcal{V}$, then the monotonicity condition holds only if $f_2^*(x) \geq f_2^*(z)$, which is always the case because $f_2^*(z) \leq b$. Define $A_{\mathcal{V}} := \left\{x \in \mathbb{R}^k : x_j = 0\ \forall j \notin \mathcal{V} \text{ and } x_j \geq 0\ \forall j \in \mathcal{V}\right\}$. If $\mathcal{V} \subsetneq \{1,\ldots,k\}$ ,then $A_{\mathcal{V}}$ is a measure zero set under the Lebesgue measure[11]. Hence, under the assumption that the design points are a realization of a random variable that admits a pdf with respect to the Lebesgue measure, we may assume that $x_i \notin A_{\mathcal{V}}$ for all $i = 1,\ldots,n$. That is, we have

$$\omega^{-1}\left(b, \Lambda_{+,\mathcal{V}}(\gamma, C), \Lambda_{+,\mathcal{V}}(0,\infty)\right) = 0$$

for all $b \geq 0$. On the other hand, if $\mathcal{V} = \{1,\ldots,k\}$, we have

$$\begin{aligned} &\omega^{-1}\left(b, \Lambda_{+,\mathcal{V}}(\gamma, C), \Lambda_{+,\mathcal{V}}(0,\infty)\right) \\ &= \sum_{i=1}^n \left(1 - \frac{C}{b}\left\|(x_i)_{\mathcal{V}+}\right\|^\gamma\right)^2 \mathbb{1}\left(b - C\left\|(x_i)_{\mathcal{V}+}\right\|^\gamma > 0, x_i \in O_+\right), \end{aligned}$$

[11] Note that this is not the case when $\mathcal{V} = \{1,\ldots,k\}$

where $O_+ = \left\{x \in \mathbb{R}^k : x_j > 0 \ \forall j\right\}$. Likewise, we have

$$\begin{aligned}&\omega^{-1}\left(b, \Lambda_{+,\mathcal{V}}(0,\infty), \Lambda_{+,\mathcal{V}}(\gamma, C)\right)\\ =&\sum_{i=1}^{n}\left(1-\frac{C}{b}\left\|(x_i)_{\mathcal{V}-}\right\|^{\gamma}\right)^2 \mathbb{1}\left(b - C\left\|(x_i)_{\mathcal{V}-}\right\|^{\gamma} > 0, x_i \in O_-\right),\end{aligned}$$

where $O_- = \left\{x \in \mathbb{R}^k : x_j < 0 \ \forall j\right\}$. Hence, in this case, adaptation is possible and resulting CIs end up using only those data with design points that lie in either the positive or negative orthant.

# Appendix E Definition of the optimal upper CI

The following corollary summarizes an analogous result for the upper CI.

**Corollary E.1.** *Let* $\left(f^*_{j,\delta}, g^*_{J,\delta}\right) \in \mathcal{F}_j \times \mathcal{F}_J$ *solve the inverse modulus* $\omega\left(\delta, \mathcal{F}_j, \mathcal{F}_J\right)$ :

$$\begin{aligned}&\sum_{i=1}^{n}\left(g^*_{J,\delta}(x_i) - f^*_{j,\delta}(x_i)\right)^2 = \delta^2, \text{ and}\\ &Lg^*_{J,\delta} - Lf^*_{j,\delta} = \omega\left(\delta, \mathcal{F}_j, \mathcal{F}_J\right)\end{aligned}$$

*with* $\delta = z_\beta + z_{1-\alpha}$, *and define*

$$\begin{aligned}\hat{L}^{u,j}_{\delta} =& \frac{Lf^*_{j,\delta} + Lg^*_{J,\delta}}{2}\\ &+ \frac{\omega'\left(\delta, \mathcal{F}_j, \mathcal{F}_J\right)}{\delta} \times \sum_{i=1}^{n}\left(g^*_{J,\delta}(x_i) - f^*_{j,\delta}(x_i)\right)\left(y_i - \frac{f^*_{j,\delta}(x_i) + g^*_{J,\delta}(x_i)}{2}\right).\end{aligned}$$

*Then,* $\hat{c}^{u,j}_{\alpha} := \hat{L}^{u,j}_{\delta} + \frac{1}{2}\omega\left(\delta, \mathcal{F}_j, \mathcal{F}_J\right) - \frac{1}{2}\delta\omega'\left(\delta, \mathcal{F}_j, \mathcal{F}_J\right) + z_{1-\alpha}\omega'\left(\delta, \mathcal{F}_j, \mathcal{F}_J\right)$ *solves*

$$\min_{\hat{c}:(-\infty,\hat{c}]\in\mathcal{I}^J_{\alpha,1,+}} \sup_{f\in\mathcal{F}_j} q_{f,\beta}(\hat{c}^U - Lf).$$

*Moreover, we have*

$$\sup_{f\in\mathcal{F}_j} q_{f,\beta}\left(Lf - \hat{c}^{\ell,j}_{\alpha}\right) \le \omega\left(\delta, \mathcal{F}_j, \mathcal{F}_J\right).$$

*Especially, when $\beta = 1/2$, we have*

$$\sup_{f \in \mathcal{F}_j} \mathbf{E}_f \left( \hat{c}_\alpha^{u,j} - Lf \right) \leq \omega \left( z_{1-\alpha}, \mathcal{F}_j, \mathcal{F}_J \right).$$