CONDITIONAL QUANTILE PROCESSES BASED ON SERIES OR MANY REGRESSORS

ALEXANDRE BELLONI, VICTOR CHERNOZHUKOV, DENIS CHETVERIKOV, AND IVÁN FERNÁNDEZ-VAL

ABSTRACT. Quantile regression (QR) is a principal regression method for analyzing the impact of covariates on outcomes. The impact is described by the conditional quantile function and its functionals. In this paper we develop the nonparametric QR-series framework, covering many regressors as a special case, for performing inference on the entire conditional quantile function and its linear functionals. In this framework, we approximate the entire conditional quantile function by a linear combination of series terms with quantile-specific coefficients and estimate the function-valued coefficients from the data. We develop large sample theory for the QR-series coefficient process, namely we obtain uniform strong approximations to the QR-series coefficient process by conditionally pivotal and Gaussian processes. Based on these two strong approximations, or couplings, we develop four resampling methods (pivotal, gradient bootstrap, Gaussian, and weighted bootstrap) that can be used for inference on the entire QR-series coefficient function.

We apply these results to obtain estimation and inference methods for linear functionals of the conditional quantile function, such as the conditional quantile function itself, its partial derivatives, average partial derivatives, and conditional average partial derivatives. Specifically, we obtain uniform rates of convergence and show how to use the four resampling methods mentioned above for inference on the functionals. All of the above results are for function-valued parameters, holding uniformly in both the quantile index and the covariate value, and covering the pointwise case as a by-product. We demonstrate the practical utility of these results with an empirical example, where we estimate the price elasticity function and test the Slutsky condition of the individual demand for gasoline, as indexed by the individual unobserved propensity for gasoline consumption.

---

1. Introduction

Quantile regression (QR) is a principal method for analyzing the impact of covariates on outcomes, particularly when the impact may be heterogeneous. This impact is characterized by the quantile function of the conditional distribution of the outcome given covariates and its functionals Arias, Hallock and Sosa-Escudero [5], Buchinsky [13] and Koenker [51]. For example, we can model the log of the individual demand for some good, Y, as a function of the price of the good, the income of the individual, and other observed individual characteristics, X, and an unobserved preference for consuming the good, U, as

$$Y = Q(U, X),$$

where the function Q is strictly increasing in the unobservable U. With the normalization that $U \sim \text{Uniform}(0, 1)$ and the assumption that U and X are independent, the function $Q(u, x)$ is the u-th quantile of the conditional distribution of Y given $X = x$, i.e. $Q(u, x) = Q_{Y|X}(u|x)$. This function can be used for policy analysis. For example, we can determine how changes in taxes for the good could impact demand heterogeneously across individuals.

In this paper we develop the nonparametric QR-series framework for performing inference on the entire conditional quantile function $Q(u, x)$ and its linear functionals. In this framework, we approximate the entire conditional quantile function $Q(u, x)$ by a linear combination of series terms, $Z(x)'\beta(u)$. The vector $Z(x)$ includes transformations of x that have good approximation properties such as powers, trigonometrics, local polynomials, splines, and/or wavelets. The function $u \mapsto \beta(u)$ contains quantile-specific coefficients that can be estimated from the data using the QR estimator of Koenker and Bassett [52]. As the number of series terms grows, the approximation error $Q(u, x) - Z(x)'\beta(u)$ decreases, approaching zero in the limit. By controlling the growth of the number of terms, we can obtain consistent estimators and perform inference on the entire conditional quantile function and its linear functionals. The QR-series framework also covers as a special case the so called many regressors model, which is motivated by many new types of data that emerge in the new information age, such as scanner and online shopping data.

We describe now the main results in more detail. Let $u \mapsto \widehat{\beta}(u)$ denote the QR estimator of $u \mapsto \beta(u)$. The first set of results provides large-sample theory for the normalized *QR-series coefficient process* of increasing dimension $u \mapsto \sqrt{n}(\widehat{\beta}(u) - \beta(u))$ that can be used to perform inference on the function $u \mapsto \beta(u)$. We note that inference on the function $u \mapsto \beta(u)$, in particular simultaneous inference on the parameters $\beta(u)$ that holds uniformly over all $u \in \mathcal{U}$, where $\mathcal{U}$ is a set of quantile indices of interest, is difficult because the

standard asymptotic theory (van der Vaart and Wellner [80]) based on limit distributions does not help here as the process $u \mapsto \sqrt{n}(\widehat{\beta}(u) - \beta(u))$ in general does not have a limit distribution, even after an appropriate normalization. Instead, we develop high-quality inferential procedures based on the idea of the coupling. The coupling is a construction of two processes on the same probability space that are uniformly close to each other with high probability. Typically, one of the processes is the process of interest and the other one is a process whose distribution is known up-to a relatively small number of parameters that can be consistently estimated from the data. Thus, being able to construct an appropriate coupling means that we are able to approximate the distribution of the process of interest by simulating the distribution of the coupling process from the data.

In this paper, we develop two couplings: pivotal and Gaussian, that is, for each sample size n, we construct a pivotal process and a Gaussian process on the same probability space as the data that are uniformly close to the process $u \mapsto \sqrt{n}(\widehat{\beta}(u) - \beta(u))$ with high probability. In other words, these pivotal and Gaussian processes strongly approximate the process $u \mapsto \sqrt{n}(\widehat{\beta}(u) - \beta(u))$. In addition, we develop four resampling methods (pivotal, gradient bootstrap, Gaussian, and weighted bootstrap) that allow us to approximately simulate the distribution of the pivotal (first two methods) and the Gaussian (last two methods) processes. These results provide an inference theory for the function $u \mapsto \beta(u)$ and also help to develop a unified feasible inference theory for linear functionals of the conditional quantile functions $x \mapsto Q(u, x)$, $u \in \mathcal{U}$, using any of the four proposed resampling methods. To the best of our knowledge, all of the above results are new.

The existence of the pivotal coupling emerges from the special nature of QR, where a (sub) gradient of the sample objective function evaluated at the true values of $\beta(u)$ is pivotal conditional on the regressors, up-to an approximation error. This coupling allows us to perform high-quality inference based on pivotal and gradient bootstrap methods without even resorting to Gaussian approximations. We also show that the gradient bootstrap method, originally introduced by Parzen, Wei and Ying [68] in the parametric context, is effectively a means of carrying out the conditionally pivotal approximation without explicitly estimating Jacobian matrices, which may be difficult in the quantile regression context. The conditions for validity of the pivotal and gradient bootstrap methods require only a mild restriction on the growth of the number of series terms in relation to the sample size. To obtain the Gaussian coupling, we use chaining arguments and Yurinskii's construction. This coupling implies that one can use the Gaussian method to perform inference on the function $u \mapsto \beta(u)$, and we also use this coupling to show that the weighted bootstrap

method works to approximate the distribution of the whole QR-series coefficient process for the same reason as the Gaussian method. The conditions for validity of the Gaussian and weighted bootstrap methods, however, may be stronger than those for the pivotal and gradient bootstrap methods.

As a corollary of our results on the process $u \mapsto \sqrt{n}(\widehat{\beta}(u) - \beta(u))$, we also demonstrate that the QR-series estimator $x \mapsto Z(x)'\widehat{\beta}(u)$ of the function $x \mapsto Q(u, x)$ based on either polynomials or splines has the fastest possible rate of convergence in the L^2 norm and that the QR-series estimator based on splines has the fastest possible rate of convergence in the sup norm in Hölder smoothness classes. These findings on optimality of series estimators in the quantile regression setting complement results in the literature on optimality of series estimators in the mean regression setting; see Newey [67], Huang [49], Cattaneo and Farrell [15], Belloni et al [9], Chen and Christensen [22]. In particular, our result on optimality in the sup norm is a major extension of Huang's work [50] on mean regression, as it requires us to establish some fine properties of the QR-series approximation; see the next section for details.

The second set of results provides estimation and inference methods for linear functionals of the conditional quantile functions, including

(i) the conditional quantile function itself, $(u, x) \mapsto Q(u, x)$,

(ii) the partial derivative function, $(u, x) \mapsto \partial_{x_k} Q(u, x)$,

(iii) the average partial derivative function, $u \mapsto \int \partial_{x_k} Q(u, x) d\mu(x)$, and

(iv) the conditional average partial derivative, $(u, x_k) \mapsto \int \partial_{x_k} Q(u, x) d\mu(x|x_k)$,

where μ is a given measure and x_k is the k-th component of x. Specifically, we derive the pointwise rate of convergence and asymptotic normality of the QR-series estimators of the linear functionals. In addition, using our results on the QR-series coefficient process $u \mapsto \sqrt{n}(\widehat{\beta}(u) - \beta(u))$, we derive uniform rate of convergence and large sample inference procedures based on the pivotal, gradient bootstrap, Gaussian, and weighted bootstrap methods for the QR-series estimators of the linear functionals. These results provide solutions to a wide range of inference problems. For illustration purposes, we demonstrate how to use these results to construct uniform confidence bands for function-valued linear functionals and how to test shape constraints for the conditional quantile function $x \mapsto Q(u, x)$. It is noteworthy that all of the above results apply to function-valued parameters, holding uniformly in both the quantile index u and the covariate value x. We also emphasize that although we do not treat non-linear/non-smooth functionals in this paper, our results on

couplings and resampling methods for the process $u \mapsto \sqrt{n}(\widehat{\beta}(u) - \beta(u))$ are useful for treatment of such functionals.

The paper contributes and builds on the existing important literature on conditional quantile estimation. First and foremost, we build on the work of He and Shao [45] that studied the many regressors model and gave pointwise limit theorems for the QR estimator in the case where only one quantile index u is of interest. We go beyond the many regressors model to the series model and develop large sample estimation and inference results for the entire QR process. We also develop analogous estimation and inference results for the conditional quantile function and its linear functionals, such as derivatives, average derivatives, conditional average derivatives, and others. None of these results were available in the previous work. We also build on Lee [59] that studied QR estimation of partially linear models in the series framework for a single quantile index u, and on Horowitz and Lee [47] that studied nonparametric QR estimation of additive quantile models for a single quantile index u in a series framework. Our framework covers these partially linear models and additive models as important special cases, and allows us to perform inference on a considerably richer set of functionals, uniformly across covariate values and a continuum of quantile indices. In a more applied side, Koenker and Schorfheide [55] used a smoothing splines QR method to estimate the conditional quantile functions of global temperature over the last century. Other important work includes Stute [78], Chaudhuri [18], Chaudhuri, Doksum and Samarov [20], Härdle, Ritov, and Song [43], Horderlein and Mammen [46], Cattaneo, Crump, and Jansson [14], Kong, Linton, and Xia [57], Qu and Yoon [73], and Guerre and Sabbah [40], among others, but these papers focused on local, non-series, methods.

Our work also relies on the series literature, at least in a motivational and conceptual sense. In particular, we rely on the work of Stone [77], Andrews [2], Newey [67], Chen and Shen [24], Chen [21] and others that rigorously motivated the series framework as an approximation scheme and gave pointwise normality results for least squares estimators, and on Chen [21] and van de Geer [79] that gave (non-uniform) consistency and rate results for general series estimators, including quantile regression for the case of a single quantile index u. White [82] established non-uniform consistency of nonparametric estimators of the conditional quantile function based on a nonlinear series approximation using artificial neural networks. In contrast to the previous results, our rate results are uniform in covariate values and quantile indices, and cover both the quantile function and its functionals. Moreover, we not only provide estimation rate results, but also derive a full set of results on feasible inference based on the couplings for the process $u \mapsto \sqrt{n}(\widehat{\beta}(u) - \beta(u))$.

While relying on previous work for motivation, our results require to develop both new proof techniques and new approaches to inference. In particular, our proof techniques rely on new maximal inequalities for function classes with growing moments and uniform entropy. In addition, as explained above, and in contrast to previous papers, our inference results heavily rely on the idea of the couplings. Yurinskii's coupling was previously used by Chernozhukov, Lee and Rosen [30] to obtain a Gaussian coupling for the least squares series estimator, but the use of this technique in our context is new and much more involved. Thus, we approximate an entire QR-series coefficient process of an increasing dimension, instead of a vector of increasing dimension, by a Gaussian process. Finally, it is noteworthy that our uniform inference results on functionals, where uniformity is over covariate values, have not had analogs even in the least squares series literature until recently (the extension of our results to least squares has been recently published in Belloni et al [9]).

The results developed in this paper for series (global) estimation are of interest even though some of the results have analogs in the literature on kernel (local) estimation, because series estimators have several attractive features that are not shared by kernel estimators. First, series methods represent the estimate of the whole conditional quantile function $x \mapsto Q(u, x)$ and its linear functionals via a relatively small set of parameters, that is, the estimates $\widehat{\beta}(u)$ of the QR-series coefficients $\beta(u)$. As long as these parameters are reported, any researcher will be able to calculate the value of the estimate, for example, of $Q(u, x)$ for any $x \in \mathcal{X}$, which is very convenient. Second, series methods allow to easily impose shape constraints. For example, if it is known that the function $x \mapsto Q(u, x)$ is concave, one can simply impose this constraint on the QR-series optimization problem to obtain concave estimates (imposing shape constraints is especially convenient when B-splines are used; see DeVore and Lorentz [35]). Third, as we demonstrate in the empirical example in Section 5.1, when the vector X contains many controls in addition to one main covariate of interest, series methods are particularly convenient to impose a partial linear form on the functions $x \mapsto Q(u, x)$, which helps curb the course of dimensionality.

This paper does not deal with sparse models, where there are some key series terms and many "non-key" series terms which ideally should be omitted from estimation. In these settings, the goal is to find and indeed remove most of the "non-key" series terms before proceeding with estimation. Belloni and Chernozhukov [7] obtained rate results for quantile regression estimators in this case, but did not provide inference results. Even though our paper does not explicitly deal with inference in sparse models after model selection, the methods and bounds provided herein are useful for analyzing this problem. Investigating

this matter rigorously is a challenging issue, since it needs to take into account the model selection mistakes in estimation, and is beyond the scope of the present paper; however, it is a subject of our ongoing research, see Belloni, Chernozhukov and Kato [10].

Plan of the paper. The rest of the paper is organized as follows. In Section 2, we describe the nonparametric QR series model and estimators. In Section 3, we derive asymptotic theory for the QR series process. In Section 4, we give estimation and inference theory for linear functionals of the conditional quantile function. In Section 5, we present an empirical application to the demand of gasoline and a computational experiment calibrated to the application. The computational algorithms to implement our inference methods are collected in the Appendix. The Supplemental Material [8] contains some further results, and the proofs of the main results.

Notation. In what follows, for all $x = (x_1, \ldots, x_m)' \in \mathbb{R}^m$, we use $\|x\|$ to denote the Euclidean norm of x, that is, $\|x\| = (x_1^2 + \cdots + x_m^2)^{1/2}$, and we use $\|x\|_\infty$ to denote the sup norm of x, that is, $\|x\|_\infty = \max_{1 \leq j \leq m} |x_j|$. Also, we use S^{m-1} to denote the unit sphere in $\mathbb{R}^m$, that is, $S^{m-1} = \{x \in \mathbb{R}^m \colon \|x\| = 1\}$, and for $r > 0$, we use $B_m(0, r)$ to denote the ball in $\mathbb{R}^m$ with center at 0 and radius r, that is, $B_m(0, r) = \{x \in \mathbb{R}^m \colon \|x\| \leq r\}$. For all $m \times m$-dimensional matrices, we use $\|A\|$ to denote the operator norm of A (also known as the spectral norm), that is, $\|A\| = \sup_{\alpha \in S^{m-1}} \|A\alpha\|$. For a set I, $\text{diam}(I) = \sup_{v, \bar{v} \in I} \|v - \bar{v}\|$ denotes the diameter of I, and $\text{int}(I)$ denotes the interior of I. For any two real numbers a and b, $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. The relation $a_n \lesssim b_n$ means that the inequality $a_n \leq C b_n$ holds for all n with a constant C that is independent of n. We denote by P^* the probability measure induced by conditioning on realization of the data $\mathcal{D}_n = (Y_i, X_i)_{i=1}^n$. We say that a random variable $\Delta_n = o_{P^*}(1)$ in P-probability if for any $\epsilon > 0$, we have $P^*(|\Delta_n| > \epsilon) = o_P(1)$. We typically shall omit the qualifier "in P-probability." The operator E denotes the expectation with respect to the probability measure P, $\mathbb{E}_n$ denotes the expectation with respect to the empirical measure, and $\mathbb{G}_n$ denotes $\sqrt{n}(\mathbb{E}_n - E)$. Finally, we use ε' to denote a constant that depends only on the constant ε but is such that its value can change at each appearance.

2. Series Framework

2.1. **Model.** We consider the sequence of models indexed by the sample size n:

$$Y_{i,n} = Q_n(U_{i,n}, X_{i,n}), \quad i = 1, \ldots, n, \tag{2.1}$$

where $Y_{i,n}$ is a response variable, $X_{i,n}$ is a d_n-dimensional vector of covariates (elementary regressors) with the support $\mathcal{X}_n \subset \mathbb{R}^{d_n}$, $U_{i,n}$ is an unobservable individual ranking that is distributed uniformly on $(0,1)$ and is independent of $X_{i,n}$, and $Q_n \colon [0,1] \times \mathcal{X}_n \to \mathbb{R}$ is an unknown function such that for all $x \in \mathcal{X}_n$, the function $u \mapsto Q_n(u,x)$ is strictly increasing. Since $U_{i,n}|X_{i,n} \sim \mathrm{Uniform}(0,1)$, for each $u \in (0,1)$, $x \mapsto Q_n(u,x)$ is the uth quantile function of the conditional distribution of $Y_{i,n}$ given $X_{i,n}$, which we refer to as the conditional u-quantile function. For a compact set of quantile indices $\mathcal{U} \subset (0,1)$, we are interested in estimating the functions $x \mapsto Q_n(u,x)$ and their functionals. For given n, we assume that $(X_{i,n}, U_{i,n}, Y_{i,n})_{i=1}^n$ is a random sample from the distribution of the triple (X^n, U^n, Y^n), where n denotes the sample size.

The model (2.1) covers two specifications of major interest:

1. **Nonparametric model:** In this model, the data generating process does not depend on n, that is, for all $n \geq 1$, we have $Y^n = Y$, $X^n = X$, $U^n = U$, and $Q_n = Q$ for a response variable Y, a d-dimensional vector of covariates X with the support $\mathcal{X} \subset \mathbb{R}^d$, an unobservable individual ranking U satisfying $U|X \sim \mathrm{Uniform}(0,1)$, and a function $Q \colon [0,1] \times \mathcal{X} \to \mathbb{R}^d$ with the property that for all $x \in \mathcal{X}$, the function $u \mapsto Q(u,x)$ is strictly increasing. We assume that the functions $x \mapsto Q(u,x)$ are smooth but we do not impose any parametric structure on them. Throughout the paper, we refer to this specification as the NP model.

2. **Many regressors model:** In this model, the dimension d_n is allowed to grow with n but the function Q_n is linear in its second argument: $Q_n(u,x) = x'\beta_n(u)$ for some vector of coefficients $\beta_n(u) \in \mathbb{R}^{d_n}$ and all $u \in \mathcal{U}$ and $x \in \mathcal{X}_n$. Throughout the paper, we refer to this specification as the MR model.

Both models are of interest in econometrics. The NP model is important because it is very flexible as it does not impose any parametric structure on the functions $x \mapsto Q(u,x)$ and also does not require the function $(u,x) \mapsto Q(u,x)$ to be separately additive in u; see Matzkin [66] for extensive evidence on importance of this flexibility in economics. The MR model is also important, and different versions of this model have recently attracted much attention in the literature due to emergence of datasets with information on many variables; see Cattaneo, Jansson, and Newey [16] for some recent advances and also Mammen [64] for some classical results on the mean regression version of this model. As we demonstrate in this paper, both models can be treated in a unifying quantile regression series framework.

For brevity of notation, we shall omit the index n whenever it does not lead to confusion, that is, we write Y, X, U, Q, d, and $\mathcal{X}$ instead of Y^n, X^n, U^n, Q_n, d_n, and $\mathcal{X}_n$, respectively, even though we implicitly assume that all these quantities are allowed to depend on n. Also, we write $(X_i, U_i, Y_i)_{i=1}^n$ instead of $(X_{i,n}, U_{i,n}, Y_{i,n})_{i=1}^n$.

2.2. **QR-series estimator.** Next, we introduce the QR-series estimator of the function $x \mapsto Q(u, x)$. We start with preparing some notation. Fix $u \in \mathcal{U}$ and let $x \mapsto Z(x) = (Z_1(x), \ldots, Z_m(x))'$ be a vector of series approximating functions of dimension $m = m_n$, where each function $x \mapsto Z_j(x)$ maps $\mathcal{X}$ into $\mathbb{R}$. Define the vector of coefficients $\beta(u) = (\beta_1(u), \ldots, \beta_m(u))'$ as a solution to the *QR-series approximation problem*:

$$\min_{\beta \in \mathbb{R}^m} E\Big[\rho_u(Y - Z(X)'\beta) - \rho_u(Y - Q(u, X))\Big], \tag{2.2}$$

where $\rho_u(z) = (u - 1\{z < 0\})z$ is the check function (Koenker [51]).[1] For the MR model, we assume that $Z(x) = x$ for all $x \in \mathcal{X}$, so that $m = d$ and the vector $\beta(u)$ defined in (2.2) coincides with the vector $\beta_n(u)$ in the definition of the model, $Q_n(u, x) = x'\beta_n(u)$. For the NP model, we assume that the vector Z consists of series functions with good approximation properties such as indicators, B-splines (or regression splines), polynomials, Fourier series, and/or compactly supported wavelets.[2,3] We refer the reader to Newey [67] and Chen [21] for a careful and detailed description of these series functions; see also Belloni et al [9] for an overview of recent advances on series approximating functions.

Further, define the *QR-series approximating function* $x \mapsto Z(x)'\beta(u)$ mapping $\mathcal{X}$ into $\mathbb{R}$, and, for all $x \in \mathcal{X}$, let

$$R(u, x) = Q(u, x) - Z(x)'\beta(u)$$

be the *QR-series approximation error*. We will assume that the QR-series approximation error asymptotically vanishes, that is, $\sup_{x \in \mathcal{X}, u \in \mathcal{U}} |R(u, x)| \to 0$ as $n \to \infty$. For the MR model, this assumption always holds because $R(u, x) = 0$ for all $x \in \mathcal{X}$. For the NP model, we will demonstrate that this assumption holds under appropriate conditions as long as $m =$

---

[1]The optimization problem (2.2) has a finite solution if $E[|Q(u, X)|]$ is finite. In addition, since the function $z \mapsto \rho_u(z)$ is strictly convex, the solution is unique if the matrix $E[Z(X)Z(X)']$ is non-singular, which is assumed in Condition S below. The term $\rho_u(Y - Q(u, X))$ does not affect the optimization problem but guarantees the existence of the solution when $E[|Y|]$ is not finite.

[2]Interestingly, in the case of B-splines and compactly supported wavelets, the entire collection of series terms is dependent upon the sample size n.

[3]It is possible to combine these sets of approximating functions. For example, when we model gasoline consumption, we can simultaneously use Fourier series to capture seasonal effects and polynomials to capture long term growth (we thank a referee for pointing out this example).

$m_n \to \infty$ as $n \to \infty$. In turn, given that the QR-series approximation error asymptotically vanishes, it follows that the QR-series approximating function $x \mapsto Z(x)'\beta(u)$ approximates well the true conditional u-quantile function $x \mapsto Q(u,x)$. This motivates the *QR-series estimator* of the function $x \mapsto Q(u,x)$:

$$\widehat{Q}(u,x) = Z(x)'\widehat{\beta}(u), \quad x \in \mathcal{X}, \tag{2.3}$$

where $\widehat{\beta}(u)$ is the Koenker and Bassett [52] estimator of $\beta(u)$ that solves the empirical analog of the population problem (2.2):

$$\min_{\beta \in \mathbb{R}^m} \mathbb{E}_n[\rho_u(Y_i - Z_i'\beta)], \tag{2.4}$$

where we denote $Z_i = Z(X_i)$ for all $i = 1,\ldots,n$. As n gets large, both the estimation error $\widehat{Q}(u,x) - Z(x)'\beta(u)$ and the approximation error $R(u,x)$ asymptotically vanish.

Since we are interested in estimating the functions $x \mapsto Q(u,x)$ for a set of quantile indices $\mathcal{U}$, we solve the problem (2.4) for all $u \in \mathcal{U}$ to obtain the *QR-series coefficient process*

$$\widehat{\beta}(\cdot) = \{\widehat{\beta}(u) \colon u \in \mathcal{U}\}$$

and the QR-series estimator (2.3) for all $u \in \mathcal{U}$. We note that obtaining this estimator is computationally easy even if $\mathcal{U}$ contains many quantile indices and the dimension m of the vectors Z_i is large. In particular, one can use the results of Portnoy and Koenker [70], who developed interior points methods with preprocessing for the problem (2.4) that are very efficient and give the solution for multiple quantile indices simultaneously.

2.3. **Functionals.** In addition to the functions $x \mapsto Q(u,x)$, $u \in \mathcal{U}$, we are also interested in various linear functionals of these functions. If x is decomposed as (w,v) and x_k denotes the k-th component of x, examples of particularly useful linear functionals include

1. the derivative: $\quad \theta(u,x) = \partial_{x_k} Q(u,x)$;
2. the average derivative: $\quad \theta(u) = \int \partial_{x_k} Q(u,x) d\mu(x)$;
3. the conditional average derivative: $\quad \theta(u,w) = \int \partial_{x_k} Q(u,w,v) d\mu(v|w)$.

The measures μ entering the definitions above are assumed to be known but our results can also be extended to include estimated measures. To cover all examples, we denote the linear functional of interest by $\theta(u,w)$, where $w \in \mathcal{W} \subset \mathbb{R}^{d_w}$.

Let $I \subset \mathcal{U} \times \mathcal{W}$ denote the set of values of (u,w) of interest. For example, if we are interested in

- the function $\theta(u,w)$ at a particular point (u,w), then $I = \{(u,w)\}$,

- the function $u \mapsto \theta(u, w)$ having fixed w, then $I = \mathcal{U} \times \{w\}$,
- the function $w \mapsto \theta(u, w)$ having fixed u, then $I = \{u\} \times \mathcal{W}$,
- the entire function $(u, w) \mapsto \theta(u, w)$, then $I = \mathcal{U} \times \mathcal{W}$.

By the linearity of the series approximations, the function $\theta(u, w)$ can be seen as a linear functional of the quantile regression coefficients $\beta(u)$ up to an approximation error, that is,

$$\theta(u, w) = \ell(w)' \beta(u) + r(u, w), \quad (u, w) \in I, \tag{2.5}$$

where $\ell(w)' \beta(u)$ is the QR-series approximation term, with $\ell(w)$ denoting the m-dimensional vector of loadings on the coefficients, and $r(u, w)$ is the remainder term, which corresponds to the QR-approximation error. Indeed, this decomposition arises from the application of different linear operators $\mathcal{A}$ to the decomposition $Q(u, \cdot) = Z(\cdot)' \beta(u) + R(u, \cdot)$ and evaluating the resulting functions at w:

$$(\mathcal{A}Q(u, \cdot))\,[w] = (\mathcal{A}Z(\cdot))\,[w]' \beta(u) + (\mathcal{A}R(u, \cdot))\,[w]. \tag{2.6}$$

In the three examples above the operator $\mathcal{A}$ is given by, respectively,

1. a differential operator: $(\mathcal{A}g)[x] = (\partial_{x_k} g)[x]$, so that

$$\ell(x) = \partial_{x_k} Z(x), \quad r(u, x) = \partial_{x_k} R(u, x);$$

2. an integro-differential operator: $\mathcal{A}g = \int \partial_{x_k} g(x) d\mu(x)$, so that

$$\ell = \int \partial_{x_k} Z(x) d\mu(x), \quad r(u) = \int \partial_{x_k} R(u, x) d\mu(x);$$

3. a partial integro-differential operator: $(\mathcal{A}g)[w] = \int \partial_{x_k} g(w, v) d\mu(v|w)$, so that

$$\ell(w) = \int \partial_{x_k} Z(w, v) d\mu(v|w), \quad r(u, w) = \int \partial_{x_k} R(u, w, v) d\mu(v|w).$$

For notational convenience, we use the formulation (2.5) in the analysis, instead of the motivational formulation (2.6).

Given $\widehat{Q}(u, x) = Z(x)' \widehat{\beta}(u)$, we use the plug-in estimator

$$\widehat{\theta}(u, w) = \ell(w)' \widehat{\beta}(u), \quad (u, w) \in I,$$

to estimate $\theta(u, w)$. We will provide rates of convergence for $\widehat{\theta}(u, w)$ as well as the inference tools that will be valid for inference on the QR-series approximation

$$\ell(w)' \beta(u), \quad (u, w) \in I,$$

and, provided that the QR-approximation error $r(u, w)$ is small enough relative to the estimation noise, will also be valid for inference on the functional of interest:

$$\theta(u, w), \quad (u, w) \in I.$$

Thus, the QR-series approximation $\ell(w)'\beta(u)$ is an important penultimate target, whereas the functional $\theta(u, w)$ is the ultimate target.

2.4. **Main regularity conditions.** Let $\kappa \in (0, \infty]$ be some constant that is independent of n. Also, for $x \in \mathcal{X}$, let $\mathcal{Y}_x$ denote the support of the conditional distribution of Y given $X = x$. Moreover, let $\bar{\mathcal{U}}$ denote the convex hull of $\mathcal{U}$. Throughout the paper, we will use the following regularity condition:

Condition S.

S.1 *The data form a triangular array of random variables so that for any given n, the data $\mathcal{D}_n = \{(X_i, Y_i) : 1 \leq i \leq n\}$ is an i.i.d. random sample from the distribution of the pair (X, Y).*

S.2 *The conditional density $f_{Y|X}(y|x)$ is bounded from above uniformly over $y \in \mathcal{Y}_x$, $x \in \mathcal{X}$, and n; also, $f_{Y|X}(Q(u, x)|x)$ is bounded away from zero uniformly over $u \in \bar{\mathcal{U}}$, $x \in \mathcal{X}$, and n; moreover, the derivative of $y \mapsto f_{Y|X}(y|x)$ is continuous and bounded in absolute value from above uniformly over $y \in \mathcal{Y}_x$, $x \in \mathcal{X}$, and n.*

S.3 *The eigenvalues of the Gram matrix $\Sigma = E[Z(X)Z(X)']$ are bounded from above and away from zero uniformly over n.*

S.4 *The approximation error $R(u, x)$ satisfies $\sup_{x \in \mathcal{X}, u \in \mathcal{U}} |R(u, x)| \lesssim m^{-\kappa}$.*

Condition S.1 requires that the data is i.i.d. but it can be extended to standard time series models at the expense of more technicalities. Condition S.2 imposes mild smoothness assumptions on the conditional density function $f_{Y|X}(y|x)$. Since it follows from simple algebra, see for example (H.116), that

$$\frac{1}{f_{Y|X}(y|x)} = \frac{\partial Q(Q^{-1}(y, x), x)}{\partial u}, \quad y \in \mathcal{Y}_x, \ x \in \mathcal{X},$$

where $y \mapsto Q^{-1}(y, x)$ denotes the inverse of $u \mapsto Q(u, x)$, it is easy to provide a set of conditions in terms of the function $Q(u, x)$ that imply Condition S.2. Indeed, Condition S.2 follows if (i) $\partial Q(u, x)/\partial u$ is bounded away from zero uniformly over $u \in [0, 1]$, $x \in \mathcal{X}$, and n; (ii) $\partial Q(u, x)/\partial u$ is bounded from above uniformly over $u \in \bar{\mathcal{U}}$, $x \in \mathcal{X}$ and n; (iii)

$\partial^2 Q(u,x)/\partial u^2$ is bounded in absolute value from above uniformly over $u \in [0,1]$, $x \in \mathcal{X}$, and n.[4]

For the MR model, Condition S.3 implies that there is no perfect multicollinearity among covariates, and Condition S.4 is satisfied with $\kappa = \infty$ since $R(u,x) = 0$ for all $u \in \mathcal{U}$ and $x \in \mathcal{X}$.

For the MR model, all conditions can be regarded as primitive. For the NP model, Conditions S.1 and S.2 are also primitive but Conditions S.3 and S.4 depend on the vector of approximating series functions $x \mapsto Z(x)$ used for the estimation. Therefore, below we provide some discussion of these conditions in the NP model. Suppose that X is absolutely continuous with respect to the Lebesgue measure on $\mathcal{X}$ and let $f_X \colon \mathcal{X} \to \mathbb{R}$ denote its pdf. Then it is well-known that Condition S.3 holds if $f_X(x)$ is bounded from above and away from zero uniformly over $x \in \mathcal{X}$, and the eigenvalues of the matrix

$$\int_{x \in \mathcal{X}} Z(x)Z(x)' dx$$

are bounded from above and away from zero uniformly over n; see, for example, Proposition 2.1 in Belloni et al [9]. In turn, the latter condition holds if, for example, the vector Z consists of functions that are orthonormal on $\mathcal{X}$. In the case that the former condition is violated in the sense that the density $f_X(x)$ is not bounded away from zero uniformly over all $x \in \mathcal{X}$, one can consider a subset $\widetilde{\mathcal{X}}$ of $\mathcal{X}$ such that $f(x)$ is bounded away from zero uniformly over $x \in \widetilde{\mathcal{X}}$ and consider the estimation problem based on the subset of observations i satisfying $X_i \in \widetilde{\mathcal{X}}$. This will give the estimate of $Q(u,x)$ for all $x \in \widetilde{\mathcal{X}}$ and $u \in \mathcal{U}$. As the sample size gets larger, one can increase the set $\widetilde{\mathcal{X}}$ to extend the estimate of $Q(u,x)$ to a larger set of points. Developing a method how this truncation should be performed in practice, however, is beyond the scope of this paper.

To provide some primitive conditions for Condition S.4 in the NP model, we need to prepare some notation. For a d-tuple $\alpha = (\alpha_1, \ldots, \alpha_d)$ of nonnegative integers, let $D^\alpha = \partial_{x_1}^{\alpha_1} \cdots \partial_{x_d}^{\alpha_d}$. Also, for $s > 0$, let $[s]$ denote the largest integer strictly smaller than s. For the constant $C > 0$, define the Hölder ball $\Omega(s, C, \mathcal{X})$ as the set of all functions $f \colon \mathcal{X} \to \mathbb{R}$

---

[4]Note that we assume that the conditional density $f_{Y|X}(y|x)$ is bounded away from zero only for $y = Q(u,x)$, where $u \in \bar{\mathcal{U}}$ and $x \in \mathcal{X}$. This allows us to avoid the stronger condition that assumes that $f_{Y|X}(y|x)$ is bounded away from zero for all $y \in \mathcal{Y}_x$ and $x \in \mathcal{X}$. The latter condition can simplify some arguments (see the proof of Lemma 2) but it requires the conditional density of Y given X to have bounded support, thus excluding some important distributions such as the Gaussian.

such that

$$|D^\alpha f(x) - D^\alpha f(\widetilde{x})| \le C\left(\sum_{j=1}^d (x_j - \widetilde{x}_j)^2\right)^{(s-[s])/2} \text{ and } |D^\beta f(x)| \le C \qquad (2.7)$$

for all $x = (x_1, \ldots, x_d)'$ and $\widetilde{x} = (\widetilde{x}_1, \ldots, \widetilde{x}_d)'$ in $\mathcal{X}$ and all d-tuples $\alpha = (\alpha_1, \ldots, \alpha_d)$ and $\beta = (\beta_1, \ldots, \beta_d)$ of nonnegative integers satisfying $\alpha_1 + \cdots + \alpha_d = [s]$ and $\beta_1 + \cdots + \beta_d \le [s]$ (where the left-hand sides of the inequalities in (2.7) are set to be infinity if the derivatives do not exist). For example, any s-times continuously differentiable function belongs to the Hölder ball $\Omega(s, C, \mathcal{X})$ for some $C > 0$ as long as $\mathcal{X}$ is compact. Also, we say that the vector of approximating functions Z consists of tensor products of polynomials if $m = J^d$ for some integer $J > 0$ and Z consists of all functions of the form $x = (x_1, \ldots, x_d) \mapsto \prod_{j=1}^d x_j^{\alpha_j}$ for some d-tuple $\alpha = (\alpha_1, \ldots, \alpha_d)$ of nonnegative integers such that $\alpha_j \le J - 1$ for all $j = 1, \ldots, d$. Finally, we say that the vector of approximating functions Z consists of tensor products of B-splines of order s_0 if $m = J^d$ for some integer $J > 0$ and Z consists of all functions of the form $x = (x_1, \ldots, x_d) \mapsto \prod_{j=1}^d b_{\alpha_j}(x_j)$ for some d-tuple $\alpha = (\alpha_1, \ldots, \alpha_d)$ of nonnegative integers such that $\alpha_j \le J - 1$ for all $j = 1, \ldots, d$ where $b_0, \ldots, b_{J-1}$ is a sequence of J B-splines of order s_0 on the interval $[0, 1]$ with uniform knot sequence; see Chen [21] for more explanations about Hölder balls and B-splines. The next lemma provides a set of primitive conditions for Condition S.4 in the NP model.

Lemma 1 (Verification of Condition S.4 in the NP model for polynomials and B-splines). *Consider the NP model. Suppose that Conditions S.2 and S.3 hold. In addition, suppose that $\mathcal{X} = [0, 1]^d$. Moreover, suppose that $Q(u, \cdot) \in \Omega(s, C, \mathcal{X})$ for all $u \in \mathcal{U}$ and some $s, C > 0$. If the vector of approximating functions Z consists of tensor products of polynomials and $s > d$, then*

$$(E[|R(u, X)|^2])^{1/2} \lesssim m^{-s/d} \text{ and } \sup_{x \in \mathcal{X}} |R(u, x)| \lesssim m^{1-s/d}, \qquad (2.8)$$

uniformly over $u \in \mathcal{U}$. Also, if the vector of approximating functions Z consists of tensor products of B-splines of order s_0, $s \wedge s_0 > d$, and X has the pdf $f_X(x)$ bounded from above and away from zero uniformly over $x \in \mathcal{X}$, then

$$(E[|R(u, X)|^2])^{1/2} \lesssim m^{-(s\wedge s_0)/d} \text{ and } \sup_{x \in \mathcal{X}} |R(u, x)| \lesssim m^{-(s\wedge s_0)/d}, \qquad (2.9)$$

uniformly over $u \in \mathcal{U}$. Thus, under the presented conditions, Condition S.4 is satisfied with $\kappa = s/d - 1$ in the case of polynomials and with $\kappa = (s \wedge s_0)/d$ in the case of B-splines.

Comment 1 (Importance of Lemma 1). Lemma 1 plays a crucial role in our derivation of the convergence rate of the QR-series estimator for the NP model in the next

section. Indeed, it is well-known from the approximation theory that under the assumptions of the lemma, in the case of polynomials, for example, there exists $\beta^m(u)$ such that $\sup_{x \in \mathcal{X}} |Q(u,x) - Z(x)'\beta^m(u)| \lesssim m^{-s/d}$; see Chen [21]. However, this result does not help in our analysis because the QR-series estimator $\widehat{\beta}_n(u)$ converges in probability to $\beta(u)$, which may or may not be equal to $\beta^m(u)$. We therefore need to derive a bound on $\sup_{x \in \mathcal{X}} |Q(u,x) - Z(x)'\beta(u)|$.

The part of the lemma concerning the B-splines case is particularly important because it allows us to prove in the next section that the QR-series estimator based on B-splines achieves the fastest possible rate of convergence in the sup norm. This part of the lemma is a major extension of a result in Huang [50], who obtained similar inequalities with the QR-series approximation error replaced by the least-squares-series approximation error. □

Comment 2 (Other series approximating functions)**.** Other popular choices of the series approximating functions include Fourier series and compactly supported wavelets. Although we do not provide formal results for these choices, we note that under conditions similar to those in Lemma 1, one can show that Condition S.4 holds with $\kappa = 1/2 - s/d$ in the case of Fourier series and with $\kappa = -(s \wedge s_0)/d$ in the case of compactly supported wavelets, where s_0 is the order of the wavelets. □

2.5. **Some additional notation.** The properties of the QR-series coefficient process and of the QR-series estimator depend on the choice of the approximating functions and the dimension of Z. Like in the analysis of series estimators of conditional mean functions (see Newey [67]), the following quantity will play a crucial role in our analysis:

$$\zeta_m = \sup_{x \in \mathcal{X}} \|Z(x)\|.$$

Assuming that $\mathcal{X} = [0,1]^d$, it is well known that $\zeta_m \lesssim m$ if the vector Z consists of tensor products of polynomials and $\zeta_m \lesssim m^{1/2}$ if the vector Z consists of tensor products of B-splines.

As in the analysis of the parametric quantile regression, the following Jacobian matrix will also play a crucial role in the analysis:

$$J(u) = E\Big[f_{Y|X}(Q(u,X)|X) Z(X) Z(X)' \Big], \quad u \in \mathcal{U}. \tag{2.10}$$

Implementing some of our inference methods will require an estimator of $J(u)$. For the purposes of this paper, we will use Powell's [71] estimator defined by

$$\widehat{J}(u) = \frac{1}{2h} \mathbb{E}_n \Big[1\{|Y_i - Z_i'\widehat{\beta}(u)| \le h\} \cdot Z_i Z_i' \Big], \tag{2.11}$$

where h is some bandwidth value satisfying $h = h_n \to 0$. We will also use the estimator of Σ defined by

$$\widehat{\Sigma} = \mathbb{E}_n[Z_i Z_i'].\tag{2.12}$$

The properties of $\widehat{J}(u)$ and $\widehat{\Sigma}$ in our high-dimensional setting are established in Lemma 30 in Appendix H of the Supplemental Material.

3. Asymptotic Theory for QR Coefficient Processes based on Series

In this section, we study properties of the normalized QR-series coefficient process

$$\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot)) = \left\{ \sqrt{n}(\widehat{\beta}(u) - \beta(u)) \colon u \in \mathcal{U} \right\}.$$

Specifically, we derive the rate of convergence and construct two couplings for this process. The couplings give two processes that are uniformly close to $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ with high probability, but are such that their distribution can be simulated. In particular, we develop four resampling methods (pivotal, gradient bootstrap, Gaussian, and weighted bootstrap) to simulate the distribution of these processes. In the next section, these results allow us to develop a unified feasible inference theory for all the functionals of interest. We also derive rates of convergence for the QR-series estimator process $\widehat{Q}(\cdot, \cdot) = \{\widehat{Q}(u, x) \colon u \in \mathcal{U}, x \in \mathcal{X}\}$ in the NP model. In particular, we show that the QR-series estimator based on either polynomials or B-splines has the fastest possible rate of convergence in the L^2 norm and that the QR-series estimator based on B-splines has the fastest possible rate of convergence in the sup norm.

3.1. **Uniform-in-u Rate of Convergence.** As explained in the previous section, given an i.i.d. sample $(X_i, Y_i)_{i=1}^n$ from the distribution of the pair (X, Y), we estimate the coefficient function $\beta(\cdot) = \{\beta(u) \colon u \in \mathcal{U}\}$ using the QR-series coefficient process $\widehat{\beta}(\cdot) = \{\widehat{\beta}(u) \colon u \in \mathcal{U}\}$, namely, for each $u \in \mathcal{U}$, we define $\widehat{\beta}(u)$ as the Koenker and Bassett [52] estimator that solves the empirical analog (2.4) of the population problem (2.2). Our first main result is a uniform-in-u rate of convergence for the QR-series coefficient process.

Theorem 1 (Uniform-in-u rate of convergence for QR-Series coefficient process). *Suppose that Condition S holds. In addition, suppose that $m\zeta_m^2 \log^2 n = o(n)$ and $m^{-\kappa} \log n = o(1)$. Then*

$$\sup_{u \in \mathcal{U}} \|\widehat{\beta}(u) - \beta(u)\| \lesssim_P \sqrt{m/n}.$$

Theorem 1 establishes a rate of convergence of the estimator $\widehat{\beta}(u)$ in our high-dimensional setting that holds uniformly over $u \in \mathcal{U}$. The theorem complements the rate of convergence results in the literature. Indeed, Koenker and Portnoy [54] established rate of convergence results that hold uniformly over $\mathcal{U}$ in the fixed-dimensional setting, and He and Shao [45] established rates in the high-dimensional setting for the case when $\mathcal{U}$ is a singleton (pointwise-in-u rate of convergence). Importantly, the uniform-in-u rate of convergence in Theorem 1 is the same as the pointwise-in-u rate of convergence. The proof of this theorem relies on new concentration inequalities that control the behavior of the eigenvalues of the design matrix $\widehat{\Sigma}$.

Theorem 1 has an implication for the uniform-in-u rate of convergence in the L^2 norm of the QR-series estimator in the NP model. Indeed, define

$$\|h\|_{L^2(X)} = \left(E[|h(X)|^2] \right)^{1/2}, \quad \text{for } h \colon \mathcal{X} \to \mathbb{R}.$$

We then have the following corollary of Theorem 1, which is the second main result together with Corollary 2 below on the uniform-in-u rate of convergence in the sup norm of the QR-series estimator in the NP model.

Corollary 1 (Uniform-in-u L^2 rate of convergence for QR-series estimator in the NP model). *Consider the NP model. Suppose that (i) Condition S.1-3 holds. In addition, suppose that (ii) $\mathcal{X} = [0,1]^d$ and that (iii) $Q(u, \cdot) \in \Omega(s, C, \mathcal{X})$ for all $u \in \mathcal{U}$ and some $s, C > 0$. If the vector of approximating functions Z consists of tensor products of polynomials, $m^3 \log^2 n = o(n)$, and $m^{1-s/d} \log n = o(1)$, then*

$$\sup_{u \in \mathcal{U}} \|\widehat{Q}(u, \cdot) - Q(u, \cdot)\|_{L^2(X)} \lesssim_P \sqrt{m/n} + m^{-s/d}. \tag{3.13}$$

Also, if the vector of approximating functions Z consists of tensor products of B-splines of order s_0, $s \wedge s_0 > d$, $m^2 \log^2 n = o(n)$, $m^{-(s \wedge s_0)/d} \log n = o(1)$, and X has the pdf $f_X(x)$ bounded from above and away from zero uniformly over $x \in \mathcal{X}$, then

$$\sup_{u \in \mathcal{U}} \|\widehat{Q}(u, \cdot) - Q(u, \cdot)\|_{L^2(X)} \lesssim_P \sqrt{m/n} + m^{-(s \wedge s_0)/d}. \tag{3.14}$$

Comment 3 (QR-series estimator achieves the fastest possible rate of convergence in the L^2 norm). Consider the NP model and suppose that conditions (i)–(iii) of Corollary 1 hold. If Z consists of a tensor product of polynomials and $s > d$, setting $m = Cn^{d/(d+2s)}$ for some constant $C > 0$ satisfies conditions that $m^3 \log^2 n = o(n)$ and $m^{1-s/d} \log n = o(1)$, and so substituting this m into the bound (3.13) gives

$$\sup_{u \in \mathcal{U}} \|\widehat{Q}(u, \cdot) - Q(u, \cdot)\|_{L^2(X)} \lesssim_P n^{-s/(d+2s)}. \tag{3.15}$$

Similarly, if Z consists of a tensor product of B-splines of order s_0, $s_0 \geq s > d$, and X has the pdf $f_X(x)$ bounded from above and away from zero uniformly over $x \in \mathcal{X}$, setting $m = Cn^{d/(d+2s)}$ for some constant $C > 0$ satisfies conditions that $m^2 \log^2 n = o(n)$ and $m^{-(s \wedge s_0)/d} \log n = o(1)$, and so substituting this m into the bound (3.14) again gives (3.15). Note that the rate in (3.15) is the optimal L^2 rate of convergence for the estimators of nonparametric conditional quantile functions; see Chaudhuri [19]. Thus, the QR-series estimator based on either polynomials or B-splines has the fastest possible L^2 rate of convergence, and as we demonstrate, this rate is actually achieved uniformly in $u \in \mathcal{U}$. The same results can also be shown for the QR-series estimator based on Fourier series and compactly supported wavelets. This is one of the attractive properties of the QR-series estimator. □

3.2. Uniform Strong Approximations (Couplings) and Resampling Methods.

Here we derive two couplings yielding strong approximations to the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ in the form of either a pivotal or a Gaussian process, and develop four resampling methods to approximate the distribution of these processes and thus approximate also the distribution of the original process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$.

Let

$$\mathbb{U}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i(u - 1\{U_i \leq u\}), \quad u \in \mathcal{U}. \tag{3.16}$$

Note that the process $\mathbb{U}(\cdot) = \{\mathbb{U}(u) : u \in \mathcal{U}\}$ is (conditionally) pivotal since conditional on $(Z_i)_{i=1}^{n}$, the sequence $(U_i)_{i=1}^{n}$ consists of i.i.d. Uniform$(0,1)$ random variables. The following theorem, which is the third main result together with the Gaussian coupling in Theorem 5 below, shows that the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ is strongly approximated by the (conditionally) pivotal process $J^{-1}(\cdot)\mathbb{U}(\cdot) = \{J^{-1}(u)\mathbb{U}(u) : u \in \mathcal{U}\}$.

Theorem 2 (Pivotal Coupling). *Suppose that Condition S holds. In addition, suppose that* $m^3 \zeta_m^2 = o(n^{1-\varepsilon})$ *and* $m^{-\kappa+1} = o(n^{-\varepsilon})$ *for some constant* $\varepsilon > 0$. *Then*

$$\sqrt{n}\left(\widehat{\beta}(u) - \beta(u)\right) = J^{-1}(u)\mathbb{U}(u) + r(u), \quad u \in \mathcal{U},$$

where

$$\sup_{u \in \mathcal{U}} \|r(u)\| \lesssim_P \frac{m^{3/4}\zeta_m^{1/2} \log^{1/2} n}{n^{1/4}} + \sqrt{m^{1-\kappa} \log n} = o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$.

This theorem is important because it has many useful implications. One of the implications is the following result for the uniform-in-u rate of convergence in the sup norm of the QR-series estimator in the NP model.

Corollary 2 (Uniform-in-u sup-rate of convergence for QR-series estimator in the NP model). *Consider the NP model. Suppose that (i) Condition S.1-3 holds. In addition, suppose that (ii) $\mathcal{X} = [0,1]^d$ and that (iii) $Q(u,\cdot) \in \Omega(s, C, \mathcal{X})$ for all $u \in \mathcal{U}$ and some $s, C > 0$. If the vector of approximating functions Z consists of tensor products of polynomials and for some $\varepsilon > 0$, $m^5 = o(n^{1-\varepsilon})$ and $m^{2-s/d} = o(n^{-\varepsilon})$, then*

$$\sup_{u \in \mathcal{U}} \sup_{x \in \mathcal{X}} |\widehat{Q}(u,x) - Q(u,x)| \lesssim_P \sqrt{m^2 \log n / n} + m^{1-s/d}.$$

Also, if the vector of approximating functions Z consists of tensor products of B-splines of order s_0, X has the pdf $f_X(x)$ bounded from above and away from zero uniformly over $x \in \mathcal{X}$, and for some $\varepsilon > 0$, $m^4 = o(n^{1-\varepsilon})$ and $m^{1-(s \wedge s_0)/d} = o(n^{-\varepsilon})$, then

$$\sup_{u \in \mathcal{U}} \sup_{x \in \mathcal{X}} |\widehat{Q}(u,x) - Q(u,x)| \lesssim_P \sqrt{m \log n / n} + m^{-(s \wedge s_0)/d}. \tag{3.17}$$

Comment 4 (B-splines version of QR-series estimator achieves the fastest possible rate of convergence in the sup norm). Consider the NP model and suppose that conditions (i)–(iii) of Corollary 2 hold. In addition, suppose that Z consists of a tensor product of B-splines of order s_0, $s_0 \geq s > 3d/2$, and X has the pdf $f_X(x)$ bounded from above and away from zero uniformly over $x \in \mathcal{X}$. Then setting $m = C(n/\log n)^{d/(d+2s)}$ for some constant $C > 0$ satisfies conditions that $m^4 = o(n^{1-\varepsilon})$ and $m^{1-(s \wedge s_0)/d} = o(n^{-\varepsilon})$ for some $\varepsilon > 0$, and so substituting this m into the bound (3.17) gives

$$\sup_{u \in \mathcal{U}} \sup_{x \in \mathcal{X}} |\widehat{Q}(u,x) - Q(u,x)| \lesssim_P \left(\frac{\log n}{n} \right)^{s/(d+2s)},$$

which is the optimal rate of convergence in the sup norm for an estimator of the nonparametric conditional quantile function; see Chaudhuri [19]. Thus, the QR-series estimator based on B-splines has the fastest possible rate of convergence in the sup norm, and as we demonstrate, this rate is actually achieved uniformly in $u \in \mathcal{U}$.[5] This is another attractive property of the QR-series estimator. □

We also note that although the uniform convergence rate based on polynomials is not optimal, the rate derived in Corollary 2 is faster than the (trivial) uniform rate implied by the L_2 rate and the relation between the L_2-norm and sup-norm.

---

[5]The same results can also be shown for the QR-series estimator based on compactly supported wavelets.

Another implication of Theorem 2 is that it suggests the following high-quality method to approximate the distribution of the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$, which we refer to as the pivotal method. First, simulate an i.i.d. sequence $(U_i^*)_{i=1}^n$ of Uniform$(0,1)$ random variables that are independent of the data and define

$$\mathbb{U}^*(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(u - 1\{U_i^* \leq u\}), \quad u \in \mathcal{U}, \tag{3.18}$$

so that conditional on $(Z_i)_{i=1}^n$, the process $\mathbb{U}^*(\cdot) = \{\mathbb{U}^*(u) : u \in \mathcal{U}\}$ is a copy of the process $\mathbb{U}(\cdot)$, and $J^{-1}(\cdot)\mathbb{U}^*(\cdot) = \{J^{-1}(u)\mathbb{U}^*(u) : u \in \mathcal{U}\}$ is a copy of $J^{-1}(\cdot)\mathbb{U}(\cdot)$. Second, calculate the estimators $\widehat{J}(u)$ of the matrices $J(u)$ for all $u \in \mathcal{U}$ as in (2.11) of Section 2.5. Then, as shown in the next theorem, one can use the conditional distribution of the process $\widehat{J}^{-1}(\cdot)\mathbb{U}^*(\cdot) = \{\widehat{J}^{-1}(u)\mathbb{U}^*(u) : u \in \mathcal{U}\}$ given the data, which can be simulated, to approximate the distribution of the process $J^{-1}(\cdot)\mathbb{U}^*(\cdot)$, and, via Theorem 2, also of the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$.

Theorem 3 (Pivotal Method). *Suppose that Condition S holds. In addition, suppose that* $h\sqrt{m} = o(n^{-\varepsilon})$, $m^2\zeta_m^2 = o(n^{1-\varepsilon}h)$, *and* $m^{-\kappa+1/2} = o(n^{-\varepsilon})$ *for some constant* $\varepsilon > 0$. *Then*

$$\widehat{J}^{-1}(u)\mathbb{U}^*(u) = J^{-1}(u)\mathbb{U}^*(u) + r(u), \quad u \in \mathcal{U},$$

where

$$\sup_{u \in \mathcal{U}} \|r(u)\| \lesssim_P \sqrt{\frac{\zeta_m^2 m^2 \log n}{nh}} + m^{-\kappa+1/2} + h\sqrt{m} = o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$. *The stated bound continues to hold in P-probability if we replace the unconditional probability* P *by the conditional probability* P^*.

This theorem is the fourth main result together with Theorems 4, 6, and 7 below on gradient bootstrap, Gaussian, and weighted bootstrap methods. The pivotal method is closely related to another approach to inference, which we refer to as the gradient bootstrap method. This approach was previously introduced by Parzen, Wei and Ying [68] for parametric models with fixed dimension. We extend it to the considerably more general series framework studied in this paper. The main idea is to generate for all $u \in \mathcal{U}$ the gradient bootstrap estimator $\widehat{\beta}^*(u)$ as the solution to the perturbed QR problem

$$\min_{\beta \in \mathbb{R}^m} \left(\mathbb{E}_n[\rho_u(Y_i - Z_i'\beta)] - \mathbb{U}^*(u)'\beta/\sqrt{n} \right), \tag{3.19}$$

where $\mathbb{U}^*(u)$ is defined in (3.18). Then, as shown in the next theorem, one can use the conditional distribution of the process $\sqrt{n}(\widehat{\beta}^*(\cdot) - \widehat{\beta}(\cdot)) = \{\sqrt{n}(\widehat{\beta}^*(u) - \widehat{\beta}(u)) : u \in \mathcal{U}\}$

given the data, which can be simulated, to approximate the distribution of the process $J^{-1}(\cdot)\mathbb{U}^*(\cdot)$, and, via Theorem 2, also of the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$.

Theorem 4 (Gradient Bootstrap Method). *Suppose that Condition S holds. In addition, suppose that $m^3\zeta_m^2 = o(n^{1-\varepsilon})$ and $m^{-\kappa+1/2} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$. Then*

$$\sqrt{n}\left(\widehat{\beta}^*(u) - \widehat{\beta}(u)\right) = J^{-1}(u)\mathbb{U}^*(u) + r(u),$$

where

$$\sup_{u \in \mathcal{U}} \|r(u)\| \lesssim_P \frac{m^{3/4}\zeta_m^{1/2}\log^{1/2}n}{n^{1/4}} + m^{-\kappa+1/2} = o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$. The stated bound continues to hold in P-probability if we replace the unconditional probability P by the conditional probability P^*.

Comment 5 (Comparison of pivotal and gradient bootstrap methods). Both the pivotal and gradient bootstrap methods have their own advantages. Perhaps the main advantage of the gradient bootstrap method relative to the pivotal method is that it does not require estimating the matrices $J(u)$, $u \in \mathcal{U}$, which is important because estimating these matrices requires a potentially subjective choice of the bandwidth h. In fact, implementing the gradient bootstrap method does not require any choice of smoothing parameters, making it particularly convenient for empirical researchers. On the other hand, an advantage of the pivotal method relative to the gradient bootstrap method is that it is computationally simple as it does not require solving the quantile optimization problem for each simulation of the process $\mathbb{U}^*(\cdot)$. ☐

Next, we turn to a strong approximation based on a sequence of Gaussian processes. The following theorem shows that for each n, one can construct a Gaussian process $G(\cdot) = G_n(\cdot) = \{G_n(u)\colon u \in \mathcal{U}\}$ such that the process $J^{-1}(\cdot)G(\cdot) = \{J^{-1}(u)G(u)\colon u \in \mathcal{U}\}$ is with high probability uniformly close to the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$.

Theorem 5 (Gaussian Coupling). *Suppose that Condition S holds. In addition, suppose that $m^7\zeta_m^6 = o(n^{1-\varepsilon})$ and $m^{-\kappa+1} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$. Then*

$$\sqrt{n}\left(\widehat{\beta}(u) - \beta(u)\right) = J^{-1}(u)G(u) + r(u), \quad u \in \mathcal{U},$$

where $G(\cdot) = G_n(\cdot)$ is a process on $\mathcal{U}$ that, conditionally on $(Z_i)_{i=1}^n$, is zero-mean Gaussian with a.s. continuous sample paths and the covariance function

$$E\left[G(u_1)G(u_2)' \mid (Z_i)_{i=1}^n\right] = \mathbb{E}_n[Z_iZ_i'](u_1 \wedge u_2 - u_1u_2), \quad \text{for all } u_1 \text{ and } u_2 \text{ in } \mathcal{U}, \quad (3.20)$$

and

$$\sup_{u \in \mathcal{U}} \|r(u)\| = o_P(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$.

Comment 6 (Conditions of Theorem 5). Note that the strong approximation to the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ by the Gaussian process $J^{-1}(\cdot)G(\cdot)$ constructed in Theorem 5 requires the condition that $m^7 \zeta_m^6 = o(n^{1-\varepsilon})$, which is more restrictive than the corresponding condition in Theorem 2, $m^3 \zeta_m^2 = o(n^{1-\varepsilon})$, required for the strong approximation to the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ by the pivotal process $J^{-1}(\cdot)\mathbb{U}(\cdot)$. We note that this restrictive condition is sufficient but we do not know whether it is necessary. This condition is a consequence of a step in the proof of Theorem 5 that relies upon Yurinskii's coupling. Therefore, improving that step through the use of another coupling could potentially lead to significant improvements in the conditions of the theorem; see, in particular, Theorem 12 in the next section. $\square$

Although Theorem 5 requires strong conditions, it is important because it suggests that one can approximate the distribution of the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ using Gaussian and weighted bootstrap methods, which are wide-spread in the literature in other contexts and which we now describe.

Let us start with the Gaussian method. Let $\widehat{\Sigma}^{1/2}$ denote the square root of the matrix $\widehat{\Sigma}$. Note that the covariance function of the process $G(\cdot)$ conditional on $(Z_i)_{i=1}^n$, given in (3.20), is equal to that of the process $\widehat{\Sigma}^{1/2}B_m(\cdot)$, where $B_m(\cdot) = \{B_m(u) : u \in \mathcal{U}\}$ is a standard m-dimensional Brownian bridge, that is, a vector consisting of m independent scalar Brownian bridges. Since the sample path of the Brownian bridge is continuous a.s., it follows that the process $\widehat{\Sigma}^{1/2}B_m(\cdot)$ is a copy of the process $G(\cdot)$, conditional on $(Z_i)_{i=1}^n$. Hence, one can simulate a standard m-dimensional Brownian bridge $B_m^*(\cdot) = \{B_m^*(u) : u \in \mathcal{U}\}$ that is independent of the data and define

$$G^*(u) = G_n^*(u) = \widehat{\Sigma}^{1/2}B_m^*(u), \quad u \in \mathcal{U}, \tag{3.21}$$

so that conditional on $(Z_i)_{i=1}^n$, the process $G^*(\cdot) = \{G^*(u) : u \in \mathcal{U}\}$ is a copy of the process $G(\cdot)$, and $J^{-1}(\cdot)G^*(\cdot) = \{J^{-1}(u)G^*(u) : u \in \mathcal{U}\}$ is a copy of $J^{-1}(\cdot)G(\cdot)$. Let $\widehat{J}(u)$ be the estimators of the matrices $J(u)$ for all $u \in \mathcal{U}$ in (2.11). Then, as shown in the next theorem, one can use the conditional distribution of the process $\widehat{J}^{-1}(\cdot)G^*(\cdot) = \{\widehat{J}^{-1}(u)G^*(u) : u \in \mathcal{U}\}$ given the data, which can be simulated, to approximate the distribution of the process $J^{-1}(\cdot)G^*(\cdot)$, and via Theorem 5 also of the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$.

Theorem 6 (Gaussian Method). *Suppose that Condition S holds. In addition, suppose that* $h\sqrt{m} = o(n^{-\varepsilon})$, $m^2\zeta_m^2 = o(n^{1-\varepsilon}h)$, *and* $m^{-\kappa+1/2} = o(n^{-\varepsilon})$ *for some constant* $\varepsilon > 0$. *Then*

$$\widehat{J}^{-1}(u)G^*(u) = J^{-1}(u)G^*(u) + r(u), \quad u \in \mathcal{U},$$

where

$$\sup_{u \in \mathcal{U}} \|r(u)\| \lesssim_P \sqrt{\frac{m^2\zeta_m^2 \log n}{nh}} + m^{-\kappa+1/2} + h\sqrt{m} = o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$. *The stated bound continues to hold in P-probability if we replace the unconditional probability P by the conditional probability* P^*.

Another related inference method is the weighted bootstrap method. Præstgaard and Wellner [72], Hahn [41], Chamberlain and Imbens [17], and Chen and Pouzo [23] previously used this method in the point-wise case, where the set $\mathcal{U}$ is a singleton. We extend this method to obtain the distributional approximation for the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot)) = \{\sqrt{n}(\widehat{\beta}(u) - \beta(u)): u \in \mathcal{U}\}$ where $\mathcal{U}$ is not a singleton and in fact can be a continuum of quantile indices. To describe the method, consider a set of weights $\pi_1, ..., \pi_n$ that are i.i.d. draws from the distribution of a non-negative random variable π with $E[\pi] = 1$ and $E[\pi^2] = 2$, such as the standard exponential distribution, and that are independent of the data. For all $u \in \mathcal{U}$, define the weighted bootstrap estimator $\widehat{\beta}^b(u)$ as the solution to the weighted QR problem

$$\widehat{\beta}^b(u) \in \arg \min_{\beta \in \mathbb{R}^m} \mathbb{E}_n[\pi_i \rho_u(Y_i - Z_i'\beta)]. \tag{3.22}$$

Then, as shown in the next theorem, one can use the conditional distribution of the process $\sqrt{n}(\widehat{\beta}^b(\cdot) - \widehat{\beta}(\cdot)) = \{\sqrt{n}(\widehat{\beta}^b(u) - \widehat{\beta}(u)): u \in \mathcal{U}\}$ given the data, which can be simulated, to approximate the distribution of the process $J^{-1}(\cdot)G^*(\cdot)$, and via Theorem 5 also of the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$.

Theorem 7 (Weighted Bootstrap Method). *Suppose that Condition S holds. In addition, suppose that* $m^7\zeta_m^6 = o(n^{1-\varepsilon})$ *and* $m^{-\kappa+1} = o(n^{-\varepsilon})$ *for some constant* $\varepsilon > 0$. *Moreover, suppose that the random variable* π *is non-negative and satisfies* $E[\pi] = 1$, $E[\pi^2] = 2$, $E[\pi^4] \lesssim 1$. *Finally, suppose that* $\max_{1 \leq i \leq n} \pi_i \lesssim \log n$. *Then*

$$\sqrt{n}\left(\widehat{\beta}^b(u) - \widehat{\beta}(u)\right) = J^{-1}(u)G^*(u) + r(u), \tag{3.23}$$

where $G^*(\cdot) = G_n^*(\cdot)$ *is a process on* $\mathcal{U}$ *that, conditionally on* $(Z_i)_{i=1}^n$, *is zero-mean Gaussian with a.s. continuous sample paths and the covariance function* (3.20), *and*

$$\sup_{u \in \mathcal{U}} \|r(u)\| \lesssim_P o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$. Moreover, the stated bound continues to hold in P-probability if we replace the unconditional probability P by the conditional probability P^*.

Comment 7 (Comparison of Gaussian and weighted bootstrap methods)**.** The comparison of the Gaussian and weighted bootstrap methods is similar to that of the pivotal and gradient bootstrap methods. Again both methods have their own advantages. The main advantage of the weighted bootstrap method is arguably that it does not require estimating the matrices $J(u)$, $u \in \mathcal{U}$, which allows us to bypass the need to select a bandwidth h. An advantage of the Gaussian method is that it is computationally simple as it does not require solving the quantile optimization problem for each simulation of weights $(\pi_i)_{i=1}^n$. □

Comment 8 (Comparison of resampling methods based on the pivotal and Gaussian couplings)**.** Although it is difficult to compare the resampling methods based on the pivotal coupling (pivotal and gradient bootstrap methods) with those based on the Gaussian coupling (Gaussian and weighted bootstrap methods) from a theoretical point of view, our results suggest that the former methods might be more accurate than the latter ones. Indeed, the methods based on the pivotal coupling require weaker conditions (see, however, Theorem 12 in the next section, where it is possible to substantially weaken conditions required for the Gaussian coupling in some examples) and, in addition, developing the Gaussian coupling requires a "double approximation": in order to construct a Gaussian process that strongly approximates the original process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$, we first construct a coupling of the latter process with the pivotal process, and then we construct a coupling of the Gaussian process with the pivotal process, so that the Gaussian process is coupled with the original process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ via the pivotal process. In the numerical example of Section 5.2, however, we find that the performance of the four methods is similar in small samples.

In addition, the Gaussian coupling is important because of the existence of well-developed extreme value theory for Gaussian processes; see, for example, Leadbetter, Lindgren and Rootzen [58]. In combination with the Gaussian coupling, this theory can be used to develop inferential procedures for some linear functionals without relying upon resampling methods (that is, with non-bootstrap critical values) like in Rio [74]. Moreover, the Gaussian coupling is important because of existence of anti-concentration inequalities for Gaussian processes (Lemma 12), which are useful to construct uniform confidence bands for linear functionals in the next section. □

4. Estimation and Inference Methods for Linear Functionals

Here we build upon the results from the previous section to develop estimation and inference methods for the ultimate parameters of interest: linear functionals of the conditional quantile function. We consider both pointwise and uniform methods.

4.1. **Pointwise Rate of Convergence and Inference.** As explained in Section 2, we construct an estimator of the parameters $\theta(u, w)$ for $(u, w) \in I$ by applying the plug-in rule to its QR-series approximation, that is,

$$\widehat{\theta}(u, w) = \ell(w)'\widehat{\beta}(u), \quad (u, w) \in I,$$

where $\widehat{\beta}(u)$ is the QR-series coefficient estimator that solves the optimization problem (2.4) and $\ell(w)$ is the vector of loadings. In cases where $\theta(u, w)$ is known to be monotone with respect to either w or u, we show in Appendix B of the Supplemental Material how to impose this restriction after estimation to improve finite sample properties of $\widehat{\theta}(u, w)$.

We start with the rate of convergence of the estimator $\widehat{\theta}(u, w)$ at a particular quantile index value u and a particular covariate value w (pointwise rate of convergence). In principle, the point (u, w) can depend on n, but we suppress the dependence for simplicity of notation. We use the following assumption:

Condition P. *The QR-series decomposition* $\theta(u, w) = \ell(w)'\beta(u) + r(u, w)$ *satisfies*

$$\frac{\sqrt{n}|r(u, w)|}{\|\ell(w)\|} = o(1).$$

Condition P can be understood as an undersmoothing condition. Although undersmoothing conditions are widely spread in the literature, as Belloni et al [9] pointed out, there is no theoretically justified procedure in the literature that would lead to a desired level of undersmoothing for the estimators of the linear functionals even for least squares estimators.

Based on Condition P, we derive the following theorem for the pointwise rate of convergence of $\widehat{\theta}(u, w)$, which is the fifth main result together with Theorem 9 below on pointwise asymptotic normality of $\widehat{\theta}(u, w)$.

Theorem 8 (Pointwise Convergence Rate for Linear Functionals)**.** *Suppose that the conditions of Theorem 2 hold. In addition, suppose that Condition P holds. Then*

$$|\widehat{\theta}(u, w) - \theta(u, w)| \lesssim_P \frac{\|\ell(w)\|}{\sqrt{n}}.$$

Comment 9 (Rates and norm of vector of loadings)**.** The rate of convergence of $\widehat{\theta}(u, w)$ depends on the functional $\theta(u, w)$ through the norm of the vector of loadings $\ell(w)$. For example, if we are interested in the coefficient $\beta_1(u)$, so that $\theta(u, w) = \beta_1(u)$, which might be a parameter of interest in the Many regressors (MR) model, then $\ell(w) = (1, 0, \ldots, 0)'$, and so $\|\ell(w)\| = 1$, yielding a $\sqrt{n}$-consistent estimator $\widehat{\theta}(u, w)$. See Comment 11 below for additional examples of linear functionals with bounds on $\|\ell(w)\|$. $\qquad\qquad\square$

In order to perform inference, we consider the t-statistic

$$t(u, w) = \frac{\widehat{\theta}(u, w) - \theta(u, w)}{\widehat{\sigma}(u, w)},$$

where

$$\widehat{\sigma}^2(u, w) = u(1 - u)\ell(w)'\widehat{J}^{-1}(u)\widehat{\Sigma}\widehat{J}^{-1}(u)\ell(w)/n \qquad\qquad (4.24)$$

is a consistent estimator of

$$\sigma^2(u, w) = u(1 - u)\ell(w)'J^{-1}(u)\Sigma J^{-1}(u)\ell(w)/n, \qquad\qquad (4.25)$$

the asymptotic variance of $\widehat{\theta}(u, w)$, obtained by the delta method. We can carry out standard inference based on this t-statistic because $t(u, w) \rightarrow_d N(0, 1)$, as we establish below.

Theorem 9 (Pointwise Inference for Linear Functionals)**.** *Suppose that the conditions of Theorem 2 hold. In addition, suppose that Condition P holds, $h = o(1)$ and $m\zeta_m^2 \log^2 n = o(nh)$. Then*

$$t(u, w) \rightarrow_d N(0, 1).$$

Comment 10 (Using resampling methods for pointwise inference)**.** Although it is possible to establish validity of all the resampling methods from the previous section to perform pointwise inference on linear functionals, we do not show these results here because they will follow as a special case from our results below on uniform inference for linear functionals. $\quad\square$

4.2. **Uniform Rate of Convergence and Inference.** Next, we derive the rate of convergence of the estimator $\widehat{\theta}(u, w)$ that holds uniformly over $(u, w) \in I$. We use the following assumption:

Condition U.

 U.1 *The set I is such that its dimension d_I is fixed and its diameter is bounded uniformly over n.*

U.2 *For some $\varepsilon > 0$, the QR-approximation error $r(u, w)$ satisfies*

$$\sqrt{n} \sup_{(u,w) \in I} \frac{|r(u, w)|}{\|\ell(w)\|} = o(n^{-\varepsilon}).$$

U.3 *The vector of loadings $\ell(w)$ satisfies*

$$\|\ell(w)\| \leq \zeta_{m,\theta} \ and \ \left\| \frac{\ell(w)}{\|\ell(w)\|} - \frac{\ell(w')}{\|\ell(w')\|} \right\| \leq \zeta_{m,\theta}^L \|w - w'\|$$

for all $w, w' \in \mathcal{W}$, where $\log \zeta_{m,\theta}^L \lesssim \log n$.

Condition U.1 on the dimension and the diameter of the set I is mild and can be further relaxed at the expense of additional technicalities. As in the pointwise case, Condition U.2 can be understood as an undersmoothing condition. Condition U.3 requires that the vector of loadings $\ell(w)$ is bounded uniformly over $w \in \mathcal{W}$ in the Euclidean norm by $\zeta_{m,\theta}$ and the function $w \mapsto \ell(w)/\|\ell(w)\|$ is Lipschitz-continuous in the Euclidean norm with the Lipschitz constant $\zeta_{m,\theta}^L$. We note that the last condition is rather weak because the only requirement on the Lipschitz constant that we impose is that $\log \zeta_{m,\theta}^L \lesssim \log n$. We discuss some bounds on the constant $\zeta_{m,\theta}$ in a separate comment below.

Comment 11 (Primitive bounds on $\zeta_{m,\theta}$)**.** The uniform rate of convergence for the estimator $\widehat{\theta}(u, w)$ derived below in Theorem 10 will crucially depend on the constant $\zeta_{m,\theta}$ appearing in Condition U. Here we discuss some bounds on this constant. For brevity, we only discuss the case of B-splines and refer to Newey [67] and Chen [21] for other choices of approximating functions. We assume that $\mathcal{X} = [0, 1]^d$ and that the vector of approximating functions Z consists of tensor products of B-splines of order s_0. As discussed above, then $\zeta_m = \sup_{x \in \mathcal{X}} \|Z(x)\| \lesssim \sqrt{m}$ and it is also possible to verify that for all positive integers $\alpha \leq s_0$, $\sup_{x \in \mathcal{X}} \|\partial_{x_k}^\alpha Z(x)\| \lesssim m^{1/2 + \alpha/d}$; see for example Chen and Christensen [22]. Then

- For $\theta(u, w) = Q(u, x)$, $\ell(w) = Z(x)$ and $\zeta_{m,\theta} \lesssim m^{1/2}$;
- For $\theta(u, w) = \partial_{x_k} Q(u, x)$, $\ell(w) = \partial_{x_k}^\alpha Z(x)$ and $\zeta_{m,\theta} \lesssim m^{1/2 + \alpha/d}$;
- For $\theta(u) = \int \partial_{x_k} Q(u, x) d\mu(x)$ with $\text{supp}(\mu) \subset \text{int}(\mathcal{X})$ and $|\partial_{x_k} \mu(x)| \lesssim 1$, $\ell = \int \partial_{x_k} Z(x) \mu(x) \ dx = -\int Z(x) \partial_{x_k} \mu(x) \ dx$ and $\zeta_{m,\theta} \lesssim 1$;

see Newey [67] for more explanations on the last bound. $\qquad\square$

The following theorem establishes the uniform rate of convergence of the QR-series estimator $\widehat{\theta}(u, w)$, which is the sixth main result.

Theorem 10 (Uniform Convergence Rate for Linear Functionals). *Suppose that the conditions of Theorem 2 hold. In addition, suppose that Condition U hold. Then*

$$\sup_{(u,w)\in I} |\widehat{\theta}(u,w) - \theta(u,w)| \lesssim_P \sqrt{\frac{\zeta_{m,\theta}^2 \log n}{n}}.$$

The uniform rate of Theorem 10 is the same as the pointwise rate of Theorem 8 up to a small logarithmic factor. As in the pointwise rate result, the norm of the vector of loadings play a role which is controlled by $\zeta_{m,\theta}$ in Condition U.

Next, we consider inference on the function $(u,w) \mapsto \theta(u,w)$. We base inference on the t-statistic process $t(\cdot,\cdot) = \{t(u,w)\colon (u,w) \in I\}$ defined as follows:

$$t(u,w) = \frac{\widehat{\theta}(u,w) - \theta(u,w)}{\widehat{\sigma}(u,w)}, \tag{4.26}$$

where $\widehat{\sigma}^2(u,w)$, defined in (4.24), is an estimator of the asymptotic variance $\sigma^2(u,w)$ of $\widehat{\theta}(u,w)$ in (4.25). Using the results in the previous section, we construct pivotal and Gaussian couplings for this process in the following theorem, which is the seventh main result together with Theorems 12, 13, and 14 below on couplings and resampling methods for the t-statistic process.

Theorem 11 (Pivotal and Gaussian Couplings for t-statistic Process). *Suppose that Conditions S and U hold. If $h = o(n^{-\varepsilon})$, $m\zeta_m^2 = o(n^{1-\varepsilon}h)$, $m^3\zeta_m^2 = o(n^{1-\varepsilon})$, and $m^{-\kappa+1} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$, then*

$$\sup_{(u,w)\in I} \left| t(u,w) - \frac{\ell(w)'J^{-1}(u)\mathbb{U}(u)/\sqrt{n}}{\sigma(u,w)} \right| \lesssim_P o(n^{-\varepsilon'}) \tag{4.27}$$

for the process $\mathbb{U}(\cdot)$ defined in (3.16) for some $\varepsilon' > 0$. Also, if $h = o(n^{-\varepsilon})$, $m\zeta_m^2 = o(n^{1-\varepsilon}h)$, $m^7\zeta_m^6 = o(n^{1-\varepsilon})$, and $m^{-\kappa+1} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$, then

$$\sup_{(u,w)\in I} \left| t(u,w) - \frac{\ell(w)'J^{-1}(u)G(u)/\sqrt{n}}{\sigma(u,w)} \right| \lesssim_P o(n^{-\varepsilon'}) \tag{4.28}$$

for the process $G(\cdot)$ defined in Theorem 5 for some $\varepsilon' > 0$.

The Gaussian coupling is derived in this theorem under rather strong condition $m^7\zeta_m^6 = o(n^{1-\varepsilon})$. It turns out that it is possible to construct the same coupling under a different set of conditions:

Theorem 12 (Gaussian Coupling for t-statistic Process under Alternative Conditions). *Suppose that Conditions S and U hold. In addition, suppose that $h = o(n^{-\varepsilon})$, $m\zeta_m^2 =$*

$o(n^{1-\varepsilon}h)$, $m^3\zeta_m^2 = o(n^{1-\varepsilon})$, and $m^{-\kappa+1} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$. Moreover, suppose that $(1+\zeta_{m,\theta}^L)^{2d_I}\zeta_m^2 = o(n^{1-\varepsilon})$. Then (4.28) holds for the same process $G(\cdot)$ as that used in Theorem 11.

Comment 12 (Comparison of conditions for the Gaussian coupling in Theorems 11 and 12)**.** The conditions of Theorems 11 and 12 required for the Gaussian coupling are non-nested. In particular, Theorem 12 requires the condition $m^3\zeta_m^2 = o(n^{1-\varepsilon})$ that is weaker than the corresponding condition in Theorem 11, $m^7\zeta_m^6 = o(n^{1-\varepsilon})$, but it also requires the condition $(1+\zeta_{m,\theta}^L)^{2d_I}\zeta_m^2 = o(n^{1-\varepsilon})$ that is stronger than the corresponding condition in Theorem 11, $\log \zeta_{m,\theta}^L \lesssim \log n$. However, in most cases of practical importance, the conditions of Theorem 12 are substantially weaker than those of Theorem 11. For example, consider the NP model and suppose that we are interested in the conditional quantile function $Q(u,x)$ itself, so that $\ell(\omega) = Z(x)$. Further, suppose that $\mathcal{X} = [0,1]^d$ and that the vector of approximating functions z consists of tensor products of B-splines. Then $\zeta_m \lesssim \sqrt{m}$, as discussed above, and it is also possible to show that $\zeta_{m,\theta}^L \lesssim m^{1/d}$. Hence, in this case Theorem 12 requires that $m^{4\vee(2/d+3)} = o(n^{1-\varepsilon})$ since $d_I = 1 + d$ whereas Theorem 11 requires $m^{10} = o(n^{1-\varepsilon})$.

As in the previous section, we can use four resampling methods to approximately simulate the distribution of the pivotal and Gaussian processes. Specifically, define the processess $\mathbb{U}^*(\cdot)$ and $G^*(\cdot)$ as in (3.18) and (3.21), respectively. Recall that conditional on $(Z_i)_{i=1}^n$, these processes are copies of the processes $\mathbb{U}(\cdot)$ and $G(\cdot)$, respectively, and so the processes

$$\left\{ \frac{\ell(w)'J^{-1}(u)\mathbb{U}^*(u)/\sqrt{n}}{\sigma(u,w)} : (u,w) \in I \right\} \quad \text{and} \quad \left\{ \frac{\ell(w)'J^{-1}(u)\mathbb{G}^*(u)/\sqrt{n}}{\sigma(u,w)} : (u,w) \in I \right\}$$

are copies of the the pivotal and Gaussian processes

$$\left\{ \frac{\ell(w)'J^{-1}(u)\mathbb{U}(u)/\sqrt{n}}{\sigma(u,w)} : (u,w) \in I \right\} \quad \text{and} \quad \left\{ \frac{\ell(w)'J^{-1}(u)\mathbb{G}(u)/\sqrt{n}}{\sigma(u,w)} : (u,w) \in I \right\},$$

respectively. Also, define the t-statistic bootstrap process $t^*(\cdot, \cdot) = \{t^*(u, w): (u, w) \in I\}$ for each method as

$$\text{pivotal method:} \qquad t^*(u, w) = \frac{\ell(w)'\widehat{J}^{-1}(u)\mathbb{U}^*(u)/\sqrt{n}}{\widehat{\sigma}(u, w)};$$

$$\text{gradient bootstrap method:} \quad t^*(u, w) = \frac{\ell(w)'(\widehat{\beta}^*(u) - \widehat{\beta}(u))}{\widehat{\sigma}(u, w)};$$

$$\text{Gaussian method:} \qquad t^*(u, w) = \frac{\ell(w)'\widehat{J}^{-1}(u)G^*(u)/\sqrt{n}}{\widehat{\sigma}(u, w)};$$

$$\text{weighted bootstrap method:} \quad t^*(u, w) = \frac{\ell(w)'(\widehat{\beta}^b(u) - \widehat{\beta}(u))}{\widehat{\sigma}(u, w)}.$$

The following theorem shows that the conditional distribution of the t-statistic bootstrap process $t^*(\cdot, \cdot)$ given the data, which can be simulated, approximates the distribution of the pivotal (in the case of pivotal and gradient bootstrap methods) and Gaussian (in the case of Gaussian and weighted bootstrap methods) processes, and via Theorems 11 and 12 also of the original t-statistic process $t(\cdot, \cdot)$.

Theorem 13 (Validity of Resampling Methods for t-statistic Process). *Suppose that Conditions S and U hold. In addition, suppose that $h = o(n^{-\varepsilon})$ and $m\zeta_m^2 = o(n^{1-\varepsilon}h)$ for some constant $\varepsilon > 0$. Moreover, suppose that (i) the conditions of Theorem 3 hold in the case of the pivotal method, (ii) the conditions of Theorem 4 hold in the case of gradient bootstrap method, (iii) the conditions of Theorem 6 hold in the case of Gaussian method, and (iv) the conditions of Theorems 7 hold in the case of weighted bootstrap method. Then for the pivotal and gradient bootstrap methods,*

$$\sup_{(u,w)\in I} \left| t^*(u, w) - \frac{\ell(w)'J^{-1}(u)\mathbb{U}^*(u)/\sqrt{n}}{\sigma(u, w)} \right| \lesssim_P o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$. In addition, for the Gaussian and weighted bootstrap methods,

$$\sup_{(u,w)\in I} \left| t^*(u, w) - \frac{\ell(w)'J^{-1}(u)G^*(u)/\sqrt{n}}{\sigma(u, w)} \right| \lesssim_P o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$. Moreover, the stated bounds continue to hold in P-probability if we replace the unconditional probability P by the conditional probability P^*.

Note that in the case of weighted bootstrap method, the theorem above imposes the rather strong condition $m^7 \zeta_m^6 = o(n^{1-\varepsilon})$. Like in the case of Theorem 12, it turns out that it is possible to obtain the same approximation as in this theorem but under a different set of conditions:

Theorem 14 (Weighted Bootstrap Method for t-statistic Process under Alternative Conditions). *Suppose that Conditions S and U hold. In addition, suppose that $h = o(n^{-\varepsilon})$, $m\zeta_m^2 = o(n^{1-\varepsilon}h)$, $m^3\zeta_m^2 = o(n^{1-\varepsilon})$, and $m^{-\kappa+1} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$. Moreover, suppose that $(1 + \zeta_{m,\theta}^L)^{2d_I}\zeta_m^2 = o(n^{1-\varepsilon})$. Finally, suppose that the conditions of Theorem 7 on the weights π_i hold. Then for the weighted bootstrap method,*

$$\sup_{(u,w)\in I} \left| t^*(u,w) - \frac{\ell(w)'J^{-1}(u)G^*(u)/\sqrt{n}}{\sigma(u,w)} \right| \lesssim_P o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$. Moreover, the stated bound continues to hold in P-probability if we replace the unconditional probability P by the conditional probability P^*.

With the help of Theorems 11 – 14, we can solve a wide range of inference problems. For example, we can construct uniform confidence bands for linear functionals $(u,w) \mapsto \theta(u,w)$ on I, and test shape constraints for the conditional quantile functions $x \mapsto Q(u,x)$. For the former problem, let

$$V = \sup_{(u,w)\in I} |t(u,w)|$$

be the maximal t-statistic. Also, let $k(1-\alpha)$ denote the $(1-\alpha)$ quantile of the distribution of V. If $k(1-\alpha)$ were known, we would have the confidence band

$$\left\{ [\widehat{\theta}(u,w) - k(1-\alpha)\widehat{\sigma}(u,w), \widehat{\theta}(u,w) + k(1-\alpha)\widehat{\sigma}(u,w)] \colon (u,w) \in I \right\} \tag{4.29}$$

covering the whole function $\{\theta(u,w)\colon (u,w) \in I\}$ with probability $1 - \alpha$ exactly. However, $k(1-\alpha)$ is typically unknown, and the confidence band (4.29) is infeasible. Instead, we approximate $k(1-\alpha)$ using the resampling methods developed in this paper. Specifically, let

$$V^* = \sup_{(u,w)\in I} |t^*(u,w)|$$

be the bootstrap maximal t-statistic, and let $k^*(1-\alpha)$ be the $(1-\alpha)$ quantile of the conditional distribution of V^* given the data. This quantity can be computed numerically by Monte Carlo methods, as we illustrate in the next section via empirical examples and give precise algorithms in Appendix A. We then form a two-sided $(1-\alpha)$ uniform confidence

band as

$$\left\{[i(u,w), \ddot{i}(u,w)] = [\widehat{\theta}(u,w) - k^*(1-\alpha)\widehat{\sigma}(u,w), \ \widehat{\theta}(u,w) + k^*(1-\alpha)\widehat{\sigma}(u,w)] : (u,w) \in I\right\}.$$

The following theorem establishes that this confidence band covers the whole function $\{\theta(u,w) : (u,w) \in I\}$ with probability $(1-\alpha)$ in large samples.

Theorem 15 (Uniform Confidence Bands). *Suppose that Conditions S and U hold. In addition, suppose that $m^3\zeta_m^2 = o(n^{1-\varepsilon})$ and $m^{-\kappa+1} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$. Moreover, suppose that (i) $h\sqrt{m} = o(n^{-\varepsilon})$ and $m^2\zeta_m^2 = o(n^{1-\varepsilon}h)$ in the case of pivotal and Gaussian methods and (ii) $h = o(n^{-\varepsilon})$ and $m\zeta_m^2 = o(n^{1-\varepsilon}h)$ in the case of gradient and weighted bootstrap methods. Finally, suppose that the conditions of Theorem 7 on the weights π_i hold in the case of the weighted bootstrap method. (1) Then*

$$P\left(V \le k^*(1-\alpha)\right) = 1 - \alpha + o(1). \tag{4.30}$$

(2) As a consequence,

$$P\left(\theta(u,w) \in [i(u,w), \ddot{i}(u,w)], \ for \ all \ (u,w) \in I\right) = 1 - \alpha + o(1). \tag{4.31}$$

(3) The width of the confidence band $2k^*(1-\alpha)\widehat{\sigma}(u,w)$ obeys

$$2k^*(1-\alpha)\widehat{\sigma}(u,w) \lesssim_P \sqrt{\frac{\zeta_{m,\theta}^2 \log n}{n}} \tag{4.32}$$

uniformly over $(u,w) \in I$.

In addition to the validity of the uniform confidence band, Theorem 15 establishes that the width of the uniform confidence band is of the same order as the uniform rate of convergence of the estimator $\widehat{\theta}(u,w)$.

Comment 13 (Related literature). The construction of uniform confidence bands for non-parametric functions has been of large interest both in econometrics and statistics at least from the seventies. Early constructions can be traced back at least to the classic work [11] by Bickel and Rosenblatt. More recent contributions include Claeskens and Keilegom [32], Horowitz and Lee [48], Giné and Nickl [38], and Chernozhukov, Chetverikov and Kato [25], among many others. Most of the constructions in the literature rely on a two-step strategy. First, the distribution of an estimator of the function of interest is approximated by some Gaussian process uniformly over its domain. Second, extreme value theory is employed to obtain the limit distribution of the supremum of the absolute value of the Gaussian process and its appropriate quantile is used to choose the width of the confidence band. A widely

understood problem of this construction, however, is that the limit distribution on the second step may not exist and even if it does, it is often difficult to derive its explicit form. This distribution depends both on the function of interest and on the estimator considered, so that treatment of any new estimation problem requires a separate theorem, and in fact considerable efforts have been devoted to derive this distribution even in relatively simple settings, like density estimation based on projection kernels; see Giné and Nickl [38] and references therein. We avoid this problem: instead of deriving the limit distribution on the second step, we rely upon resampling methods developed in this paper. As a result, our construction yields asymptotically exact uniform confidence bands that work generically for all linear functionals of the conditional quantile functions. Our strategy is related to that used in Chernozhukov, Chetverikov and Kato [25] for the problem of density estimation and is built on Chernozhukov, Lee and Rosen [30], who proposed a related strategy for inference on the minimum of a function. □

Finally, we consider the problem of testing shape constraints for the conditional quantile functions $x \mapsto Q(u, x)$. Let x_k denote the k-th component of x. We assume that the functions $x \mapsto Q(u, x)$ are twice continuously differentiable and consider three types of shape constraints:

(i) Monotonicity of $x \mapsto Q(u, x)$ with respect to x_k: $\partial_{x_k} Q(u, x) \leq 0$ for all $x \in \mathcal{X}$ and $u \in \mathcal{U}$;

(ii) Concavity of $x \mapsto Q(u, x)$ with respect to x_k: $\partial_{x_k}^2 Q(u, x) \leq 0$ for all $x \in \mathcal{X}$ and $u \in \mathcal{U}$;

(iii) Concavity of $x \mapsto Q(u, x)$ with respect to x: $\alpha' \partial_x^2 Q(u, x)\alpha \leq 0$ for all $\alpha \in S^{d-1}$, $x \in \mathcal{X}$, and $u \in \mathcal{U}$,

where in the third example $\partial_x^2 Q(u, x)$ denotes the $d \times d$-dimensional matrix whose (j, k)-th element is given by $\partial_{x_j} \partial_{x_k} Q(u, x)$ for all $j, k = 1, \ldots, d$.[6] Note that in the first example we focus on the case where $x \mapsto Q(u, x)$ is decreasing with respect to x_k but we can also consider the case where $x \mapsto Q(u, x)$ is increasing simply by replacing Y by $-Y$ and u by $1 - u$. Similarly, we can consider the case of convexity in the second and third examples.

---

[6]Note that the twice continuously differentiable function $f \colon \mathcal{X} \to \mathbb{R}$ is concave if and only if $\alpha' \partial_x^2 f(x)\alpha \leq 0$ for all $x \in \mathcal{X}$ and $\alpha \in S^{d-1}$. To prove this claim, note that f is concave if and only if the function $t \mapsto f(x+t\alpha)$ mapping $\{t \in \mathbb{R} \colon x + t\alpha \in \mathcal{X}\}$ to $\mathbb{R}$ is concave for all $x \in \mathcal{X}$ and $\alpha \in S^{d-1}$, which in turns holds if and only if $\alpha' \partial_x^2 f(x)\alpha \leq 0$ for all $x \in \mathcal{X}$ and $\alpha \in S^{d-1}$.

34

Now, observe that all three shape constraints discussed above can be expressed using the same notation:

$$\theta(u, w) \leq 0, \quad \text{for all } (u, w) \in I,$$

where $\theta(u, w)$ is a linear functional with the vector of loadings being $\ell(w) = \partial_{x_k} Z(x)$ with $w = x$ in the first example, $\ell(w) = \partial_{x_k}^2 Z(x)$ with $w = x$ in the second example, and $\ell(w) = (\ell_1(w), \ldots, \ell_m(w))'$ where $\ell_j(w) = \alpha' \partial_x^2 Z_j(x) \alpha$ for all $j = 1, \ldots, m$ with $w = (x, \alpha)$ in the third example. Hence, we are interested in testing

$$H_0: \sup_{(u,w) \in I} \theta(u, w) \leq 0 \quad \text{against} \quad H_1: \sup_{(u,w) \in I} \theta(u, w) > 0.$$

To test H_0 against H_1, we consider the one-sided Kolmogorov-Smirnov statistic

$$T = \sup_{(u,w) \in I} \frac{\widehat{\theta}(u, w)}{\widehat{\sigma}(u, w)}.$$

Then under H_0,

$$T = \sup_{(u,w) \in I} \frac{\widehat{\theta}(u, w)}{\widehat{\sigma}(u, w)} \leq \sup_{(u,w) \in I} \frac{\widehat{\theta}(u, w) - \theta(u, w)}{\widehat{\sigma}(u, w)} = \sup_{(u,w) \in I} t(u, w).$$

Note also that too large values of T suggest that H_0 is violated. Hence, letting $\tilde{k}(1 - \alpha)$ denote the $(1 - \alpha)$ quantile of $\sup_{(u,w) \in I} t(u, w)$, we would like to reject H_0 if $T > \tilde{k}(1 - \alpha)$. However, such a test is not feasible because $\tilde{k}(1 - \alpha)$ is unknown. Instead, we approximate $\tilde{k}(1 - \alpha)$ using the resampling methods developed in this paper. Specifically, let

$$T^* = \sup_{(u,w) \in I} t^*(u, w)$$

be the bootstrap statistic, and let $\tilde{k}^*(1 - \alpha)$ be the $(1 - \alpha)$ quantile of the conditional distribution of T^* given the data. This quantity can be computed numerically by Monte Carlo methods. Then we reject H_0 in favor of H_1 if $T > \tilde{k}^*(1 - \alpha)$. The following theorem shows that this test controls size in large samples.

Theorem 16 (Test of Shape Constraints)**.** *Suppose that the conditions of Theorem 15 hold. Then under H_0,*

$$P\left(T > \tilde{k}^*(1 - \alpha)\right) \leq \alpha + o(1).$$

Moreover, if $\mathcal{M} = \mathcal{M}_n$ is a set of data-generating processes that satisfy H_0 and is such that the conditions of Theorem 15 hold uniformly over this set, then

$$\sup_{M \in \mathcal{M}} P_M\left(T > \tilde{k}^*(1 - \alpha)\right) \leq \alpha + o(1),$$

where P_M denotes probability under the data-generating process M.

5. Examples

This section illustrates the finite sample performance of the estimation and inference methods with two examples. All the calculations were carried out with the software R ([75]), using the package `quantreg` for quantile regression (Koenker [53]).

5.1. **Empirical Example.** To illustrate our methods with real data, we consider an empirical application to nonparametric estimation of the demand for gasoline. Blundell, Horowitz and Parey [12], Hausman and Newey [44], Schmalensee and Stoker [76], and Yatchew and No [83] estimated nonparametrically the average demand function. We estimate nonparametrically the quantile demand and price elasticity functions and apply our inference methods to construct confidence bands for the average quantile price elasticity function and to test the Slutsky condition of consumer demand. We use the same data set as in Yatchew and No [83], which comes from the National Private Vehicle Use Survey, conducted by Statistics Canada between October 1994 and September 1996.[7] The main advantage of this data set, relative to similar data sets for the U.S., is that it is based on fuel purchase diaries and contains detailed household level information on prices, fuel consumption patterns, vehicles and demographic characteristics. (See Yatchew and No [83] for a more detailed description of the data.) Our sample selection and variable construction also follow Yatchew and No [83]. We select into the sample households with non-zero licensed drivers, vehicles, and distance driven. We focus on regular grade gasoline consumption. This selection results in a sample of 5,001 households. Fuel consumption and expenditure are recorded by the households at the purchase level.

We consider a partially linear specification for the demand function:[8]

$$Y = Q(U, X), \quad Q(U, X) = g(U, W) + V'\beta(U), \quad X = (W, V),$$

where Y is the log of total gasoline consumption in liters per month; W is the log of price in Canadian dollars per liter; U is the unobservable preference of the household to consume gasoline; and V is a vector of 28 covariates. Following Yatchew and No [83], the covariate vector includes the log of age, a dummy for the top coded value of age, the log

---

[7]The data set can be downloaded from Adonis Yatchew's web site at www.economics.utoronto.ca/yatchew/.

[8]This partially linear specification of the demand function arises from household preferences characterized by the indirect utility function $V(w, v, u) = v^{1-\beta(u)}/[1 - \beta(u)] - G(u, w)$, where w is real gasoline price, v is real income, and $g(u, w) = \partial_w G(u, w)$ (see Lewbel [60], Th. 1). We thank Arthur Lewbel for pointing this out.

of income, a set of dummies for household size, a dummy for urban dwellers, a dummy for young-single (age less than 36 and household size of one), the number of drivers, a dummy for more than 4 drivers, 5 province dummies, and 12 monthly dummies. To estimate the function $w \mapsto g(w, u)$ at each u, we consider three different vectors of series approximating functions $w \mapsto Z(w)$: linear, a power orthogonal polynomial of degree 6, and a cubic B-spline with 5 knots at the $\{0, 1/4, 1/2, 3/4, 1\}$ quantiles of the observed values of W. The series approximation to the function $(u, x) \mapsto Q(u, x)$ takes the following form:

$$Q(u, x) = Z(w)'\delta(u) + v'\gamma(u) = Z(x)'\beta(u), \quad Z(x) = (Z(w), v), \quad \beta(u) = (\delta(u), \gamma(u)).$$

The number of series terms in the power and B-spline specifications is selected by under-smoothing over the specifications chosen by applying cross validation to the corresponding least squares estimators.[9] In the next section, we analyze the size of the specification error of these series approximations in a numerical experiment calibrated to mimic this example.

The empirical results for the B-spline specification are reported in Figures 1 and 2.[10] The first two panels of fig. 1 plot the initial and monotonized estimates of the quantile demand surface for gasoline as a function of price and the quantile index, that is

$$(u, \exp(w)) \mapsto \theta(u, w) = \exp(g(w, u) + v'\beta(u)),$$

where the value of v is fixed at the sample median values of the ordinal variables and one for the dummies corresponding to the sample modal values of the rest of the variables.[11] The monotonized estimates are obtained using the average rearrangement over both the price

---

[9]There is potentially a large set of methods that can be used to choose the number of series terms (cross-validation, penalization, the method of Lepski, among others). Indeed, the problem of selecting the number of series terms is a special case of the problem of model selection, and there are several textbooks/monographs in the literature on model selection in abstract settings; for example, Massart [65] and Koltchinskii [56]. However, to the best of our knowledge, there are no papers in the literature that apply to the problem of selecting the number of series terms in the nonparametric quantile regression problem studied here. Hence, we have opted to use an ad hoc method that consists of performing cross-validation as if we were to estimate the conditional mean function $x \mapsto E[Y|X = x]$, which is estimated by the series least squares method. Under the implicit assumption that the smoothness of the functions $x \mapsto Q(u, x)$ is similar to that of the function $x \mapsto E[Y|X = x]$, such a cross-validation would yield the number of series terms that approximately equalize variance and bias terms in estimating the functions $x \mapsto Q(u, x)$. We then slightly increase the number of series terms so that the bias term is of smaller order relative to the variance term (that is, to achieve undersmoothing, as stated in Condition U.2), so that valid inference can be performed.

[10]The results for the linear and power specifications are not reported for the sake of brevity. They are similar to the results for the B-spline specification.

[11]The median values of the ordinal covariates are $\$40K$ for income, 46 for age, and 2 for the number of drivers. The modal values for the rest of the covariates are 0 for the top-coding of age, 2 for household size,

and quantile dimensions proposed in Chernozhukov, Fernández-Val, and Galichon [29]; see Appendix B. The demand surface show most noticeably non-monotone areas with respect to price at high quantiles, which are removed by the rearrangement. The last panel of fig. 1 shows the estimate of the quantile price elasticity surface as a function of price and the quantile index, that is:

$$(u, \exp(w)) \mapsto \theta(u, w) = \partial_w g(u, w).$$

The estimates show substantial heterogeneity of the elasticity across quantiles and prices, with individuals at the upper quantiles being less sensitive to high prices.[12]

Fig. 2 shows 90% uniform confidence bands for the average quantile price elasticity function

$$u \mapsto \theta(u) = \int \partial_w \ g(u, w) d\mu(w),$$

over the quantile indices $\mathcal{I} = [0.1, 0.9]$, where μ is the empirical distribution of W. The panels of the figure correspond to the pivotal, gradient bootstrap, Gaussian and weighted bootstrap methods. For the pivotal and Gaussian methods the distribution of the maximal t-statistic is obtained by 1,000 simulations. The gradient bootstrap uses 199 repetitions. The weighted bootstrap uses standard exponential weights and 199 repetitions. The confidence bands show that the evidence of heterogeneity in the elasticities across quantiles is not statistically significant, because we can trace a horizontal line within the bands. They also show that there is significant evidence of negative price sensitivity at most quantiles as the bands are bounded away from zero for most quantiles.

The Slutsky condition of consumer demand states that the compensated price elasticity is negative for all the households. Dette, Hoderlein, and Neumeyer [34] showed that this condition has testable implications for the quantile demand function and its derivatives in heterogeneous demand systems with multiple goods and possible infinite dimensional unobservables. The one good version of their test is:

$$H_0 : S(u, x) \leq 0, \text{ for all } (u, x) \in \mathcal{U} \times \mathcal{X}, \text{ vs } H_1 : S(u, x) > 0, \text{ for some } (u, x) \in \mathcal{U} \times \mathcal{X},$$

$$(5.33)$$

---

1 for urban dwellers, 0 for young-single, 0 for the dummy of more than 4 drivers, 4 (Prairie) for province, and 11 (November) for month.

[12]These estimates are smoothed by local weighted polynomial regression across the price dimension (Cleveland [33]), because the unsmoothed elasticity estimates display very erratic behavior.

where $S(u,x)$ is the compensated quantile price elasticity that in our logarithmic specification takes the form

$$S(u,x) = \exp(\ell)\partial_w Q(u,x) + \exp(Q(u,x) + w)\partial_\ell Q(u,x), \ x = (w, \ell, c),$$

which is a smooth function of the quantile demand and derivatives. Here we have partitioned the covariate vector X into (W, L, C), where W is log of price, L is the log of income, and C includes the rest of the covariates.

To test the functional hypothesis (5.33) we use the two sided Kolmogorov-Smirnov statistic:

$$K = \max_{(u,x)\in I} \frac{\widehat{S}(u,x)}{\widehat{\sigma}_S(u,x)},$$

where

$$\widehat{S}(u,x) = \exp(\ell)\partial_w \widehat{Q}(u,x) + \exp(\widehat{Q}(u,x) + w)\partial_\ell \widehat{Q}(u,x),$$

is the plug-in series estimator of $S(u,x)$, $\widehat{Q}(u,x)$ is the series estimator of $Q(u,x)$, $\widehat{\sigma}_S(u,x)$ is a delta method estimator of the asymptotic standard deviation of $\widehat{S}(u,x)$, and $I \subseteq \mathcal{U} \times \mathcal{X}$ denotes the set of values of interest. We reject H_0 if the p-value of K under H_0 is less than α, i.e. $\sup_{P\in H_0} P(K > k) < \alpha$ where k is the realized value of K. Dette, Hoderlein, and Neumeyer [34] proposed an alternative test based on kernel estimators of the quantile function and its derivatives.

We estimate the distribution of K under H_0 by weighted bootstrap with moment selection to reduce the asymptotic non-similarity on the boundary of composite one sided functional tests (Linton, Song, and Whang [61]). The weighted bootstrap version of K is

$$K^*(c_n) = \max_{(u,x)\in I} \frac{\widehat{S}^*(u,x) - \widehat{S}(u,x)}{\widehat{\sigma}_S(u,x)} 1[|\widehat{S}(u,x)| < c_n \widehat{\sigma}_S(u,x)],$$

where

$$\widehat{S}^*(u,x) = \exp(\ell)\partial_w \widehat{Q}^*(u,x) + \exp(\widehat{Q}^*(u,x) + w)\partial_\ell \widehat{Q}^*(u,x),$$

is the bootstrap version of $\widehat{S}(u,x)$, $\widehat{Q}^*(u,x)$ is the series estimator of $Q(u,x)$ in the weighted sample, $1[|\widehat{S}(u,x)| < c_n \widehat{\sigma}_S(u,x)]$ is the moment selector (Chernozhukov, Hong, and Tamer [31], and Andrews and Soares [3]), and c_n is a sequence of thresholds that can grow with n. The centering by $\widehat{S}(u,x)$ imposes the least favorable null hypothesis $S(u,x) = 0$ at all the points $(u,x) \in I$ in the bootstrap to control the size of the test, whereas the moment selector discards points that are far from this hypothesis with very high probability to increase power. We consider three sequences for c_n: no moment selection with $c_n = 0$, BIC moment selection with $c_n^2 = \log n$, and LIL selection with $c_n^2 = 2\log\log n$. The estimator of

the p-value for a realization of the statistic k is the probability that $K^*(c_n)$ is greater than k conditional on the data.

Table 1 presents the results of the test of the Slutsky condition in our data set. We set the region I to the product of $\{0.01, 0.02, ..., 0.99\}$ and the observed support of X in the data. We obtain the p-values by weighted bootstraps with standard exponential weights and 199 replications. Here, we do not find sufficient evidence to reject the Slutsky condition at standard significance levels in any of the specifications with or without the moment selection.

TABLE 1. Test of Slutsky Condition

Specification	K stat	P-value†		
		No selection	BIC selection	LIL selection
Linear	0.47	0.95	0.76	0.58
Power	3.63	0.30	0.30	0.28
B-spline	2.30	0.96	0.96	0.96

†P-values obtained by weighted bootstrap with standard exponential weights and 199 replications.

5.2. **Numerical Example.** To evaluate the performance of our estimation and inference methods in finite samples, we conduct a Monte Carlo experiment designed to mimic the previous empirical example. We consider the following design for the data generating process:

$$Y = g(W) + V'\beta + \sigma\Phi^{-1}(U), \tag{5.34}$$

where $g(w) = \alpha_0 + \alpha_1 w + \alpha_2 \sin(2\pi w) + \alpha_3 \cos(2\pi w) + \alpha_4 \sin(4\pi w) + \alpha_5 \cos(4\pi w)$, V is the same covariate vector as in the empirical example, $U \sim U(0,1)$, and Φ^{-1} denotes the inverse of the CDF of the standard normal distribution. The parameters of $g(w)$ and β are calibrated by applying least squares to the data set in the empirical example and σ is calibrated to the least squares residual standard deviation. We consider linear, power and B-spline series methods to approximate $g(w)$, with the same number of series terms and other tuning parameters as in the empirical example.

Figures 3 and 4 in the Supplemental Material examine the quality of the series approximations in the population. They compare the true quantile function

$$(u, \exp(w)) \mapsto \theta(u, w) = g(w) + v'\beta + \sigma\Phi^{-1}(u),$$

and the quantile price elasticity function

$$(u, \exp(w)) \mapsto \theta(u, w) = \partial_w g(w),$$

to the estimands of the series approximations. In the quantile demand function the value of v is fixed at the sample median values of the ordinal variables and at one for the dummies corresponding to the sample modal values of the rest of the variables (see footnote 11). The estimands are obtained numerically from a mega-sample (a proxy for infinite population) of $100 \times 5,001$ observations with the values of (W, V) as in the data set (repeated 100 times) and with Y generated from the DGP (5.34). Although the derivative function does not depend on u in our design, we do not impose this restriction on the estimands. Both figures show that the power and B-spline estimands are close to the true target functions, whereas the more parsimonious linear approximation misses important curvature features of the target functions, especially in the elasticity function.

To analyze the properties of the inference methods in finite samples, we draw 500 samples from the DGP (5.34) with 4 sample sizes, n: $10,002$, $5,001$, $1,000$, and 500 observations. For $n = 10,002$ we fix W to the values in the data set repeated twice, for $n = 5,001$ we fix W to the values in the data set, whereas for the smaller sample sizes we draw W with replacement from the values in the data set and keep them fixed across samples. To speed up computation, we drop the vector V by fixing it at the sample median values of the ordinal components and at one for the dummies corresponding to the sample modal values for all the individuals. We focus on the average quantile price elasticity function

$$u \mapsto \theta(u) = \int \partial_w g(w) d\mu(w),$$

over the region $I = [0.1, 0.9]$. We estimate this function using linear, power and B-spline quantile regression with the same number of terms and other tuning parameters as in the empirical example. Although $\theta(u)$ does not change with u in our design, again we do not impose this restriction on the estimators. For inference, we compare the performance of 90% confidence bands for the entire elasticity function. These bands are constructed using the pivotal, gradient bootstrap, Gaussian, and weighted bootstrap methods, all implemented in the same fashion as in the empirical example. The interval I is approximated by a finite grid of 91 quantiles $\tilde{I} = \{0.10, 0.11, ..., 0.90\}$.

Table 2 reports estimation and inference results averaged across 500 simulations. The true value of the elasticity function is $\theta(u) = -0.74$ for all $u \in \tilde{I}$. Bias and RMSE are the absolute bias and root mean squared error integrated over $\tilde{I}$. SE/SD reports the ratios of empirical average standard errors to empirical standard deviations. SE/SD uses the

analytical standard errors from expression (4.24). The bandwidth for $\widehat{J}(u)$ is chosen using the Hall-Sheather option of the `quantreg R` package (Hall and Sheather [42]). Length gives the empirical average of the length of the confidence band. SE/SD and length are integrated over the grid of quantiles $\tilde{I}$. Cover reports empirical coverage of the confidence bands with nominal level of 90%. Stat is the empirical average of the 90% quantile of the maximal t-statistic used to construct the bands. Table 2 shows that the linear estimator has higher absolute bias than the more flexible power and B-spline estimators, but displays lower rmse, especially for small sample sizes. The analytical standard errors provide good approximations to the standard deviations of the estimators. The confidence bands have empirical coverage close to the nominal level of 90% for all the estimators and sample sizes considered; and both bootstrap bands tend to have larger average length than the pivotal and Gaussian bands. The source of this difference in coverage might be that the bootstrap methods resample the distribution of the covariates, whereas the pivotal and Gaussian methods condition on the distribution in the sample.

All in all, these results strongly confirm the practical value of the theoretical results and methods developed in the paper. They also support the empirical example by verifying that our estimation and inference methods work quite nicely in a very similar setting.

Appendix A. Implementation Algorithms

Throughout this section we assume that we have a random sample $\{(Y_i, Z_i) : 1 \le i \le n\}$. We are interested in approximating the distribution of the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ or of the statistics associated with functionals of it. Recall that for each quantile $u \in \mathcal{U} \subset (0, 1)$, we estimate $\beta(u)$ by quantile regression $\widehat{\beta}(u) = \arg\min_{\beta \in \mathbb{R}^m} \mathbb{E}_n[\rho_u(Y_i - Z_i'\beta)]$, the Gram matrix Σ by $\widehat{\Sigma} = \mathbb{E}_n[Z_i Z_i']$, and the Jacobian matrix $J(u)$ by Powell [71] estimator $\widehat{J}(u) = \mathbb{E}_n[1\{|Y_i - Z_i'\widehat{\beta}(u)| \le h\} \cdot Z_i Z_i']/2h$, where we recommend choosing the bandwidth h as in the `quantreg R` package with the Hall-Sheather option (Hall and Sheather [42]).

We begin describing the algorithms to implement the methods to approximate the distribution of the process $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ indexed by $\mathcal{U}$. Let B be a pre-specified number of bootstrap or simulation repetitions.

Algorithm 1 (Pivotal method). *(1) For $b = 1, \dots, B$, draw $U_1^b, \dots, U_n^b$ i.i.d. from $U \sim$ Uniform$(0, 1)$ and compute $\mathbb{U}^b(u) = n^{-1/2} \sum_{i=1}^n Z_i(u - 1\{U_i^b \le u\}), \quad u \in \mathcal{U}$. (2) Approximate the distribution of $\{\sqrt{n}(\widehat{\beta}(u) - \beta(u)) : u \in \mathcal{U}\}$ by the empirical distribution of $\{\widehat{J}^{-1}(u)\mathbb{U}^b(u) : u \in \mathcal{U}, 1 \le b \le B\}$.*

Algorithm 2 (Gradient bootstrap method). *(1) For $b = 1, \ldots, B$, draw $U_1^b, \ldots, U_n^b$ i.i.d. from $U \sim Uniform(0, 1)$ and compute $\mathbb{U}^b(u) = n^{-1/2} \sum_{i=1}^n Z_i(u - 1\{U_i^b \leq u\})$, $u \in \mathcal{U}$. (2) For $b = 1, \ldots, B$, estimate the quantile regression process $\widehat{\beta}^b(u) = \arg\min_{\beta \in \mathbb{R}^m} \sum_{i=1}^n \rho_u(Y_i - Z_i'\beta) + \rho_u(Y_{n+1} - X_{n+1}^b(u)'\beta)$, $u \in \mathcal{U}$, where $X_{n+1}^b(u) = -\sqrt{n}\ \mathbb{U}_n^b(u)/u$, and $Y_{n+1} = n \max_{1 \leq i \leq n} |Y_i|$ to ensure $Y_{n+1} > X_{n+1}^b(u)'\widehat{\beta}^b(u)$, for all $u \in \mathcal{U}$. (3) Approximate the distribution of $\{\sqrt{n}(\widehat{\beta}(u) - \beta(u)) : u \in \mathcal{U}\}$ by the empirical distribution of $\{\sqrt{n}(\widehat{\beta}^b(u) - \widehat{\beta}(u)) : u \in \mathcal{U}, 1 \leq b \leq B\}$.*

Algorithm 3 (Gaussian method). *(1) For $b = 1, \ldots, B$, generate a m-dimensional standard Brownian bridge on $\mathcal{U}$, $B_m^b(\cdot)$. Define $G_n^b(u) = \widehat{\Sigma}^{1/2} B_m^b(u)$ for $u \in \mathcal{U}$. (2) Approximate the distribution of $\{\sqrt{n}(\widehat{\beta}(u) - \beta(u)) : u \in \mathcal{U}\}$ by the empirical distribution of $\{\widehat{J}^{-1}(u)G_n^b(u) : u \in \mathcal{U}, 1 \leq b \leq B\}$.*

Algorithm 4 (Weighted bootstrap method). *(1) For $b = 1, \ldots, B$, draw $\pi_1^b, \ldots, \pi_n^b$ i.i.d. from the standard exponential distribution and compute the weighted quantile regression process $\widehat{\beta}^b(u) = \arg\min_{\beta \in \mathbb{R}^m} \sum_{i=1}^n \pi_i^b \rho_u(Y_i - Z_i'\beta)$, $u \in \mathcal{U}$. (2) Approximate the distribution of $\{\sqrt{n}(\widehat{\beta}(u) - \beta(u)) : u \in \mathcal{U}\}$ by the empirical distribution of $\{\sqrt{n}(\widehat{\beta}^b(u) - \widehat{\beta}(u)) : u \in \mathcal{U}, 1 \leq b \leq B\}$.*

The previous algorithms provide approximations to the distribution of $\sqrt{n}(\widehat{\beta}(u) - \beta(u))$ that are uniformly valid in $u \in \mathcal{U}$. We can use these approximations directly to make inference on linear functionals of $Q(u, x)$ including the conditional quantile functions itself, provided the approximation error is small as stated in Conditions P and U. Each linear functional is represented by $\{\theta(u, w) = \ell(w)'\beta(u) + r_n(u, w) : (u, w) \in I\}$, where $\ell(w)'\beta(u)$ is the series approximation, $\ell(w) \in \mathbb{R}^m$ is a loading vector, $r_n(u, w)$ is the remainder term, and I is the set of pairs of quantile indices and covariates values of interest, see Section 4 for details and examples. Next we provide algorithms to conduct pointwise or uniform inference over linear functionals.

Algorithm 5 (Pointwise Inference for Linear Functionals). *(1) Compute the variance estimate $\widehat{\sigma}^2(u, w) = u(1 - u)\ell(w)'\widehat{J}^{-1}(u)\widehat{\Sigma}\widehat{J}^{-1}(u)\ell(w)/n$. (2) Using any of the Algorithms 1-4, compute vectors $V_1(u), \ldots, V_B(u)$ whose empirical distribution approximates the distribution of $\sqrt{n}(\widehat{\beta}(u) - \beta(u))$. (3) For $b = 1, \ldots, B$, compute the t-statistic $t^{*b}(u, w) = \left| \frac{\ell(w)'V_b(u)}{\sqrt{n}\widehat{\sigma}(u,w)} \right|$. (4) Form a $(1 - \alpha)$-confidence interval for $\theta(u, w)$ as $\ell(w)'\widehat{\beta}(u) \pm k(1 - \alpha)\widehat{\sigma}(u, w)$, where $k(1 - \alpha)$ is the $1 - \alpha$ sample quantile of $\{t^{*b}(u, w) : 1 \leq b \leq B\}$.*

Algorithm 6 (Uniform Inference for Linear Functionals). *(1) Compute the variance estimates $\widehat{\sigma}^2(u, w) = u(1 - u)\ell(w)'\widehat{J}^{-1}(u)\widehat{\Sigma}\widehat{J}^{-1}(u)\ell(w)/n$ for $(u, w) \in I$. (2) Using any of the*

Algorithms 1-4, compute the processes $V_1(\cdot), \dots, V_B(\cdot)$ whose empirical distribution approximates the distribution of $\{\sqrt{n}(\widehat{\beta}(u) - \beta(u)) \colon u \in \mathcal{U}\}$. (3) For $b = 1, \dots, B$, compute the maximal t-statistic $\|t^{*b}\|_I = \sup_{(u,w)\in I} \left| \frac{\ell(w)'V_b(u)}{\sqrt{n}\widehat{\sigma}(u,w)} \right|$. (4) Form a $(1-\alpha)$-confidence band for $\{\theta(u,w) \colon (u,w) \in I\}$ as $\{\ell(w)'\widehat{\beta}(u) \pm k(1-\alpha)\widehat{\sigma}(u,w) \colon (u,w) \in I\}$, where $k(1-\alpha)$ is the $1-\alpha$ sample quantile of $\{\|t^{*b}\|_I : 1 \le b \le B\}$.

Table 2. Finite Sample Properties of Estimation and Inference Methods for Average Quantile Elasticity Function

	Bias	RMSE	SE/SD	Pivotal			Gradient Bootstrap			Gaussian			Weighted Bootstrap		
				Cover	Length	Stat	Cover	Length	Stat	Cover	Length	Stat	Cover	Length	Stat
n = 10,002															
Linear	0.04	0.10	0.98	85	0.54	2.64	86	0.57	2.83	85	0.54	2.64	88	0.57	2.82
Power	0.00	0.11	0.96	90	0.60	2.65	93	0.64	2.81	89	0.59	2.65	91	0.63	2.80
B-splines	0.01	0.11	0.96	88	0.61	2.65	94	0.65	2.80	88	0.61	2.64	90	0.65	2.79
n = 5,001															
Linear	0.04	0.14	1.03	90	0.77	2.64	90	0.83	2.87	90	0.77	2.64	91	0.83	2.88
Power	0.00	0.15	1.01	91	0.85	2.65	94	0.95	2.88	91	0.85	2.65	94	0.91	2.84
B-splines	0.01	0.15	1.02	91	0.86	2.65	93	0.95	2.85	90	0.86	2.65	92	0.93	2.83
n = 1,000															
Linear	0.03	0.30	1.06	89	1.78	2.64	90	1.93	3.01	90	1.78	2.64	91	1.91	2.99
Power	0.01	0.33	1.07	92	2.00	2.66	100	5.28	3.92	91	2.00	2.65	97	2.25	2.98
B-splines	0.01	0.35	1.05	90	2.07	2.65	98	2.90	3.20	90	2.07	2.65	95	2.38	2.97
n = 500															
Linear	0.02	0.44	1.05	90	2.61	2.64	95	2.84	3.08	91	2.61	2.64	90	2.77	3.05
Power	0.01	0.52	1.06	92	3.13	2.65	100	21.48	4.79	92	3.14	2.66	98	3.37	3.08
B-splines	0.01	0.54	1.06	91	3.26	2.65	100	7.52	3.67	91	3.26	2.65	96	3.46	3.04

Notes: 500 repetitions. Simulation standard error for coverage probability is 1%.
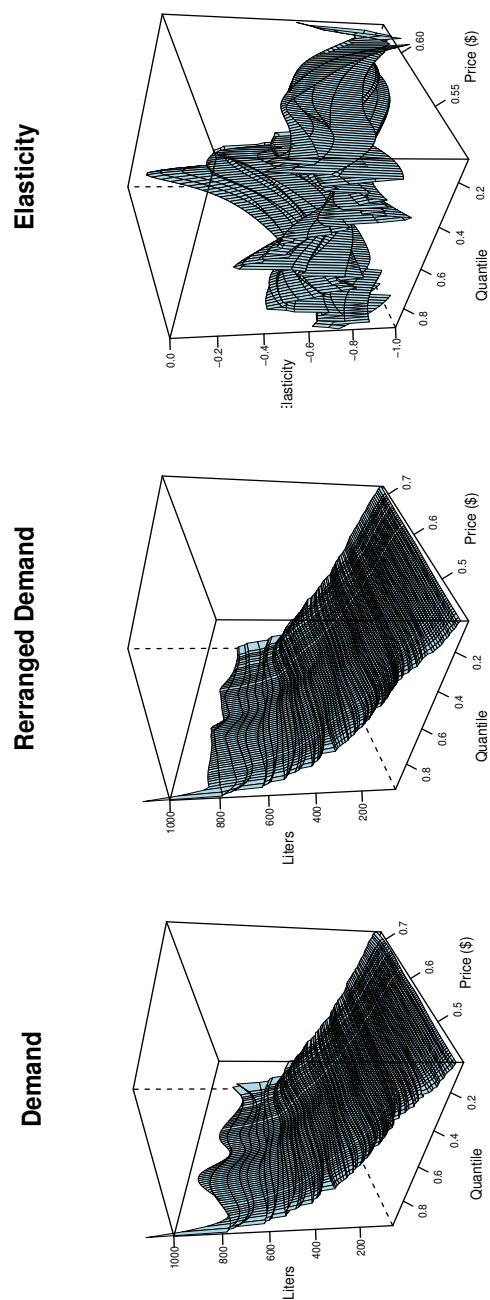
45

FIGURE 1. Demand and price elasticity surfaces as a function of price and the quantile index using B-spline specification. The elasticity estimates are smoothed by local weighted polynomial regression with bandwidth 0.5.
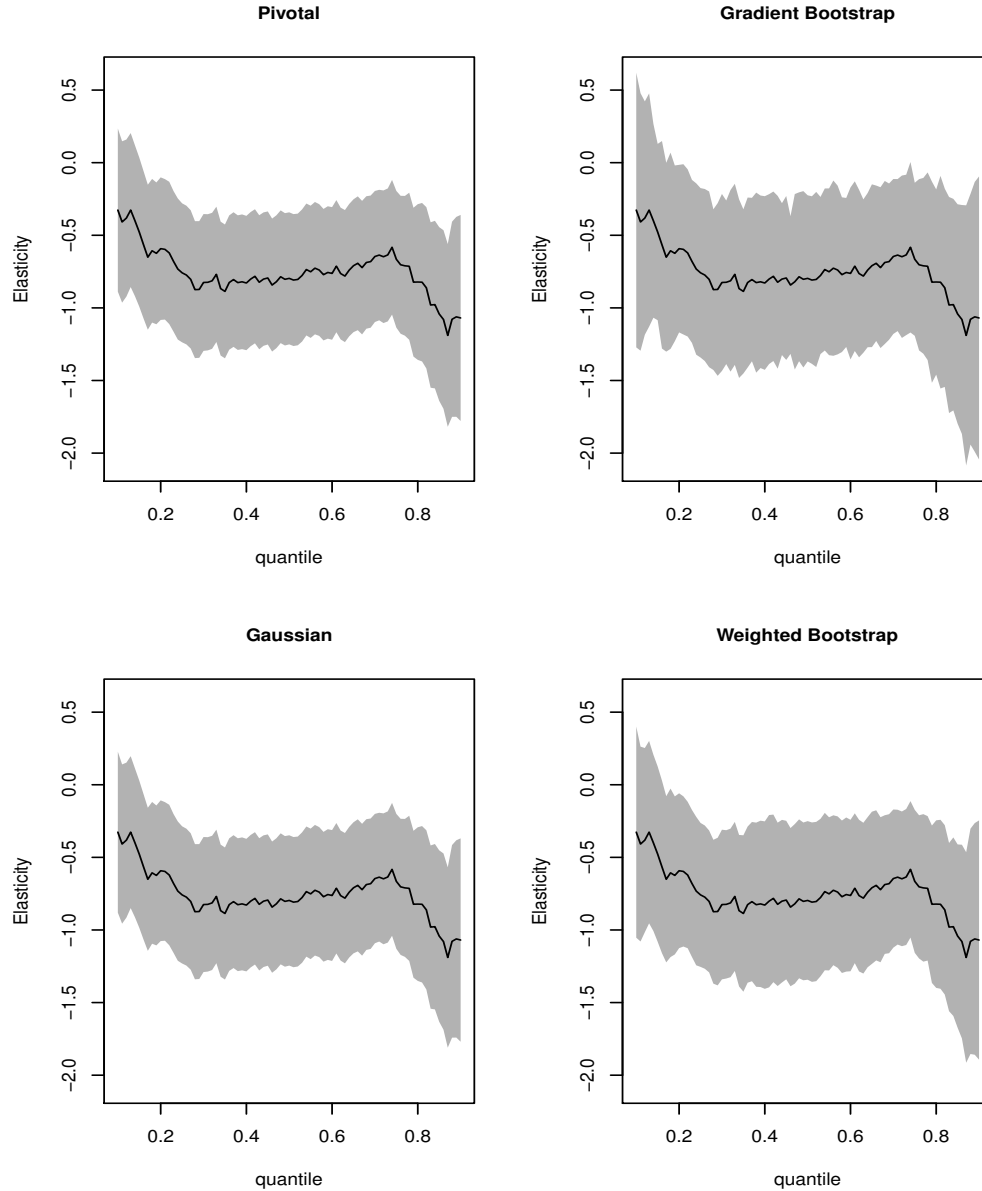
FIGURE 2. 90% Confidence bands for the average quantile price elasticity function using B-spline specification. Pivotal and Gaussian bands are obtained by 1,000 simulations. Gradient bootstrap bands are based on 199 bootstrap repetitions. Weighted bootstrap bands are based on 199 bootstrap repetitions with standard exponential weights.

Supplement to "Conditional Quantile Processes Based on Series or Many Regressors"

By Alexandre Belloni, Victor Chernozhukov, Denis Chetverikov, and Iván Fernández-Val

Appendix B. Imposing Monotonicity on Linear Functionals

Consider the setting of Section 4. The functionals of interest might be naturally monotone in some of their arguments. For example, the conditional quantile function is increasing in the quantile index and the conditional quantile demand function is decreasing in price and increasing in the quantile index. Therefore, it might be desirable to impose the same requirements on the estimators of these functions.

Let $\theta(u, w)$, where $(u, w) \in I$, be a weakly increasing function in (u, w), i.e. $\theta(w', u') \leq \theta(u, w)$ whenever $(w', u') \leq (u, w)$ componentwise.[13] Let $\widehat{\theta}$ and $[\check{i}, \hat{i}]$ be the point and band estimators of θ, constructed using one of the methods described in the previous sections. These estimators might not satisfy the monotonicity requirement due to either estimation error or imperfect approximation. However, we can monotonize these estimates and perform inference using the method suggested in Chernozhukov, Fernández-Val and Galichon [29].

Let $q, f \colon I \to K$, where K is a bounded subset of $\mathbb{R}$, and consider any monotonization operator $\mathcal{M}$ that satisfies: (1) a monotone-neutrality condition

$$\mathcal{M}q = q \quad \text{if } q \text{ monotone;}\tag{B.35}$$

(2) a distance-reducing condition

$$\|\mathcal{M}q - \mathcal{M}f\|_I \leq \|q - f\|_I;\tag{B.36}$$

and (3) an order-preserving condition

$$q \leq f \quad \text{implies} \quad \mathcal{M}q \leq \mathcal{M}f.\tag{B.37}$$

Examples of operators that satisfy these conditions include the multivariate rearrangement (Chernozhukov, Fernández-Val and Galichon [29]), isotonic projection (Barlow et al [6]), convex combinations of rearrangement and isotonic regression (Chernozhukov, Fernández-Val and Galichon [29]), and convex combinations of monotone minorants and monotone majorants.

---

[13]If $\theta(u, w)$ is decreasing in w, we take the transformation $\tilde{w} = -w$ and $\tilde{\theta}(\tilde{w}, u) = \theta(-\tilde{w}, u)$, where $\tilde{\theta}(\tilde{w}, u)$ is increasing in $\tilde{w}$.

Let $\mathcal{M}\widehat{\theta}$ and $[\mathcal{M}i, \mathcal{M}\ddot{\imath}]$ be the monotonized QR-series process and confidence band. We establish that $\mathcal{M}\widehat{\theta}$ has smaller estimation error than the original QR-series process $\widehat{\theta}$, and that $[\mathcal{M}i, \mathcal{M}\ddot{\imath}]$ has higher coverage and smaller length than the original confidence band $[i, \ddot{\imath}]$. The following result is a corollary from Theorems 10 and 15 using the same arguments as in Propositions 2 and 3 in Chernozhukov, Fernández-Val, and Galichon [29].

Corollary 3 (Inference for Monotone Linear Functionals). *Let $\theta : I \to K$ be weakly increasing over I and $\widehat{\theta}$ be the QR-series process of Theorem 10. If $\mathcal{M}$ satisfies the conditions (B.35) and (B.36), then the monotonized QR-series process is necessarily closer to the true value:*

$$\|\mathcal{M}\widehat{\theta} - \theta\|_I \leq \|\widehat{\theta} - \theta\|_I.$$

Let $[i, \ddot{\imath}]$ be a confidence band for θ of Theorem 15. If $\mathcal{M}$ satisfies the conditions (B.35) and (B.37), the monotonized confidence bands maintain at least the asymptotic level of the original bands:

$$P\Big(\theta(u, w) \in [\mathcal{M}i(u, w), \mathcal{M}\ddot{\imath}(u, w)] : (u, w) \in I\Big) \geq 1 - \alpha + o(1).$$

If $\mathcal{M}$ satisfies the condition (B.36), the monotonized confidence bands are shorter in length than the original bands:

$$\|\mathcal{M}\ddot{\imath} - \mathcal{M}i\|_I \leq \|\ddot{\imath} - i\|_I.$$

APPENDIX C. QR-SERIES APPROXIMATION ERROR AND PROOF OF LEMMA 1

In this section we study properties of the QR-series approximating functions and, in particular, we prove Lemma 1. In what follows, for any function $h\colon \mathcal{X} \to \mathbb{R}$, we define

$$Q_u(h) = E\Big[\rho_u(Y - h(X)) - \rho_u(Y - Q(u, X))\Big],$$

so that $\beta(u) \in \arg\min_{\beta \in \mathbb{R}^m} Q_u(Z'\beta)$. For brevity and with some abuse of notation, depending on the context, we use h to denote either a function or a random variable $h = h(X)$ (the same applies to the vector of functions Z). Also, let $\bar{f} = \sup_{x \in \mathcal{X}, y \in \mathcal{Y}_x} f_{Y|X}(y|x)$ and $\underline{f} = \inf_{x \in \mathcal{X}, u \in \mathcal{U}} f_{Y|X}(Q(u, x)|x)$. In addition, let $f'_{Y|X}(y|x)$ denote the derivative of the function $y \mapsto f_{Y|X}(y|x)$ and define $\bar{f}' = \sup_{x \in \mathcal{X}, y \in \mathcal{Y}_x} |f'_{Y|X}(y|x)|$. Observe that $\bar{f}$ and $\bar{f}'$ are bounded from above and $\underline{f}$ is bounded away from zero uniformly over n by Condition S.2. Moreover, let $\lambda_{\min}$ denote the minimal eigenvalue of the matrix $\Sigma = E[ZZ']$. Observe that $\lambda_{\min}$ is bounded away from zero uniformly over n by Condition S.3.

For $u \in \mathcal{U}$, consider the best L_2-approximation to the conditional u-quantile function $x \mapsto Q(u, x)$ by a linear combination of functions in the vector Z, namely

$$\widetilde{\beta}(u) = \arg \min_{\beta \in \mathbb{R}^m} E\Big[|Z(X)'\beta - Q(u, X)|^2\Big]. \tag{C.38}$$

Define

$$c_{u,2}^2 = E\Big[|Z(X)'\widetilde{\beta}(u) - Q(u, X)|^2\Big] \text{ and } c_{u,\infty} = \sup_{x \in \mathcal{X}} \Big|Z(x)'\widetilde{\beta}(u) - Q(u, x)\Big|. \tag{C.39}$$

The next lemma relates the QR-series approximation error $R(u, x) = Q(u, x) - Z(x)'\beta(u)$ to $c_{u,2}$ and $c_{u,\infty}$. Define $\zeta_m = \sup_{x \in \mathcal{X}} \|Z(x)\|$.

Lemma 2 (Bounds on QR-series approximation error). *Assume that Conditions S.2-S.3 hold. In addition assume that*

$$c_{u,\infty}^2 \leq \frac{\underline{f}^3}{12\overline{f}(\overline{f'})^2} \quad \text{and} \quad \zeta_m^2 c_{u,2}^2 \leq \frac{\underline{f}^3 \lambda_{\min}}{12\overline{f}(\overline{f'})^2}.$$

Then we have for all $u \in \mathcal{U}$ that

$$E\Big[|Z(X)'\beta(u) - Q(u, X)|^2\Big] \leq 3(\overline{f}/\underline{f})c_{u,2}^2, \tag{C.40}$$

$$\sup_{x \in \mathcal{X}} \Big|Z(x)'\beta(u) - Q(u, x)\Big| \leq \lambda_{\min}^{-1/2} \zeta_m \Big(1 + \sqrt{3\overline{f}/\underline{f}}\Big)c_{u,2} + c_{u,\infty}. \tag{C.41}$$

Proof of Lemma 2. Fix $u \in \mathcal{U}$. Let $g_u : \mathcal{X} \to \mathbb{R}$ be the function defined by $g_u(x) = Q(u, x)$ for all $x \in \mathcal{X}$.

Step 1 (Main argument). Recall that $Z = Z(X)$. For notational convenience, let

$$\bar{q} = \frac{\underline{f}^{3/2}\Big(E\Big[|Z'\beta(u) - g_u(X)|^2\Big]\Big)^{3/2}}{\overline{f'} E\Big[|Z'\beta(u) - g_u(X)|^3\Big]}. \tag{C.42}$$

By Steps 2 and 3 below, we have respectively that

$$Q_u(Z'\beta(u)) - Q_u(g_u) \leq \overline{f}c_{u,2}^2, \tag{C.43}$$

$$Q_u(Z'\beta(u)) - Q_u(g_u) \geq \frac{\underline{f}E[|Z'\beta(u) - g_u(X)|^2]}{3} \wedge \Big(\frac{\bar{q}}{3}\sqrt{\underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right]}\Big). \tag{C.44}$$

Thus,

$$\frac{\underline{f}E[|Z'\beta(u) - g_u(X)|^2]}{3} \wedge \Big(\frac{\bar{q}}{3}\sqrt{\underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right]}\Big) \leq \overline{f}c_{u,2}^2. \tag{C.45}$$

Therefore, if the minimum in (C.45) is achieved by the first term, it immediately follows that $E[|Z'\beta(u) - g_u(X)|^2] \leq 3(\overline{f}/\underline{f})c_{u,2}^2$, which is (C.40). On the other hand, if the minimum

in (C.45) is achieved by the second term, it follows from the inequality $\bar{f}c_{u,2}^2 \leq \bar{q}^2/3$, which is established in Step 4 below, that

$$\frac{\bar{q}}{3}\sqrt{\underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right]} \leq \frac{\bar{q}}{\sqrt{3}} \cdot (\bar{f}c_{u,2}^2)^{1/2},$$

which again gives (C.40). Hence, (C.40) follows.

To prove (C.41), note that it follows from the triangle inequality, (C.39), and (C.40) that

$$\sqrt{E\left[|Z'\beta(u) - Z'\widetilde{\beta}(u)|^2\right]} \leq \sqrt{E\left[|Z'\beta(u) - g_u(X)|^2\right]} + \sqrt{E\left[|g_u(X) - Z'\widetilde{\beta}(u)|^2\right]}$$

$$\leq \sqrt{3\bar{f}/\underline{f}}c_{u,2} + c_{u,2} = \left(1 + \sqrt{3\bar{f}/\underline{f}}\right)c_{u,2},$$

and so

$$\sup_{x \in \mathcal{X}} |Z(x)'\beta(u) - g_u(x)| \leq \sup_{x \in \mathcal{X}} |Z(x)'\beta(u) - Z(x)'\widetilde{\beta}(u)| + \sup_{x \in \mathcal{X}} |Z(x)'\widetilde{\beta}(u) - g_u(x)|$$

$$\leq \zeta_m \|\beta(u) - \widetilde{\beta}(u)\| + c_{u,\infty}$$

$$\leq \lambda_{\min}^{-1/2}\zeta_m\sqrt{E\left[|Z'\beta(u) - Z'\widetilde{\beta}(u)|^2\right]} + c_{u,\infty}$$

$$\leq \lambda_{\min}^{-1/2}\zeta_m\left(1 + \sqrt{3\bar{f}/\underline{f}}\right)c_{u,2} + c_{u,\infty}.$$

Hence, (C.40) follows. This completes Step 1.

Step 2 (Upper Bound). Here we establish (C.43). Observe that for any two scalars w and v, we have

$$\rho_u(w - v) - \rho_u(w) = -v(u - 1\{w \leq 0\}) + \int_0^v (1\{w \leq t\} - 1\{w \leq 0\})dt. \qquad \text{(C.46)}$$

Therefore, for any function $h\colon \mathcal{X} \to \mathbb{R}$, using the law of iterated expectations and (C.46) with $w = Y - g_u(X)$ and $v = h(X) - g_u(X)$, we obtain

$$Q_u(h) - Q_u(g_u) = E\left[\int_0^{h-g_u}(F_{Y|X}(g_u + t|X) - F_{Y|X}(g_u|X))dt\right]$$

$$= E\left[\int_0^{h-g_u} tf_{Y|X}(g_u + \tilde{t}_{X,t}|X)dt\right] \leq (\bar{f}/2)E[|h - g_u|^2]$$

for some $\tilde{t}_{X,t}$ between 0 and t. Thus,

$$Q_u(Z'\beta(u)) - Q_u(g_u) \leq Q_u(Z'\widetilde{\beta}(u)) - Q_u(g_u) \leq \bar{f}c_{u,2}^2,$$

which gives (C.43). This completes Step 2.

Step 3 (Lower Bound). Here we establish (C.44). For any function $h\colon \mathcal{X} \to \mathbb{R}$, we have

$$Q_u(h) - Q_u(g_u) = E\left[\int_0^{h-g_u}(F_{Y|X}(g_u + t|X) - F_{Y|X}(g_u|X))dt\right] \tag{C.47}$$

$$= E\left[\int_0^{h-g_u} tf_{Y|X}(g_u|X) + (t^2/2)f'_{Y|X}(g_u + \tilde{t}_{X,t}|X)dt\right] \tag{C.48}$$

$$\geq (\underline{f}/2)E[|h - g_u|^2] - (1/6)\bar{f}'E[|h - g_u|^3] \tag{C.49}$$

for some $\tilde{t}_{X,t}$ between 0 and t. Consider the case $(\underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right])^{1/2} \leq \bar{q}$ for $\bar{q}$ defined in (C.42). By definition of $\bar{q}$, we have

$$\bar{f}'E\left[|Z'\beta(u) - g_u(X)|^3\right] \leq \underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right],$$

and so using the inequality (C.49) with $h = Z'\beta(u)$ yields

$$Q_u(Z'\beta(u)) - Q_u(g_u) \geq \frac{\underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right]}{3},$$

which implies (C.44).

Now consider the case $(\underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right])^{1/2} > \bar{q}$. Let

$$h_u(x) = (1 - \alpha)Z(x)'\beta(u) + \alpha g_u(x), \quad x \in \mathcal{X},$$

where $\alpha \in (0, 1)$ is picked so that $(\underline{f}E\left[|h_u - g_u|^2\right])^{1/2} = \bar{q}$ (we can assume that $\bar{q} > 0$, so that such α exists, since otherwise the claim of the lemma is trivial). Then

$$1 - \alpha = \frac{\bar{q}}{\sqrt{\underline{f}E[|Z'\beta(u) - g_u(X)|^2]}},$$

and by convexity of Q_u, we have

$$Q_u(Z'\beta(u)) - Q_u(g_u) \geq \frac{1}{1 - \alpha} \cdot (Q_u(h_u) - Q_u(g_u))$$

$$= \frac{\sqrt{\underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right]}}{\bar{q}} \cdot (Q_u(h_u) - Q_u(g_u)).$$

Also, note that

$$\bar{q} = \frac{\underline{f}^{3/2}}{\bar{f}'} \frac{(E[|Z'\beta(u) - g_u(X)|^2])^{3/2}}{E[|Z'\beta(u) - g_u(X)|^3]} = \frac{\underline{f}^{3/2}}{\bar{f}'} \frac{(E[|h_u - g_u|^2])^{3/2}}{E[|h_u - g_u|^3]} = \frac{1}{\bar{f}'} \frac{\bar{q}^3}{E[|h_u - g_u|^3]},$$

so that

$$\bar{f}'E[|h_u - g_u|^3] = \bar{q}^2.$$

Hence, using the inequality (C.49) with $h = h_u$ gives

$$Q_u(h_u) - Q_u(g_u) \geq (\underline{f}/2)E[|h_u - g_u|^2] - (1/6)\bar{f}'E[|h_u - g_u|^3] = \bar{q}^2/2 - \bar{q}^2/6 = \bar{q}^2/3.$$

Combining the inequalities above yields

$$Q_u(Z'\beta(u)) - Q_u(g_u) \geq \frac{\bar{q}}{3}\sqrt{\underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right]},$$

which implies (C.44). This completes Step 3.

Step 4 (Auxiliary inequality). Here we show that $\bar{f}c_{u,2}^2 \leq \bar{q}^2/3$. Recall that $\zeta_m = \sup_{x \in \mathcal{X}} \|Z(x)\|$. Also, note that since $\widetilde{\beta}(u)$ solves the problem (C.38), we have

$$E\left[(g_u(X) - Z'\widetilde{\beta}(u)) \cdot (Z'(\beta(u) - \widetilde{\beta}(u)))\right] = 0,$$

and so

$$E\left[(g_u(X) - Z'\beta(u))^2\right] = E\left[(g_u(X) - Z'\widetilde{\beta}(u))^2\right] + E\left[(Z'(\beta(u) - \widetilde{\beta}(u)))^2\right].$$

Hence,

$$E\left[(g_u(X) - Z'\beta(u))^2\right] \geq E\left[(g_u(X) - Z'\widetilde{\beta}(u))^2\right],$$

$$E\left[(g_u(X) - Z'\beta(u))^2\right] \geq E\left[(Z'(\beta(u) - \widetilde{\beta}(u)))^2\right].$$

Therefore, the ratio $(E[|Z'\beta(u) - g_u(X)|^2])^{3/2}/E[|Z'\beta(u) - g_u(X)|^3]$ is bounded from below by

$$\frac{E[|Z'\beta(u) - g_u(X)|^2]^{1/2}}{\sup_{x \in \mathcal{X}} |Z(x)'\beta(u) - g_u(x)|}$$

$$\geq \frac{1}{2} \frac{E[|Z'\beta(u) - Z'\widetilde{\beta}(u)|^2]^{1/2} + E[|Z'\widetilde{\beta}(u) - g_u(X)|^2]^{1/2}}{\sup_{x \in \mathcal{X}} |Z(x)'\beta(u) - Z(x)'\widetilde{\beta}(u)| + \sup_{x \in \mathcal{X}} |Z(x)'\widetilde{\beta}(u) - g_u(x)|}$$

$$\geq \frac{1}{2} \left(\frac{(E[|Z'\beta(u) - Z'\widetilde{\beta}(u)|^2])^{1/2}}{\sup_{x \in \mathcal{X}} |Z(x)'\beta(u) - Z(x)'\widetilde{\beta}(u)|} \wedge \frac{(E[|Z'\widetilde{\beta}(u) - g_u(X)|^2])^{1/2}}{\sup_{x \in \mathcal{X}} |Z(x)'\widetilde{\beta}(u) - g_u(x)|} \right)$$

$$\geq \frac{1}{2} \left(\frac{\lambda_{\min}^{1/2}\|\beta(u) - \widetilde{\beta}(u)\|}{\zeta_m\|\beta(u) - \widetilde{\beta}(u)\|} \wedge \frac{c_{u,2}}{c_{u,\infty}} \right) = \frac{1}{2}\left(\frac{\lambda_{\min}^{1/2}}{\zeta_m} \wedge \frac{c_{u,2}}{c_{u,\infty}} \right),$$

where the inequality in the last line follows from (C.39). So,

$$\frac{\bar{q}^2}{3} \geq \frac{\underline{f}^3}{12(\bar{f}')^2}\left(\frac{\lambda_{\min}}{\zeta_m^2} \wedge \frac{c_{u,2}^2}{c_{u,\infty}^2} \right) \geq \bar{f}c_{u,2}^2$$

since $c_{u,\infty}^2 \leq \underline{f}^3/(12(\bar{f}')^2\bar{f})$ and $\zeta_m^2 c_{u,2}^2 \leq \underline{f}^3\lambda_{\min}/(12(\bar{f}')^2\bar{f})$. This completes Step 4 and the proof of the lemma. $\square$

Proof of Lemma 1. Fix $u \in \mathcal{U}$. All inequalities and convergence statements in the proof apply uniformly over $u \in \mathcal{U}$ but we fix u so that we do not have to repeat "uniformly over $u \in \mathcal{U}$" many times.

First, consider the case of polynomials. We have $\zeta_m \lesssim m$; see Newey [67]. Also, since $Q(u, \cdot) \in \Sigma(s, C, \mathcal{X})$ for some constant $C > 0$, it follows that

$$\inf_{\beta \in \mathbb{R}^m} \sup_{x \in \mathcal{X}} |Z(x)'\beta - Q(u, x)| \lesssim m^{-s/d};$$

see Chen [21]. Therefore, by (C.38) and (C.39),

$$c_{u,2} = \left(E[(Z'\widetilde{\beta}(u) - Q(u, X))^2] \right)^{1/2} = \min_{\beta \in \mathbb{R}^m} \left(E\left[|Z(X)'\beta - Q(u, X)|^2 \right] \right)^{1/2}$$

$$\lesssim \inf_{\beta \in \mathbb{R}^m} \sup_{x \in \mathcal{X}} |Z(x)'\beta - Q(u, x)| \lesssim m^{-s/d}.$$

Also, it follows from Proposition 3.1 in Belloni et al [9] that

$$c_{u,\infty} = \sup_{x \in \mathcal{X}} |Z(x)'\widetilde{\beta}(u) - Q(u, x)| \lesssim (1 + \zeta_m)m^{-s/d} \lesssim m^{1-s/d}$$

since $\zeta_m \lesssim m$. Moreover, under the assumed condition $s > d$, since $c_{u,2} \lesssim m^{-s/d}$ and $c_{u,\infty} \lesssim m^{1-s/d}$, we have that $c_{u,\infty} \to 0$ and $\zeta_m c_{u,2} \to 0$ if $m \to \infty$. Hence, the inequalities (2.8) follow from Lemma 2.

Next, consider the case of B-splines. We have $\zeta_m \lesssim \sqrt{m}$; see Newey [67]. Also, since $Q(u, \cdot) \in \Sigma(s, C, \mathcal{X})$ for some constant $C > 0$, it follows that

$$\inf_{\beta \in \mathbb{R}^m} \sup_{x \in \mathcal{X}} |Z(x)'\beta - Q(u, x)| \lesssim m^{-(s \wedge s_0)/d};$$

see Chen [21]. Therefore, by the same argument as above,

$$c_{u,2} = \left(E[(Z'\widetilde{\beta}(u) - Q(u, X))^2] \right)^{1/2} \lesssim \inf_{\beta \in \mathbb{R}^m} \sup_{x \in \mathcal{X}} |Z(x)'\beta - Q(u, x)| \lesssim m^{-(s \wedge s_0)/d}.$$

Also, it follows from Proposition 3.1 in Belloni et al [9] that

$$c_{u,\infty} = \sup_{x \in \mathcal{X}} |Z(x)'\widetilde{\beta}(u) - Q(u, x)| \lesssim (1 + \zeta_m)m^{-(s \wedge s_0)/d} \lesssim m^{1/2-(s \wedge s_0)/d}$$

since $\zeta_m \lesssim \sqrt{m}$. Moreover, under the assumed condition $s \wedge s_0 > d$, since $c_{u,2} \lesssim m^{-(s \wedge s_0)/d}$ and $c_{u,\infty} \lesssim m^{1/2-(s \wedge s_0)/d}$, we have that $c_{u,\infty} \to 0$ and $\zeta_m c_{u,2} \to 0$ if $m \to \infty$. Hence, the first inequality in (2.9) follows from Lemma 2.

To prove the second inequality in (2.9), note that Lemma 2 also implies that

$$\sup_{x \in \mathcal{X}} |Z(x)'\beta(u) - Q(u, x)| \lesssim m^{1/2-(s \wedge s_0)/d}.$$

Further, let

$$f_{u0} = \int_{x \in \mathcal{X}} f_X(x) f_{Y|X}(Q(u,x)|x) dx \tag{C.50}$$

and let X_u be a random variable with the support $\mathcal{X}$ and the pdf

$$f_u(x) = f_{u0}^{-1} f_X(x) f_{Y|X}(Q(u,x)|x), \quad x \in \mathcal{X}. \tag{C.51}$$

Note that $f_u(x)$ is bounded from above and away from zero uniformly over $x \in \mathcal{X}$. Define

$$\bar{\beta}(u) = \arg \min_{\beta \in \mathbb{R}^m} E\left[|Z(X_u)'\beta - Q(u,X_u)|^2\right]. \tag{C.52}$$

Applying Theorem A.1 and Lemma 5.1 in Huang [50] shows that

$$\sup_{x \in \mathcal{X}} |Z(x)'\bar{\beta}(u) - Q(u,x)| \lesssim \inf_{\beta \in \mathbb{R}^m} \sup_{x \in \mathcal{X}} |Z(x)'\beta - Q(u,x)| \lesssim m^{-(s \wedge s_0)/d}. \tag{C.53}$$

(Note that applying Theorem A.1 in Huang [50] requires verifying his Conditions A.1–A.3. Condition A.1 holds because $f_u(x)$ is bounded from above and away from zero uniformly over $x \in \mathcal{X}$; Conditions A.2 and A.3 hold because the vector of approximating functions Z consists of tensor products of B-splines with uniform knot sequence, see the discussion in Huang's paper for clarifications.) Hence, by the triangle inequality,

$$\sup_{x \in \mathcal{X}} |Z(x)'(\beta(u) - \bar{\beta}(u))| \to 0 \tag{C.54}$$

as $m \to \infty$. Moreover, since $\bar{\beta}(u)$ solves the optimization problem (C.52), we have

$$E\left[|Z(X_u)'(\beta(u) - \bar{\beta}(u))|^2\right]$$
$$= E\left[|Z(X_u)'\beta(u) - Q(u,X_u)|^2\right] - E\left[|Z(X_u)'\bar{\beta}(u) - Q(u,X_u)|^2\right]. \tag{C.55}$$

Also, let $\underline{f}_u = \inf_{x \in \mathcal{X}} f_u(x)$ and $\bar{f}_X = \sup_{x \in \mathcal{X}} f_X(x)$. Note that both $\underline{f}_u$ and f_{u0} are strictly positive and both $\bar{f}_X$ and $\bar{f}'$ are finite (recall that f_{u0} is defined in (C.50) and $\bar{f}'$ is defined in the beginning of this section).

Now, as in (C.47)-(C.48) of the proof of Lemma 2, for $g_u(\cdot) = Q(u,\cdot)$, we have

$$Q_u(Z'\beta(u)) - Q_u(g_u) = E\left[\int_0^{Z'\beta(u)-g_u} (t f_{Y|X}(g_u|X) + (t^2/2) f'_{Y|X}(g_u + \tilde{t}_{X,t}|X)) dt\right]$$
$$\geq (f_{u0}/2) E\left[|Z(X_u)'\beta(u) - Q_u(u,X_u)|^2\right]$$
$$- (1/6)(\bar{f}_X \bar{f}'/\underline{f}_u) E\left[|Z(X_u)'\beta(u) - Q_u(u,X_u)|^3\right]$$

and

$$Q_u(Z'\bar{\beta}(u)) - Q_u(g_u) = E\left[\int_0^{Z'\bar{\beta}(u)-g_u} (t f_{Y|X}(g_u|X) + (t^2/2) f'_{Y|X}(g_u + \tilde{t}_{X,t}|X)) dt\right]$$

$$\leq (\bar{f}_{u0}/2) E\left[|Z(X_u)'\bar{\beta}(u) - Q_u(u, X_u)|^2\right]$$

$$+ (1/6)(\bar{f}_X \bar{f}'/\underline{f}_u) E\left[|Z(X_u)'\bar{\beta}(u) - Q_u(u, X_u)|^3\right],$$

where inequalities follow from the definition of the pdf of X_u in (C.51). Combining these inequalities with (C.55) and using the fact that $Q_u(Z'\beta(u)) \leq Q_u(Z'\bar{\beta}(u))$ gives

$$E\left[|Z(X_u)'(\beta(u) - \bar{\beta}(u))|^2\right]$$

$$\leq \frac{2\bar{f}_X \bar{f}'}{3 \bar{f}_{u0} \underline{f}_u}\left(E\left[|Z(X_u)'\beta(u) - Q_u(u, X_u)|^3\right] + E\left[|Z(X_u)'\bar{\beta}(u) - Q_u(u, X_u)|^3\right]\right). \quad \text{(C.56)}$$

In turn, $(\bar{f}_X \bar{f}')/(\bar{f}_{u0} \underline{f}_u)$ is finite, and so the right-hand of the inequality (C.56) is bounded from above up-to a constant by

$$E\left[|Z(X_u)'(\beta(u) - \bar{\beta}(u))|^3\right] + E\left[|Z(X_u)'\bar{\beta}(u) - Q_u(u, X_u)|^3\right]$$

$$\leq c E\left[|Z(X_u)'(\beta(u) - \bar{\beta}(u))|^2\right] + \sup_{x\in\mathcal{X}} |Z(x)'\bar{\beta}(u) - Q_u(u, x)|^3,$$

where c is a constant that is arbitrarily small (uniformly over $u \in \mathcal{U}$) if m is large enough; see (C.54). Hence,

$$E\left[|Z(X_u)'(\beta(u) - \bar{\beta}(u))|^2\right] \lesssim \sup_{x\in\mathcal{X}} |Z(x)'\bar{\beta}(u) - Q_u(u, x)|^3 \lesssim m^{-3(s\wedge s_0)/d}, \quad \text{(C.57)}$$

where the second inequality follows from (C.53). Therefore,

$$\|\beta(u) - \bar{\beta}(u)\|^2 \lesssim E\left[|Z(X_u)'(\beta(u) - \bar{\beta}(u))|^2\right] \lesssim m^{-3(s\wedge s_0)/d}.$$

Conclude that

$$\sup_{x\in\mathcal{X}} |Q(u, x) - Z(x)'\beta(u)| \leq \sup_{x\in\mathcal{X}} |Q(u, x) - Z(x)'\bar{\beta}(u)| + \sup_{x\in\mathcal{X}} |Z(x)'(\bar{\beta}(u) - \beta(u))|$$

$$\lesssim m^{-(s\wedge s_0)/d} + \zeta_m \|\beta(u) - \bar{\beta}(u)\| \lesssim m^{-(s\wedge s_0)/d} + m^{1/2 - (3/2)(s\wedge s_0)/d} \lesssim m^{-(s\wedge s_0)/d},$$

since $(s \wedge s_0) > d$. This gives the second inequality in (2.9). The last claim of the lemma follows from (2.8) and (2.9). This completes the proof of the lemma. $\qquad\square$

APPENDIX D. PROOFS FOR SECTION 3

In this section we gather proofs of Theorems 1-7 and Corollaries 1 and 2 from the main text. We adopt the standard notation of the empirical process literature (van der Vaart and Wellner [80]). In addition, for notational convenience, we write $\psi_i(\beta, u) = Z_i(1\{Y_i \leq Z_i'\beta\} - u)$, where $Z_i = Z(X_i)$ and $i = 1, \ldots, n$. Also, for a sequence of scalars $(r_n)_{n \geq 1}$ such that $r_n \to 0$, we define the set

$$R_{n,m} := \left\{ (u, \beta) \in \mathcal{U} \times \mathbb{R}^m \colon \|\beta - \beta(u)\| \leq r_n \right\} \tag{D.58}$$

and the following error terms:

$$\epsilon_0(m, n) := \sup_{u \in \mathcal{U}} \left\| \mathbb{G}_n(\psi_i(\beta(u), u)) \right\|,$$

$$\epsilon_1(m, n) := \sup_{(u, \beta) \in R_{n,m}} \left\| \mathbb{G}_n(\psi_i(\beta, u)) - \mathbb{G}_n(\psi_i(\beta(u), u)) \right\|, \tag{D.59}$$

$$\epsilon_2(m, n) := \sup_{(u, \beta) \in R_{n,m}} n^{1/2} \left\| E[\psi_i(\beta, u)] - E[\psi_i(\beta(u), u)] - J(u)(\beta - \beta(u)) \right\|,$$

for the matrix $J(u)$ defined in (2.10). In what follows, we say that the data are in general position if for any $\gamma \in \mathbb{R}^m$,

$$P\left(Y_i = Z_i'\gamma, \text{for at least one } i \right) = 0, \tag{D.60}$$

which holds under Condition S. Finally, we emphasize that although we formally work with an i.i.d. setting in this paper, most of the results in this appendix do not rely on this assumption and can be extended to more general settings.

D.1. **Proof of Theorem 1.** The following technical lemma will be used in the proof of Theorem 1.

Lemma 3 (Rates in Euclidian Norm for Perturbed QR Process). *Suppose that for all $u \in \mathcal{U}$, $\widehat{\beta}(u)$ is a minimizer of*

$$\mathbb{E}_n[\rho_u(Y_i - Z_i'\beta)] + \mathcal{A}_n(u)'\beta, \tag{D.61}$$

where $\mathcal{A}_n(u)$ is the perturbation term, so that the QR-series coefficient $\widehat{\beta}(u)$ corresponds to the unperturbed case $\mathcal{A}_n(u) = 0$. Also, suppose that Condition S holds. Moreover, suppose that there exists a sequence of scalars $(\eta_n)_{n \geq 1}$ such that $\sup_{u \in \mathcal{U}} \|\mathcal{A}_n(u)\| \lesssim_P \eta_n$ and for any constant $B > 0$, defining $R_{n,m}$, $\epsilon_0(m, n)$, $\epsilon_1(m, n)$, and $\epsilon_2(m, n)$ above with $r_n = B\eta_n$ gives

R1. $\epsilon_0(m, n) \lesssim_P \sqrt{n}\eta_n$,

R2. $\epsilon_1(m,n) = o_P(\sqrt{n}\eta_n)$,

R3. $\epsilon_2(m,n) = o_P(\sqrt{n}\eta_n)$.

Then

$$\sup_{u\in\mathcal{U}} \left\| \widehat{\beta}(u) - \beta(u) \right\| \lesssim_P \eta_n. \tag{D.62}$$

Proof of Lemma 3. Due to convexity of the objective function (D.61), it suffices to show that for any $\varepsilon > 0$, there exists $B < \infty$ such that

$$P\left(\inf_{u\in\mathcal{U}} \inf_{\alpha\in S^{m-1}} \alpha' \left[\mathbb{E}_n \left[\psi_i \left(\beta, u \right) \right] + \mathcal{A}_n(u) \right] |_{\beta=\beta(u)+B\eta_n\alpha} > 0 \right) \geq 1 - \varepsilon \tag{D.63}$$

for all sufficiently large n since $\mathbb{E}_n \left[\psi_i \left(\beta, u \right) \right] + \mathcal{A}_n(u)$ is a sub-gradient of the objective function at β. To show (D.63), let B be some large constant to be chosen later and observe that uniformly in $u \in \mathcal{U}$ and $\alpha \in S^{m-1}$,

$$\sqrt{n}\alpha'\mathbb{E}_n \left[\psi_i \left(\beta(u) + B\eta_n\alpha, u \right) \right]$$
$$\geq \mathbb{G}_n(\alpha'\psi_i(\beta(u), u)) + \sqrt{n}\alpha' J(u)\alpha B\eta_n - \epsilon_1(m,n) - \epsilon_2(m,n),$$

by setting $r_n = B\eta_n$ and observing that $E\left[\psi_i(\beta(u), u)\right] = 0$ by definition of $\beta(u)$ (see the argument in the proof of Lemma 4 below). Further, we have uniformly in $u \in \mathcal{U}$ and $\alpha \in S^{m-1}$ that

$$|\mathbb{G}_n(\alpha'\psi_i(\beta(u), u))| \leq \sup_{u\in\mathcal{U}} \|\mathbb{G}_n(\psi_i(\beta(u), u))\| = \epsilon_0(m,n).$$

Also, it follows from Condition S that all eigenvalues of the matrix $J(u)$ are bounded below from zero uniformly over $u \in \mathcal{U}$, so that

$$\inf_{u\in\mathcal{U}} \inf_{\alpha\in S^{m-1}} \alpha' J(u)\alpha \geq c$$

for some constant $c > 0$. Thus, the event of interest in (D.63) is implied by the event

$$\left\{ \sqrt{n}cB\eta_n - \epsilon_0(m,n) - \epsilon_1(m,n) - \epsilon_2(m,n) - \sqrt{n}\sup_{u\in\mathcal{U}} \|\mathcal{A}_n(u)\| > 0 \right\},$$

and the probability of this event can be made arbitrarily close to one for all sufficiently large n by setting B sufficiently large since (i) $\epsilon_0(m,n) \lesssim_P \sqrt{n}\eta_n$ by R1 and $\epsilon_0(m,n)$ does not depend on B, (ii) $\sup_{u\in\mathcal{U}} \|\mathcal{A}_n(u)\| \lesssim_P \eta_n$ by assumptions of the lemma and $\sup_{u\in\mathcal{U}} \|\mathcal{A}_n(u)\|$ does not depend on B, and (iii) $\epsilon_1(m,n) + \epsilon_2(m,n) = o_P(\sqrt{n}\eta_n)$ by R2 and R3. This completes the proof of the lemma. $\qquad \square$

58

Proof of Theorem 1. The result follows from Lemma 3 with $\mathcal{A}_n(u) = 0$ and $\eta_n = \sqrt{m/n}$ provided we can show that if for any constant $B > 0$, we define $\epsilon_0(m,n)$, $\epsilon_1(m,n)$, and $\epsilon_2(m,n)$ with $r_n = B\sqrt{m/n}$, then

$$\epsilon_0(m,n) \lesssim_P \sqrt{m} \quad \text{and} \quad \epsilon_1(m,n) + \epsilon_2(m,n) = o_P(\sqrt{m}), \tag{D.64}$$

In turn, to show (D.64), note that by Lemma 32,

$$\epsilon_0(m,n) \lesssim_P \sqrt{m}\left(1 + \sqrt{m^{-\kappa}\log n} + \sqrt{m}\zeta_m \log n/\sqrt{n}\right) \lesssim \sqrt{m}$$

since $m^{-\kappa}\log n = o(1)$ and $m\zeta_m^2 \log^2 n = o(n)$. Moreover, by Lemma 33,

$$\epsilon_1(m,n) \lesssim_P \sqrt{m\zeta_m r_n \log n} + m\zeta_m \log n/\sqrt{n} = o(\sqrt{m}),$$

where we again used $m\zeta_m^2 \log^2 n = o(n)$, and

$$\epsilon_2(m,n) \lesssim \sqrt{n}\zeta_m r_n^2 + \sqrt{n}m^{-\kappa}r_n = o(\sqrt{m})$$

since $m\zeta_m^2 = o(n)$ and $m \to \infty$. This completes the proof of the theorem. $\qquad\square$

D.2. **Proof of Corollary 1.** We have uniformly over $u \in \mathcal{U}$ that

$$\|\widehat{Q}(u,\cdot) - Q(u,\cdot)\|_{L^2(X)} \le \|Z(\cdot)'(\widehat{\beta}(u) - \beta(u))\|_{L^2(X)} + \|R(u,\cdot)\|_{L^2(X)}$$
$$\lesssim \|\widehat{\beta}(u) - \beta(u)\| + \|R(u,\cdot)\|_{L^2(X)},$$

where the first line follows from the triangle inequality and the second from Condition S.3. Further, it follows from Lemma 1 that $\sup_{u \in \mathcal{U}}\|R(u,\cdot)\|_{L^2(X)} \lesssim m^{-s/d}$ in the case of polynomials and $\sup_{u \in \mathcal{U}}\|R(u,\cdot)\|_{L^2(X)} \lesssim m^{-(s \wedge s_0)/d}$ in the case of B-splines (note that in the case of polynomials, an application of Lemma 1 requires the condition that $s > d$ but this condition follows from the assumption that $m^{1-s/d}\log n = o(1)$ imposed in the corollary).

To bound $\sup_{u \in \mathcal{U}}\|\widehat{\beta}(u) - \beta(u)\|$, we apply Theorem 1. Lemma 1 implies that Condition S.4 holds with $\kappa = s/d - 1$ in the case of polynomials and with $\kappa = (s \wedge s_0)/d$ in the case of B-splines. Also, we have $\zeta_m \lesssim m$ in the case of polynomials and $\zeta_m \lesssim m^{1/2}$ in the case of B-splines. Thus, in both cases, the conditions that $\zeta_m^2 m \log^2 n = o(n)$ and $m^{-\kappa}\log n = o(1)$ required in Theorem 1 follow from the assumptions of the corollary. Therefore, an application of Theorem 1 gives

$$\sup_{u \in \mathcal{U}}\|\widehat{\beta}(u) - \beta(u)\| \lesssim_P \sqrt{m/n}$$

in both cases. Combining presented bounds gives both claims of the corollary. $\qquad\square$

D.3. **Proof of Theorem 2.** For $u \in \mathcal{U}$, consider the function in (D.61), where $\mathcal{A}_n(u)$ is the perturbation term, and let $\widehat{\beta}(u)$ be a minimizer of this function, so that the QR-series coefficient $\widehat{\beta}(u)$ corresponds to the unperturbed case $\mathcal{A}_n(u) = 0$. Define the following approximation error:

$$\epsilon_3(m,n) := \sup_{u \in \mathcal{U}} n^{1/2} \left\| \mathbb{E}_n[\psi_i(\widehat{\beta}(u), u)] + \mathcal{A}_n(u) \right\|. \tag{D.65}$$

We have the following lemma.

Lemma 4 (Uniform Linear Approximation). *Consider the setting specified above. Suppose that the data are in general position so that (D.60) holds, and further that the conditions of Lemma 3 hold for some sequence of scalars $(\eta_n)_{n \geq 1}$. Then*

$$\sqrt{n} J(u) \left(\widehat{\beta}(u) - \beta(u) \right) = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_i(\beta(u), u) - \mathcal{A}_n(u) + r_n(u), \tag{D.66}$$

where $r_n(u)$ is such that for any $\varepsilon \in (0,1)$,

$$\sup_{u \in \mathcal{U}} \|r_n(u)\| \leq \epsilon_1(m,n) + \epsilon_2(m,n) + \epsilon_3(m,n), \tag{D.67}$$

with probability at least $1 - \varepsilon$, where $\epsilon_1(m,n)$ and $\epsilon_2(m,n)$ are defined in (D.59) using $R_{n,m}$ in (D.58) with $r_n = B\eta_n$ for some sufficiently large constant $B = B(\varepsilon)$.

Proof of Lemma 4. First, note that $E\left[\psi_i(\beta(u), u)\right] = 0$ by definition of $\beta(u)$. Indeed, despite the possible approximation error, $\beta(u)$ minimizes $E[\rho_u(Y - Z'\beta)]$ so that $E\left[\psi_i(\beta(u), u)\right] = 0$ by the first order conditions. Therefore, equation (D.66) can be recast as

$$r_n(u) = n^{1/2} J(u)(\widehat{\beta}(u) - \beta(u)) + \mathbb{G}_n \left(\psi_i(\beta(u), u) \right) + \mathcal{A}_n(u).$$

Second, by Lemma 3, for any $\varepsilon \in (0,1)$, there is a constant $B = B(\varepsilon)$ such that with probability $1 - \varepsilon$, we have $(u, \widehat{\beta}(u)) \in R_{n,m}$ for all $u \in \mathcal{U}$ and all n, where $R_{n,m}$ is defined in (D.58) with $r_n = B\eta_n$. Therefore, we have by the triangle inequality that for all $u \in \mathcal{U}$,

$$\begin{aligned}
\|r_n(u)\| &\leq \left\| \mathbb{G}_n(\psi_i(\beta(u), u)) - \mathbb{G}_n(\psi_i(\widehat{\beta}(u), u)) \right\| \\
&\quad + n^{1/2} \left\| E\left[\psi_i(\widehat{\beta}(u), u)\right] - E\left[\psi_i(\beta(u), u)\right] - J(u)\left(\widehat{\beta}(u) - \beta(u)\right) \right\| \\
&\quad + n^{1/2} \left\| \mathbb{E}_n\left[\psi_i(\widehat{\beta}(u), u)\right] + \mathcal{A}_n(u) \right\| \\
&\leq \epsilon_1(m,n) + \epsilon_2(m,n) + \epsilon_3(m,n)
\end{aligned}$$

by the definitions of $\epsilon_1(m,n)$, $\epsilon_2(m,n)$, and $\epsilon_3(m,n)$. The asserted claim follows. $\qquad \square$

Proof of Theorem 2. To prove the asserted claim, we apply Lemma 4 with $\mathcal{A}_n(\cdot) = 0$. Since conditions $m^3\zeta_m^2 = o(n^{1-\varepsilon})$ and $m^{-\kappa+1} = o(n^{-\varepsilon})$, which are assumed in Theorem 2, imply conditions $m\zeta_m^2 \log^2 n = o(n)$ and $m^{-\kappa} \log n = o(1)$, which are assumed in Theorem 1, it follows that conditions of Lemma 3 hold with $\eta_n = \sqrt{m/n}$ by the same argument as that used in the proof of Theorem 1. Thus, since under Condition S, the data are in general position, the conclusion of Lemma 4 holds, and so for any $\varphi \in (0,1)$, there exists $B > 0$ such that (D.67) holds with probability at least $1 - \varphi$, where $\epsilon_1(m,n)$ and $\epsilon_2(m,n)$ are defined in (D.59) using $R_{n,m}$ in (D.58) with $r_n = B\sqrt{m/n}$.

Next, we control $\epsilon_1(m,n)$, $\epsilon_2(m,n)$, and $\epsilon_3(m,n)$ for given r_n. By Lemma 33,

$$\epsilon_1(m,n) \lesssim_P \sqrt{m\zeta_m r_n \log n} + \frac{m\zeta_m \log n}{\sqrt{n}} \lesssim \frac{m^{3/4}\zeta_m^{1/2} \log^{1/2} n}{n^{1/4}},$$

where the second inequality holds since $m^3\zeta_m^2 = o(n^{1-\varepsilon})$. Also by Lemma 33,

$$\epsilon_2(m,n) \lesssim_P \sqrt{n}\zeta_m r^2 + \sqrt{n}m^{-\kappa}r_n \lesssim \frac{m\zeta_m}{\sqrt{n}} + m^{1/2-\kappa}.$$

Further, by Lemma 34, we have with probability one that

$$\epsilon_3(m,n) \leq \frac{m\zeta_m}{\sqrt{n}}.$$

Combining the inequalities above and using (D.66) and (D.67) shows that,

$$\sup_{u\in\mathcal{U}} \left\| \sqrt{n}J(u)\big(\widehat{\beta}(u) - \beta(u)\big) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\beta(u),u) \right\| \lesssim_P \frac{m^{3/4}\zeta_m^{1/2} \log^{1/2} n}{n^{1/4}} + m^{1/2-\kappa}.$$

Finally, observe that

$$\widetilde{r}_u := \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i\Big(1\{Y_i \leq Q(u,X_i)\} - 1\{Y_i \leq Z_i'\beta(u)\}\Big), \quad u \in \mathcal{U},$$

satisfies

$$\sup_{u\in\mathcal{U}} \|\widetilde{r}_u\| \lesssim_P \sqrt{m^{1-\kappa}\log n} + \frac{m\zeta_m \log n}{\sqrt{n}}$$

by Lemma 31 and note that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\beta(u),u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(1\{Y_i \leq Z_i'\beta(u)\} - u), \quad u \in \mathcal{U}$$

and

$$\mathbb{U}(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(u - 1\{Y_i \leq Q(u,X_i)\}), \quad u \in \mathcal{U}.$$

The asserted claims now follow by noting that all eigenvalues of the matrix $J(u)$ are bounded below from zero uniformly over $u \in \mathcal{U}$. $\qquad\square$

D.4. Proof of Corollary 2. By the triangle inequality, we have for all $u \in \mathcal{U}$ and $x \in \mathcal{X}$ that

$$|\widehat{Q}(u,x) - Q(u,x)| \leq |Z(x)'(\widehat{\beta}(u) - \beta(u))| + |R(u,x)|. \tag{D.68}$$

By Lemma 1,

$$\sup_{u \in \mathcal{U}} \sup_{x \in \mathcal{X}} |R(u,x)| \lesssim m^{1-s/d} \tag{D.69}$$

in the case of polynomials and

$$\sup_{u \in \mathcal{U}} \sup_{x \in \mathcal{X}} |R(u,x)| \lesssim m^{-(s \wedge s_0)/d} \tag{D.70}$$

in the case of B-splines.

To bound $\sup_{u \in \mathcal{U}} \sup_{x \in \mathcal{X}} |Z(x)'(\widehat{\beta}(u) - \beta(u))|$, we apply Theorem 2. Note that by Lemma 1, Condition S.4 is satisfied with $\kappa = s/d - 1$ in the case of polynomials and with $\kappa = (s \wedge s_0)/d$ in the case of B-splines. Also, we have $\zeta_m \lesssim m$ in the case of polynomials and $\zeta_m \lesssim m^{1/2}$ in the case of B-splines, so that the conditions $m^3 \zeta_m^2 = o(n^{1-\varepsilon})$ and $m^{-\kappa+1} = o(n^{-\varepsilon})$, required in Theorem 2, are satisfied in both cases. Therefore, an application of Theorem 2 gives

$$Z(x)'(\widehat{\beta}(u) - \beta(u)) = \frac{1}{\sqrt{n}} Z(x)' J^{-1}(u) \mathbb{U}(u) + n^{-1/2} \zeta_m \|r(u)\|$$

uniformly over $u \in \mathcal{U}$ and $x \in \mathcal{X}$ in both cases where we used $|Z(x)'r(u)| \leq \|Z(x)\| \|r(u)\| \leq \zeta_m \|r(u)\|$. Note that

$$E[Z(x)' J^{-1}(u) \mathbb{U}(u)] = 0$$

and

$$E[(Z(x)' J^{-1}(u) \mathbb{U}(u))^2] \leq |Z(x)' J^{-1}(u) \Sigma J^{-1}(u) Z(x)| \lesssim \|Z(x)\|^2 \leq \zeta_m^2$$

uniformly over $u \in \mathcal{U}$ and $x \in \mathcal{X}$ by Conditions S.2 and S.3. Therefore, the argument like that used in the proof of Theorem 10 below shows that

$$|Z(x)' J^{-1}(u) \mathbb{U}(u)| \lesssim_P \zeta_m \sqrt{\log n}$$

uniformly over $u \in \mathcal{U}$ and $x \in \mathcal{X}$. Substituting the bound $\zeta_m \lesssim m$ in the case of polynomials and $\zeta_m \lesssim m^{1/2}$ in the case of B-splines shows that uniformly over $u \in \mathcal{U}$ and $x \in \mathcal{X}$, under the conditions corresponding to each case we have $\sup_{u \in \mathcal{U}} \|r(u)\| = o_P(1)$, so that

$$|Z(x)'(\widehat{\beta}(u) - \beta(u))| \lesssim_P \sqrt{m^2 \log n / n}$$

in the former case and

$$|Z(x)'(\widehat{\beta}(u) - \beta(u))| \lesssim_P \sqrt{m \log n / n}$$

in the latter case. Combining these bounds with those in (D.68), (D.69), and (D.70) gives the asserted claim. □

D.5. Proof of Theorems 3, 4, 5, 6, and 7.

Proof of Theorem 3. Note that

$$\sup_{u\in\mathcal{U}} \|r(u)\| \le \sup_{u\in\mathcal{U}} \left\| (\widehat{J}^{-1}(u) - J^{-1}(u))\mathbb{U}^*(u) \right\| \le \sup_{u\in\mathcal{U}} \left\| \widehat{J}^{-1}(u) - J^{-1}(u) \right\| \cdot \sup_{u\in\mathcal{U}} \|\mathbb{U}^*(u)\|.$$

In addition, by Lemma 32, we have $\sup_{u\in\mathcal{U}} \|\mathbb{U}(u)\| \lesssim_P \sqrt{m}$, and so

$$\sup_{u\in\mathcal{U}} \|\mathbb{U}^*(u)\| \lesssim_P \sqrt{m} \tag{D.71}$$

since $\mathbb{U}^*(\cdot)$ is a copy of $\mathbb{U}(\cdot)$ conditional on $(Z_i)_{i=1}^n$. Next, we bound $\sup_{u\in\mathcal{U}} \|\widehat{J}^{-1}(u) - J^{-1}(u)\|$. Note that conditions of Theorem 3 imply conditions of Lemma 30, and so

$$\sup_{u\in\mathcal{U}} \|\widehat{J}(u) - J(u)\| \lesssim_P \sqrt{\frac{\varsigma_m^2 m \log n}{nh}} + m^{-\kappa} + h = o(n^{-\varepsilon'}/\sqrt{m})$$

for some $\varepsilon' > 0$. Hence, with probability approaching one, all eigenvalues of $\widehat{J}(u)$ are bounded below from zero uniformly over $u \in \mathcal{U}$, and so the matrix identity $A^{-1} - B^{-1} = B^{-1}(B - A)A^{-1}$ implies that

$$\|J^{-1}(u) - \widehat{J}^{-1}(u)\| = \|\widehat{J}^{-1}(u)\| \cdot \|\widehat{J}(u) - J(u)\| \cdot \|J^{-1}(u)\| = o_P(n^{-\varepsilon'}/\sqrt{m})$$

uniformly over $u \in \mathcal{U}$. Combining this bound with (D.71) gives the first asserted claim.

Note also that the results continue to hold in P-probability if we replace P by P^*, which is the second asserted claim, since if a random variable $B_n = O_P(1)$, then $B_n = O_{P^*}(1)$. Indeed, the first relation means that $P(|B_n| > \ell_n) = o(1)$ for any $\ell_n \to \infty$, while the second means that $P^*(|B_n| > \ell_n) = o_P(1)$ for any $\ell_n \to \infty$. But the second follows from the first from the Markov inequality, observing that $E[P^*(|B_n| > \ell_n)] = P(|B_n| > \ell_n) = o(1)$. This completes the proof of the theorem. □

Proof of Theorem 4. The proof is similar to the proof of Theorem 2 but it applies Lemma 4 twice, one to the unperturbed problem and one to the perturbed problem with $\mathcal{A}_n(u) = -\mathbb{U}^*(u)/\sqrt{n}$, for every $u \in \mathcal{U}$.

Since conditions of Theorem 4 imply conditions of Theorem 1, we have by Theorem 1 that $\sup_{u\in\mathcal{U}} \|\widehat{\beta}(u) - \beta(u)\| \lesssim_P \eta_n = \sqrt{m/n}$. Similarly, using Lemma 3 like in the proof of Theorem 1, we obtain $\sup_{u\in\mathcal{U}} \|\widehat{\beta}^*(u) - \beta(u)\| \lesssim_P \eta_n$, since

$$\{\mathbb{U}^*(u)\}_{u\in\mathcal{U}} =_d \{\mathbb{U}(u)\}_{u\in\mathcal{U}} \lesssim_P \sqrt{m}$$

by Lemma 32. Then, by applying Lemma 4 twice, we have for all $u \in \mathcal{U}$ that

$$\sqrt{n} J(u) \left(\widehat{\beta}^*(u) - \widehat{\beta}(u) \right) = \sqrt{n} J(u) \left(\widehat{\beta}^*(u) - \beta(u) \right) - \sqrt{n} J(u) \left(\widehat{\beta}(u) - \beta(u) \right)$$

$$= \mathbb{U}^*(u) + r_n^{pert}(u) - r_n^{unpert}(u),$$

where $r_n^{pert}(u)$ and $r_n^{unpert}(u)$ are defined by (D.66) with $\mathcal{A}_n(u) = -\mathbb{U}^*(u)/\sqrt{n}$ and $\mathcal{A}_n(u) = 0$, respectively, and the same arguments as those used in the proof of Theorem 2 show that

$$\sup_{u \in \mathcal{U}} \left(\|r_n^{pert}(u)\| + \|r_n^{unpert}(u)\| \right) \lesssim_P \sqrt{m \zeta_m \eta_n \log n} + \frac{m \zeta_m \log n}{\sqrt{n}} + m^{-\kappa} \sqrt{m}.$$

The first asserted claim now follows by substituting $\eta_n = \sqrt{m/n}$ and using the growth conditions $m^3 \zeta_m^2 = o(n^{1-\varepsilon})$ and $m^{-\kappa+1/2} = o(n^{-\varepsilon})$. The second asserted claim follows from the same argument as that used in the proof of Theorem 3. This completes the proof of the theorem. $\qquad \square$

Proof of Theorem 5. The proof relies on the following lemma:

Lemma 5. *Suppose that Condition S holds. In addition, suppose that $m^7 \zeta_m^6 = o(n^{1-\varepsilon})$ for some constant $\varepsilon > 0$. Then there exists a process $G = G_n$ such that*

$$\sup_{u \in \mathcal{U}} \|\mathbb{U}(u) - G(u)\| \lesssim_P o(n^{-\varepsilon'}) \tag{D.72}$$

and the process G is conditionally on $(Z_i)_{i=1}^n$ zero-mean Gaussian with a.s. continuous sample paths and the covariance function

$$E\left[G(u_1) G(u_2)' \mid (Z_i)_{i=1}^n \right] = \mathbb{E}_n[Z_i Z_i'](u_1 \wedge u_2 - u_1 u_2), \text{ for all } u_1 \text{ and } u_2 \text{ in } \mathcal{U}.$$

Proof of Lemma 5. Since under our conditions, we have $\zeta_m^2 \log n / n = o(1)$ and $\lambda_{\max} = \sup_{\alpha \in S^{m-1}} E[(Z'\alpha)^2] \lesssim 1$, it follows from Corollary 4 in Appendix H that there exists a constant $C > 0$ such that $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \leq C$ with probability approaching one. Thus, the asserted claim of the lemma follows from applying Lemma 14 conditional on $(Z_i)_{i=1}^n$ on the event $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \leq C$. $\qquad \square$

Getting back to the proof of Theorem 5, let G be a process constructed in Lemma 5. Then

$$\sup_{u \in \mathcal{U}} \|\sqrt{n}(\widehat{\beta}(u) - \beta(u)) - J^{-1}(u) G(u)\|$$

$$\leq \sup_{u \in \mathcal{U}} \|J^{-1}(u) \mathbb{U}(u) - J^{-1}(u) G(u)\| + o_P(n^{-\varepsilon'})$$

$$\leq \sup_{u \in \mathcal{U}} \|J^{-1}(u)\| \cdot \sup_{u \in \mathcal{U}} \|\mathbb{U}(u) - G(u)\| + o_P(n^{-\varepsilon'}) = o_P(n^{-\varepsilon'}),$$

where in the second line we invoked Theorem 2 and in the third line we used (D.72) and the fact that all eigenvalues of $J(u)$ are bounded below from zero uniformly over $u \in \mathcal{U}$. The asserted claim follows. □

Proof of Theorem 6. The proof relies on the following lemma:

Lemma 6. *Suppose that Condition S holds. In addition, suppose that $\zeta_m^2 \log n = o(n)$. Then*

$$\sup_{u \in \mathcal{U}} \|G^*(u)\| \lesssim_P \sqrt{m}.$$

Proof of Lemma 6. Recall that the process $G^*(\cdot)$ is given by

$$G^*(u) = \widehat{\Sigma}^{1/2} B_m(u), \quad u \in \mathcal{U},$$

where $B_m(\cdot)$ is an m-dimensional vector of independent Brownian bridges $B_{m,j}(\cdot)$, $j = 1, \ldots, m$. In addition,

$$
\begin{aligned}
E\left[\sup_{u \in \mathcal{U}} \|B_m(u)\|\right] &= E\left[\sup_{u \in \mathcal{U}} \left(\sum_{j=1}^m B_{m,j}(u)^2\right)^{1/2}\right] \\
&\leq E\left[\left(\sum_{j=1}^m \sup_{u \in \mathcal{U}} B_{m,j}(u)^2\right)^{1/2}\right] \\
&\leq \left(\sum_{i=1}^m E\left[\sup_{u \in \mathcal{U}} B_{m,j}(u)^2\right]\right)^{1/2} \lesssim \sqrt{m}.
\end{aligned}
$$

Moreover,

$$\|\widehat{\Sigma}^{1/2}\| = \|\widehat{\Sigma}\|^{1/2} \leq \left(\|\Sigma\| + \|\widehat{\Sigma} - \Sigma\|\right)^{1/2} \lesssim_P 1,$$

by Condition S and Lemma 30. Hence,

$$\sup_{u \in \mathcal{U}} \|G^*(u)\| = \sup_{u \in \mathcal{U}} \left\|\widehat{\Sigma}^{1/2} B_m(u)\right\| \lesssim_P \sup_{u \in \mathcal{U}} \|B_m(u)\| \lesssim_P \sqrt{m}.$$

This completes the proof of the lemma. □

Getting back to the proof of Theorem 6, note that by Lemma 30,

$$\sup_{u \in \mathcal{U}} \|J(u) - \widehat{J}(u)\| \lesssim_P \sqrt{\frac{\zeta_m^2 m \log n}{nh}} + m^{-\kappa} + h = o(n^{-\varepsilon'}/\sqrt{m}),$$

and so

$$\sup_{u \in \mathcal{U}} \|J^{-1}(u) - \widehat{J}^{-1}(u)\| \lesssim_P o(n^{-\varepsilon'}/\sqrt{m})$$

like in the proof of Theorem 3. The first asserted claim follows from combining this inequality with the bound

$$\sup_{u \in \mathcal{U}} \|r(u)\| = \sup_{u \in \mathcal{U}} \|\widehat{J}^{-1}(u)G^*(u) - J^{-1}(u)G^*(u)\|$$

$$\leq \sup_{u \in \mathcal{U}} \|\widehat{J}^{-1}(u) - J^{-1}(u)\| \cdot \sup_{u \in \mathcal{U}} \|G^*(u)\|$$

and using Lemma 6. The second asserted claim follows from the same argument as that used in the proof of Theorem 3. This completes the proof of theorem. $\qquad\square$

Proof of Theorem 7. The proof relies on the following lemma:

Lemma 7. *Suppose that Condition S holds. In addition, suppose that $m^3 \zeta_m^2 = o(n^{1-\varepsilon})$ and $m^{-\kappa+1} = o(n^{-\varepsilon})$ for some constant $\varepsilon > 0$. Moreover, suppose that the random variable π is non-negative and satisfies $E[\pi] = 1$ and $E[\pi^4] \lesssim 1$. Finally, suppose that $\max_{1 \leq i \leq n} \pi_i \lesssim_P \log n$. Then*

$$\sqrt{n} \left(\widehat{\beta}^b(u) - \widehat{\beta}(u) \right) = \frac{J^{-1}(u)}{\sqrt{n}} \sum_{i=1}^{n} (\pi_i - 1) Z_i(u - 1\{U_i \leq u\}) + r(u),$$

where

$$\sup_{u \in \mathcal{U}} \|r(u)\| \lesssim_P \frac{m^{3/4} \zeta_m^{1/2} \log n}{n^{1/4}} + \sqrt{m^{1-\kappa} \log n} = o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$.

Proof of Lemma 7. First, note that it follows from the first-order conditions of the optimization problem (3.22) that $\widehat{\beta}^b(u)$ solves the quantile regression problem (2.4) for the rescaled data $(\pi_i Y_i, \pi_i Z_i)_{i=1}^n$. Second, recall that the non-negative random variable π is such that $E[\pi] = 1$ and $E[\pi^4] \lesssim 1$ and the sequence $(\pi_i)_{i=1}^n$, which consists of i.i.d. random variables with the distribution of π, is independent of the data and satisfies $\max_{1 \leq i \leq n} \pi_i \lesssim_P \log n$. Using these observations, we can follow the proof of Theorem 2 with $(Z_i, Y_i)_{i=1}^n$ replaced by $(\pi_i Z_i, \pi_i Y_i)_{i=1}^n$, so that

$$\psi_i(\beta, u) = \pi_i Z_i \left(1\{\pi_i Y_i \leq \pi_i Z_i'\beta\} - u \right) = \pi_i Z_i \left(1\{Y_i \leq Z_i'\beta\} - u \right),$$

but keeping the same vectors $\beta(u)$ and matrices $J(u)$ to show that

$$\sqrt{n} \left(\widehat{\beta}^b(u) - \beta(u) \right) = \frac{J^{-1}(u)}{\sqrt{n}} \sum_{i=1}^{n} \pi_i Z_i \left(u - 1\{Y_i \leq Z_i'\beta(u)\} \right) + r^b(u),$$

where

$$\sup_{u \in \mathcal{U}} \|r^b(u)\| \lesssim_P \frac{m^{3/4} \zeta_m^{1/2} \log n}{n^{1/4}} + \sqrt{m^{1-\kappa} \log n} = o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$, where in the proof we replace all applications of the third maximal inequality in Lemma 22 by applications of the second maximal inequality in the same lemma, and we also replace ζ_m by $\zeta_m \log n$, so that $\max_{1 \leq i \leq n} \pi_i \|Z_i\| \lesssim_P \zeta_m \log n$. The asserted claim of the lemma follows by combining this result with Theorem 2. $\qquad \square$

Getting back to the proof of Theorem 7, we apply Lemma 14 with $v_i = \pi_i - 1$, so that $E[v_i] = 0$, $E[v_i^2] = 1$, $E[|v_i|^4] \lesssim 1$, and $\max_{1 \leq i \leq n} |v_i| \lesssim_P \log n$. The lemma implies that there is a Gaussian process $G^*(\cdot) = G_n^*(\cdot)$ with the covariance structure $\mathbb{E}_n[Z_i Z_i'](u \wedge u' - uu')$ such that

$$\sup_{u \in \mathcal{U}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n (\pi_i - 1) Z_i \Big(u - 1\{U_i \leq u\} \Big) - G^*(u) \right\| \lesssim_P o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$. Combining this result with Lemma 7 gives the first asserted claim. The second asserted claim follows from the same argument as that used in the proof of Theorem 3. This completes the proof of theorem. $\qquad \square$

APPENDIX E. PROOFS OF THEOREMS 8 AND 9

Proof of Theorem 8. By Theorem 2 and Condition P,

$$|\widehat{\theta}(u, w) - \theta(u, w)| \leq |\ell(w)'(\widehat{\beta}(u) - \beta(u))| + |r(u, w)|$$
$$\leq \frac{|\ell(w)'J^{-1}(u)\mathbb{U}(u)|}{\sqrt{n}} + o_P\Big(\frac{\|\ell(w)\|}{\sqrt{n}}\Big) + o\Big(\frac{\|\ell(w)\|}{\sqrt{n}}\Big).$$

In addition,

$$E\Big[|\ell(w)'J^{-1}(u)\mathbb{U}(u)|^2\Big] \lesssim \|\ell(w)\|^2 \|J^{-1}(u)\|^2 \sup_{\alpha \in S^{m-1}} E[(\alpha'Z)^2] \lesssim \|\ell(w)\|^2$$

by Condition S. Combining these bounds gives the asserted claim. $\qquad \square$

Proof of Theorem 9. By Theorem 2 and Condition P,

$$t(u, w) = \frac{\ell(w)'(\widehat{\beta}(u) - \beta(u))}{\widehat{\sigma}(u, w)} - \frac{r(u, w)}{\widehat{\sigma}(u, w)}$$
$$= \frac{\ell(w)'J^{-1}(u)\mathbb{U}(u)}{\sqrt{n}\widehat{\sigma}(u, w)} + o_P\Big(\frac{\|\ell(w)\|}{\sqrt{n}\widehat{\sigma}(u, w)}\Big). \qquad (\text{E.73})$$

In addition,

$$\widehat{\sigma}(u, w) = (1 + o_P(1))\sigma(u, w)$$

by Lemma 30 and Condition S. Moreover, $\sigma(u, w) \gtrsim \|\ell(w)\|/\sqrt{n}$ under Condition S, and so the second term in (E.73) is $o_P(1)$. Hence, the asserted claim follows provided we can show that

$$\frac{\ell(w)'J^{-1}(u)\mathbb{U}(u)}{\sqrt{n}\sigma(u, w)} \to_d N(0, 1).$$

However, this follows from the Lindeberg-Feller central limit theorem (see Theorem 9.6.1 in Dudley [37]) since for any $\epsilon > 0$, there exists a constant C such that

$$E\left[\frac{(\ell(w)'J^{-1}(u)Z(X))^2}{n\sigma^2(u, w)}1\left\{\frac{|\ell(w)'J^{-1}(u)Z(X)|}{n\sigma(u, w)} > \epsilon\right\}\right]$$
$$\leq E\left[\frac{(\ell(w)'J^{-1}(u)Z(X))^2}{n\sigma^2(u, w)}1\left\{\zeta_m > C\sqrt{n}\epsilon\right\}\right] \to 0$$

as $n \to \infty$ under our conditions. This completes the proof of the theorem. $\qquad\square$

APPENDIX F. PROOFS OF THEOREMS 10-15

Lemma 8 (Entropy Bound). *Suppose that Conditions S and U hold. Consider the class of functions*

$$\mathcal{L}_m = \left\{Z \mapsto (\ell(w)/\|\ell(w)\|)'J^{-1}(u)Z \colon (u, w) \in I\right\},$$

mapping $B_m(0, \zeta_m)$ into $\mathbb{R}$, and let $L = 1 \vee \sup_{f \in \mathcal{L}_m} |f|$ denote its envelope. Then the uniform entropy numbers of $\mathcal{L}_m$ satisfy

$$\sup_Q \log N(\epsilon\|L\|_{Q,2}, \mathcal{L}_m, L_2(Q)) \lesssim \log(n/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1.$$

Proof. For $w \in \mathcal{W}$, denote $\xi(w) = \ell(w)/\|\ell(w)\|$. By Condition S,

$$\sup_{u \in \mathcal{U}} \|J^{-1}(u)\| \lesssim 1$$

and by Lemma 19,

$$\|J(u) - J(\tilde{u})\| \lesssim |u - \tilde{u}|, \quad \text{uniformly over } u \text{ and } \tilde{u} \text{ in } I.$$

Hence, uniformly over (u, w) and $(\tilde{u}, \tilde{w})$ in I and $Z \in B(0, \zeta_m)$,

$$
\begin{aligned}
\left| \xi(w)' J^{-1}(u) Z - \xi(\tilde{w})' J^{-1}(\tilde{u}) Z \right| &= \left| (\xi(w) - \xi(\tilde{w}))' J^{-1}(u) Z + \xi(\tilde{w})' (J^{-1}(u) - J^{-1}(\tilde{u})) Z \right| \\
&= \left| (\xi(w) - \xi(\tilde{w}))' J^{-1}(u) Z \right| \\
&\quad + \left| \xi(\tilde{w})' J^{-1}(\tilde{u}) (J(\tilde{u}) - J(u)) J^{-1}(u) Z \right| \\
&\lesssim \zeta_{m,\theta}^L \|w - \tilde{w}\| \cdot \|J^{-1}(u)\| \cdot \|Z\| \\
&\quad + \|\xi(\tilde{w})\| \cdot \|J^{-1}(\tilde{u})\| \cdot |\tilde{u} - u| \cdot \|J^{-1}(u)\| \cdot \|Z\| \\
&\leq \zeta_m (1 + \zeta_{m,\theta}^L)(\|w - \tilde{w}\| + |\tilde{u} - u|)
\end{aligned}
$$

since $\|\xi(\tilde{w})\| = 1$. Hence, the asserted claim follows from a standard argument since $L \geq 1$ and the set I is such that its dimension is independent of n and its diameter is bounded from above uniformly over n. $\qquad\square$

Proof of Theorem 10. The proof relies on the following lemma:

Lemma 9. *Suppose that Conditions S and U hold. Then*

$$
\sup_{(u,w) \in I} \left| (\ell(w)/\|\ell(w)\|)' J^{-1}(u) \mathbb{U}(u) \right| \lesssim_P (\log n)^{1/2}.
$$

Proof of Lemma 9. Consider the class of functions

$$
\mathcal{L}_m = \left\{ (Z, U) \mapsto (\ell(w)/\|\ell(w)\|)' J^{-1}(u) Z : (u, w) \in I \right\},
$$

mapping $B_m(0, \zeta_m) \times [0, 1]$ into $\mathbb{R}$. The function $L = C\zeta_m$ for sufficiently large constant C is its envelope and by Lemma 8, its uniform entropy numbers satisfy

$$
\sup_Q \log N(\epsilon \|L\|_{Q,2}, \mathcal{L}_m, L_2(Q)) \lesssim \log(n/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1.
$$

Also, consider the class of functions

$$
\mathcal{G} = \left\{ (Z, U) \mapsto 1\{U \leq u\} - u : u \in \mathcal{U} \right\},
$$

mapping $B_m(0, \zeta_m) \times [0, 1]$ into $\mathbb{R}$. The uniform entropy numbers of $\mathcal{G}$ satisfy

$$
\sup_Q \log N(\epsilon \|G\|_{Q,2}, \mathcal{G}, L_2(Q)) \lesssim \log(1/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1, \qquad \text{(F.74)}
$$

where $G(Z, U) = 1$ is its envelope. Hence, by 24, the uniform entropy numbers of the class of functions $\mathcal{L}_m \mathcal{G}$ satisfy

$$
\sup_Q \log N(\epsilon \|LG\|_{Q,2}, \mathcal{L}_m \mathcal{G}, L_2(Q)) \lesssim \log(n/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1.
$$

In addition, by Condition S,

$$E\left[\left((\ell(w)/\|\ell(w)\|)'J^{-1}(u)Z(X)(1\{U \le u\} - u)\right)^2\right]$$

$$= u(1-u)(\ell(w)/\|\ell(w)\|)'J^{-1}(u)\Sigma J^{-1}(u)(\ell(w)/\|\ell(w)\|) \lesssim 1$$

uniformly over $(u,w) \in I$, so that

$$\sup_{f \in \mathcal{L}_m\mathcal{G}} E[f^2] \lesssim 1.$$

Hence, applying the third maximal inequality of Lemma 22 gives

$$\sup_{(u,w)\in I} \left|(\ell(w)/\|\ell(w)\|)'J^{-1}(u)\mathbb{U}(u)\right| \lesssim_P \left(1 + \frac{\zeta_m^2 \log n}{n}\right)^{1/2} \log^{1/2} n \lesssim (\log n)^{1/2}. \quad \text{(F.75)}$$

This completes the proof of the lemma. $\qquad\square$

Getting back to the proof of the theorem, we have by the triangle inequality that

$$\sup_{(u,w)\in I} |\widehat{\theta}(u,w) - \theta(u,w)| \le \sup_{(u,w)\in I} |\ell(w)'(\widehat{\beta}(u) - \beta(u))| + \sup_{(u,w)\in I} |r(u,w)|,$$

where the second term satisfies

$$\sup_{(u,w)\in I} |r(u,w)| \le \zeta_{m,\theta} \sup_{(u,w)\in I} \frac{|r(u,w)|}{\|\ell(w)\|} = o\left(\frac{\zeta_{m,\theta}}{\sqrt{n}\log n}\right)$$

by Condition U. Also, by Theorem 2, the first term can be bounded uniformly over $(u,w) \in I$ as

$$|\ell(w)'(\widehat{\beta}(u) - \beta(u))| \lesssim_P \frac{|\ell(w)'J^{-1}(u)\mathbb{U}(u)|}{\sqrt{n}} + o_P\left(\frac{\zeta_{m,\theta}}{\sqrt{n}\log n}\right)$$

$$\le \left|(\ell(w)/\|\ell(w)\|)'J^{-1}(u)\mathbb{U}(u)\right| \cdot \sqrt{\frac{\zeta_{m,\theta}^2}{n}} + o_P\left(\frac{\zeta_{m,\theta}}{\sqrt{n}\log n}\right)$$

$$\lesssim_P \sqrt{\frac{\zeta_{m,\theta}^2 \log n}{n}} + o_P\left(\frac{\zeta_{m,\theta}}{\sqrt{n}\log n}\right),$$

where the second line follows from $\|\ell(w)\| \le \zeta_{m,\theta}$, holding by Condition U, and the third line from Lemma 9. The asserted claim follows. $\qquad\square$

Proof of Theorem 11. By Lemma 30,

$$\|\widehat{\Sigma} - \Sigma\| \lesssim_P \sqrt{\frac{\zeta_m^2 \log n}{n}} = o(n^{-\varepsilon'})$$

and

$$\sup_{u \in \mathcal{U}} \|\widehat{J}(u) - J(u)\| \lesssim_P \sqrt{\frac{m\zeta_m^2 \log n}{nh}} + m^{-\kappa} + h = o(n^{-\varepsilon'}).$$

Combining these inequalities with Condition S gives

$$\sup_{(u,w)\in I}\left|\frac{\widehat{\sigma}(u,w)}{\sigma(u,w)}-1\right|=o_P(n^{-\varepsilon'}). \tag{F.76}$$

Further, by Condition S,

$$\sigma(u,w)\gtrsim\|\ell(w)\|/\sqrt{n},\quad\text{uniformly over }(u,w)\in I, \tag{F.77}$$

and so

$$\widehat{\sigma}(u,w)\gtrsim_P\|\ell(w)\|/\sqrt{n},\quad\text{uniformly over }(u,w)\in I. \tag{F.78}$$

Next, by Condition U and (F.78), we have uniformly in $(u,w)\in I$ that

$$t(u,w)=\frac{\widehat{\theta}(u,w)-\theta(u,w)}{\widehat{\sigma}(u,w)}=\frac{\ell(w)'(\widehat{\beta}(u)-\beta(u))}{\widehat{\sigma}(u,w)}+o_P(n^{-\varepsilon'}). \tag{F.79}$$

Also, by Theorem 2 and (F.78), we have uniformly over $(u,w)\in I$ that

$$\frac{\ell(w)'(\widehat{\beta}(u)-\beta(u))}{\widehat{\sigma}(u,w)}=\frac{\ell(w)'J^{-1}(u)\mathbb{U}(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}+o_P(n^{-\varepsilon'}). \tag{F.80}$$

Moreover, by Lemma 9,

$$\sup_{(u,w)\in I}\left|(\ell(w)/\|\ell(w)\|)'J^{-1}(u)\mathbb{U}(u)\right|\lesssim_P(\log n)^{1/2}. \tag{F.81}$$

Hence,

$$\sup_{(u,w)\in I}\left|\frac{\ell(w)'J^{-1}\mathbb{U}(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}-\frac{\ell(w)'J^{-1}\mathbb{U}(u)/\sqrt{n}}{\sigma(u,w)}\right|$$
$$\leq\sup_{(u,w)\in I}\left|\frac{\ell(w)'J^{-1}\mathbb{U}(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}\right|\cdot\sup_{(u,w)\in I}\left|1-\frac{\widehat{\sigma}(u,w)}{\sigma(u,w)}\right|=o_P(n^{-\varepsilon'}) \tag{F.82}$$

by (F.76) and (F.78). Combining (F.79), (F.80), and (F.82), we have uniformly over $(u,w)\in I$ that

$$t(u,w)=\frac{\ell(w)'J^{-1}(u)\mathbb{U}(u)/\sqrt{n}}{\sigma(u,w)}+o_P(n^{-\varepsilon'}). \tag{F.83}$$

Hence, (4.27) follows.

To prove (4.28), note that by Lemma 5,

$$\sup_{u\in\mathcal{U}}\|\mathbb{U}(u)-G(u)\|\lesssim_P o(n^{-\varepsilon'});$$

see (D.72). Combining this inequality (F.77) and (F.83), we have uniformly over $(u,w)\in I$ that

$$t(u,w)=\frac{\ell(w)'J^{-1}(u)\mathbb{U}(u)/\sqrt{n}}{\sigma(u,w)}+o_P(n^{-\varepsilon'})=\frac{\ell(w)'J^{-1}(u)G(u)/\sqrt{n}}{\sigma(u,w)}+o_P(n^{-\varepsilon'}).$$

Hence, (4.28) follows. This complete the proof of the theorem. $\qquad\square$

Proof of Theorem 12. The proof relies on the following lemma:

Lemma 10. *Suppose that Conditions S and U hold. In addition, define*

$$\ell(u, w) = \left(\frac{\ell(w)' J^{-1}(u)/\sqrt{n}}{\sigma(u, w)} \right)', \quad (u, w) \in I.$$

Then

$$\sup_{(u,w) \in I} \|\ell(u, w)\| \lesssim 1$$

and

$$\|\ell(u, w) - \ell(\tilde{u}, \tilde{w})\| \lesssim (1 + \zeta_{m,\theta}^L)\|(u, w) - (\tilde{u}, \tilde{w})\|$$

uniformly over (u, w) and $(\tilde{u}, \tilde{w})$ in I.

Proof of Lemma 10. By Condition S,

$$\sigma(u, w) \gtrsim \|\ell(w)\|/\sqrt{n}, \quad \text{uniformly over } (u, w) \in I \tag{F.84}$$

The first asserted claim follows from this inequality and the fact that all eigenvalues of the matrix $J(u)$ are bounded below from zero uniformly over $u \in \mathcal{U}$ by Condition S.

To prove the second asserted claim, note that by Lemma 19,

$$\|J(u) - J(\tilde{u})\| \lesssim |u - \tilde{u}|, \quad \text{uniformly over } u \text{ and } \tilde{u} \text{ in } \mathcal{U}.$$

Combining this inequality with Conditions S and U, it follows from the triangle inequality that

$$\left| \frac{n\sigma^2(u, w)}{\|\ell(w)\|^2} - \frac{n\sigma^2(\tilde{u}, \tilde{w})}{\|\ell(\tilde{w})\|^2} \right| \lesssim (1 + \zeta_{m,\theta}^L)\|(u, w) - (\tilde{u}, \tilde{w})\|$$

uniformly over (u, w) and $(\tilde{u}, \tilde{w})$ in I. In turn, this inequality in combination with (F.84) imply that

$$\left| \frac{\sqrt{n}\sigma(u, w)}{\|\ell(w)\|} - \frac{\sqrt{n}\sigma(\tilde{u}, \tilde{w})}{\|\ell(\tilde{w})\|} \right| \lesssim (1 + \zeta_{m,\theta}^L)\|(u, w) - (\tilde{u}, \tilde{w})\|$$

uniformly over (u, w) and $(\tilde{u}, \tilde{w})$ in I since for any $a, b > 0$, we have

$$|a - b| = \frac{|a^2 - b^2|}{a + b}.$$

Further, combining the inequality above and (F.84) gives

$$\left| \frac{\|\ell(w)\|}{\sqrt{n}\sigma(u, w)} - \frac{\|\ell(\tilde{w})\|}{\sqrt{n}\sigma(\tilde{u}, \tilde{w})} \right| \lesssim (1 + \zeta_{m,\theta}^L)\|(u, w) - (\tilde{u}, \tilde{w})\| \tag{F.85}$$

uniformly over (u, w) and $(\tilde{u}, \tilde{w})$ in I since for any $a, b > 0$, we have

$$\left| \frac{1}{a} - \frac{1}{b} \right| = \frac{|a - b|}{ab}.$$

Similarly,

$$\left\| (\ell(w)/\|\ell(w)\|)' J^{-1}(u) - (\ell(\tilde{w})/\|\ell(\tilde{w})\|)' J^{-1}(\tilde{u}) \right\| \lesssim (1 + \zeta_{m,\theta}^L) \|(u,w) - (\tilde{u},\tilde{w})\| \qquad \text{(F.86)}$$

uniformly over (u,w) and $(\tilde{u},\tilde{w})$. The second asserted claim follows from combining inequalities (F.84), (F.85), and (F.86). This completes the proof of the lemma. $\qquad\square$

Getting back to the proof of the theorem, like in the proof of Theorem 11, we have uniformly over $(u,w) \in I$ that

$$t(u,w) = \frac{\ell(w)' J^{-1}(u) \mathbb{U}(u)/\sqrt{n}}{\sigma(u,w)} + o_P(n^{-\varepsilon'}); \qquad \text{(F.87)}$$

see (F.83). Next, it follows from Corollary 4 in Appendix H and Condition S that there exists a constant $C > 0$ such that $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \le C$ with probability approaching one. Thus, the asserted claim follows from using (F.87) and applying Lemma 36 conditional on $(Z_i)_{i=1}^n$ on the event $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \le C$ with

$$\ell(u,w) = \left(\frac{\ell(w)' J^{-1}(u)/\sqrt{n}}{\sigma(u,w)} \right)', \quad (u,w) \in I;$$

note that the conditions of Lemma 36 follow from Lemma 10 and the assumption $(1 + \zeta_{m,\theta}^L)^{2d_I} \zeta_m^2 = o(n^{1-\varepsilon})$. This completes the proof of the theorem. $\qquad\square$

Proof of Theorem 13. The proof relies on the following lemma:

Lemma 11. *Suppose that Conditions S and U hold. Then there exists a constant $C_V > 0$ such that with probability $1 - o(1)$,*

$$E\left[\sup_{(u,w) \in I} \left| (\ell(w)/\|\ell(w)\|)' J^{-1}(u) G(u) \right| \mid (Z_i)_{i=1}^n \right] \le C_V \sqrt{\log n}. \qquad \text{(F.88)}$$

In addition,

$$\sup_{(u,w) \in I} \left| (\ell(w)/\|\ell(w)\|)' J^{-1}(u) G(u) \right| \lesssim_P \sqrt{\log n}.$$

Proof of Lemma 11. The second asserted claim follows immediately from the first one. Hence, it suffices to prove the first asserted claim. To do so, note that by Corollary 4 in Appendix H, there exists a constant $C > 0$ such that $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \le C$ with probability $1 - o(1)$. We will show that (F.88) holds on the event $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \le C$.

For all $w \in \mathcal{W}$, denote $\xi(w) = \ell(w)/\|\ell(w)\|$. Observe that uniformly over $(u,w) \in I$, on the event $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \le C$,

$$E\left[|\xi(w)' J^{-1}(u) G(u)|^2 \mid (Z_i)_{i=1}^n \right] = u(1-u)\xi(w)' J^{-1}(u) \widehat{\Sigma} J^{-1}(u) \xi(w) \le C_r$$

for some constant C_r by Condition S. Moreover, by Conditions S and U and Lemma 19,

$$\sup_{(u,w)\in I} \|J^{-1}(u)\xi(w)\| \lesssim 1$$

and

$$\left\|J^{-1}(u)\xi(w) - J^{-1}(\tilde{u})\xi(\tilde{w})\right\| \leq \|J^{-1}(u) - J^{-1}(\tilde{u})\| \cdot \|\xi(w)\| + \|J^{-1}(\tilde{u})\| \cdot \|\xi(w) - \xi(\tilde{w})\|$$

$$\lesssim (1 + \zeta_{m,\theta}^L)\|(u,w) - (\tilde{u},\tilde{w})\|$$

uniformly over (u,w) and $(\tilde{u},\tilde{w})$ in I. Hence, by the calculation in the proof of Lemma 38, on the event $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha'Z_i)^2] \leq C$,

$$E\left[\left(\xi(w)'J^{-1}(u)G(u) - \xi(\tilde{w})'J^{-1}(\tilde{u})G(\tilde{u})\right)^2 \mid (Z_i)_{i=1}^n\right] \leq C_e(1 + \zeta_{m,\theta}^L)\|(u,w) - (\tilde{u},\tilde{w})\|,$$

where C_e is some constant. Combining these inequalities with $\log \zeta_{m,\theta}^L \lesssim \log n$ and using the fact that the dimension of the set I is independent of n and its diameter is bounded uniformly over n, (F.88) follows from Dudley's inequality applied conditional on $(Z_i)_{i=1}^n$ on the event $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha'Z_i)^2] \leq C$; see Corollary 2.2.8 in van der Vaart and Wellner [80]). $\qquad\square$

Getting back to the proof of the theorem, note that under the conditions $h = o(n^{-\varepsilon})$, $m\zeta_m^2 = o(n^{1-\varepsilon}h)$, and $m^{-\kappa} = o(n^{-\varepsilon})$, like in the proof of Theorem 11, we have

$$\sup_{(u,w)\in I} \left|\frac{\widehat{\sigma}(u,w)}{\sigma(u,w)} - 1\right| = o_P(n^{-\varepsilon'})$$

and

$$\widehat{\sigma}(u,w) \gtrsim_P \|\ell(w)\|/\sqrt{n}, \quad \text{uniformly over } (u,w) \in I; \qquad (F.89)$$

see (F.76) and (F.78). Also, by Lemma 9,

$$\sup_{(u,w)\in I} \left|(\ell(w)/\|\ell(w)\|)'J^{-1}(u)\mathbb{U}(u)\right| \lesssim_P (\log n)^{1/2},$$

and so

$$\sup_{(u,w)\in I} \left|\frac{\ell(w)'J^{-1}(u)\mathbb{U}(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}\right| \lesssim_P (\log n)^{1/2}.$$

Hence,

$$\sup_{(u,w)\in I} \left|\frac{\ell(w)'J^{-1}\mathbb{U}(u)/\sqrt{n}}{\widehat{\sigma}(u,w)} - \frac{\ell(w)'J^{-1}\mathbb{U}(u)/\sqrt{n}}{\sigma(u,w)}\right|$$

$$\leq \sup_{(u,w)\in I} \left|\frac{\ell(w)'J^{-1}\mathbb{U}(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}\right| \cdot \sup_{(u,w)\in I} \left|1 - \frac{\widehat{\sigma}(u,w)}{\sigma(u,w)}\right| = o_P(n^{-\varepsilon'}).$$

Further, since the process $\mathbb{U}^*(\cdot)$ is a copy of the process $\mathbb{U}(\cdot)$ conditional on $(Z_i)_{i=1}^n$, we also have

$$\sup_{(u,w)\in I}\left|\frac{\ell(w)'J^{-1}\mathbb{U}^*(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}-\frac{\ell(w)'J^{-1}\mathbb{U}^*(u)/\sqrt{n}}{\sigma(u,w)}\right|=o_P(n^{-\varepsilon'}). \qquad \text{(F.90)}$$

Now, consider the case of the pivotal method. We have uniformly over $(u,w)\in I$ that

$$
\begin{aligned}
t^*(u,w)&=\frac{\ell(w)'\widehat{J}^{-1}(u)\mathbb{U}^*(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}\\
&=\frac{\ell(w)'J^{-1}(u)\mathbb{U}^*(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}+o_P(n^{-\varepsilon'})=\frac{\ell(w)'J^{-1}(u)\mathbb{U}^*(u)/\sqrt{n}}{\sigma(u,w)}+o_P(n^{-\varepsilon'}),
\end{aligned}
$$

where the second equality follows from Theorem 3 and (F.89) and the third from (F.90). The first asserted claim for this case follows.

Next, consider the case of the gradient bootstrap method. We have uniformly over $(u,w)\in I$ that

$$
\begin{aligned}
t^*(u,w)&=\frac{\ell(w)'(\widehat{\beta}^*(u)-\widehat{\beta}(u))}{\widehat{\sigma}(u,w)}\\
&=\frac{\ell(w)'J^{-1}(u)\mathbb{U}^*(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}+o_P(n^{-\varepsilon'})=\frac{\ell(w)'J^{-1}(u)\mathbb{U}^*(u)/\sqrt{n}}{\sigma(u,w)}+o_P(n^{-\varepsilon'}),
\end{aligned}
$$

where the second equality follows from Theorem 4 and (F.89) and the third from (F.90). The first asserted claim for this case follows.

Next, consider the case of the Gaussian method. By Lemma 11, we have

$$\sup_{(u,w)\in I}\left|(\ell(w)/\|\ell(w)\|)'J^{-1}(u)G(u)\right|\lesssim_P (\log n)^{1/2}.$$

Hence, using the same steps as above, we obtain

$$\sup_{(u,w)\in I}\left|\frac{\ell(w)'J^{-1}G^*(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}-\frac{\ell(w)'J^{-1}G^*(u)/\sqrt{n}}{\sigma(u,w)}\right|=o_P(n^{-\varepsilon'}). \qquad \text{(F.91)}$$

So, we have uniformly over $(u,w)\in I$ that

$$
\begin{aligned}
t^*(u,w)&=\frac{\ell(w)'\widehat{J}^{-1}(u)G^*(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}\\
&=\frac{\ell(w)'J^{-1}(u)\mathbb{G}^*(u)/\sqrt{n}}{\widehat{\sigma}(u,w)}+o_P(n^{-\varepsilon'})=\frac{\ell(w)'J^{-1}(u)\mathbb{G}^*(u)/\sqrt{n}}{\sigma(u,w)}+o_P(n^{-\varepsilon'}),
\end{aligned}
$$

where the second equality follows from Theorem 6 and (F.89) and the third from (F.91). The first asserted claim for this case follows.

Finally, consider the case of the weighted bootstrap method. We have uniformly over $(u, w) \in I$ that

$$
\begin{aligned}
t^*(u, w) &= \frac{\ell(w)'(\widehat{\beta}^b(u) - \widehat{\beta}(u))}{\widehat{\sigma}(u, w)} \\
&= \frac{\ell(w)' J^{-1}(u) \mathbb{G}^*(u)/\sqrt{n}}{\widehat{\sigma}(u, w)} + o_P(n^{-\varepsilon'}) = \frac{\ell(w)' J^{-1}(u) \mathbb{G}^*(u)/\sqrt{n}}{\sigma(u, w)} + o_P(n^{-\varepsilon'}),
\end{aligned}
$$

where the second equality follows from Theorem 7 and (F.89) and the third from (F.91). The first asserted claim for this case follows.

The second asserted claim follows from the same argument as that used in the proof of Theorem 3. This completes the proof of theorem. □

Proof of Theorem 14. Like in the proof of Theorem 11,

$$
\sup_{(u,w) \in I} \left| \frac{\widehat{\sigma}(u, w)}{\sigma(u, w)} - 1 \right| = o_P(n^{-\varepsilon'}), \tag{F.92}
$$

$$
\widehat{\sigma}(u, w) \gtrsim_P \|\ell(w)\|/\sqrt{n}, \quad \text{uniformly over } (u, w) \in I; \tag{F.93}
$$

see (F.76) and (F.78). Also, by the same argument as that used in Lemma 9,

$$
\sup_{(u,w) \in I} \left| \frac{(\ell(w)/\|\ell(w)\|)' J^{-1}(u)}{\sqrt{n}} \sum_{i=1}^n (\pi_i - 1) Z_i(u - 1\{U_i \leq u\}) \right| \lesssim_P (\log n)^{1/2}.
$$

In addition, by Condition S, $\|\ell(w)\|/(\sqrt{n}\sigma(u, w)) \lesssim 1$ uniformly over $(u, w) \in I$, and so

$$
\sup_{(u,w) \in I} \left| \frac{\ell(w)' J^{-1}(u)}{\sqrt{n}\sigma(u, w)} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n (\pi_i - 1) Z_i(u - 1\{U_i \leq u\}) \right| \lesssim_P (\log n)^{1/2}.
$$

Combining this inequality with (F.92) gives

$$
\sup_{(u,w) \in I} \left| \left(\frac{1}{\sigma(u, w)} - \frac{1}{\widehat{\sigma}(u, w)} \right) \frac{\ell(w)' J^{-1}(u)}{\sqrt{n}} \frac{1}{\sqrt{n}} \sum_{i=1}^n (\pi_i - 1) Z_i(u - 1\{U_i \leq u\}) \right| = o_P(n^{-\varepsilon'}).
$$

Hence, uniformly over $(u, w) \in I$,

$$
\begin{aligned}
t^*(u, w) &= \frac{\ell(w)'(\widehat{\beta}^b(u) - \widehat{\beta}(u))}{\widehat{\sigma}(u, w)} \\
&= \frac{\ell(w)' J^{-1}(u)}{\sqrt{n}\widehat{\sigma}(u, w)} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n (\pi_i - 1) Z_i(u - 1\{U_i \leq u\}) + o_P(n^{-\varepsilon'}) \\
&= \frac{\ell(w)' J^{-1}(u)}{\sqrt{n}\sigma(u, w)} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^n (\pi_i - 1) Z_i(u - 1\{U_i \leq u\}) + o_P(n^{-\varepsilon'}), \tag{F.94}
\end{aligned}
$$

where the second line follows from Lemma 7 and (F.93). Next, it follows from Corollary 4 in Appendix H and Condition S that there exists a constant $C > 0$ such that $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \leq C$ with probability approaching one. Thus, the first asserted claim follows from using (F.94) and applying Lemma 36 conditional on $(Z_i)_{i=1}^n$ on the event $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \leq C$ with

$$\ell(u, w) = \left(\frac{\ell(w)' J^{-1}(u)/\sqrt{n}}{\sigma(u, w)} \right)', \quad (u, w) \in I;$$

note that the conditions of Lemma 36 follow from Lemma 10 and the assumption $(1 + \zeta_{m,\theta}^L)^{2d_I} \zeta_m^2 = o(n^{1-\varepsilon})$. The second asserted claim follows from the same argument as that used in the proof of Theorem 3. This completes the proof of theorem. $\qquad \square$

Proof of Theorem 15. We split the proof into five steps.

Step 1 (Coupling for V). Here for the random variable

$$V = \sup_{(u,w) \in I} \frac{|\widehat{\theta}(u, w) - \theta(u, w)|}{\widehat{\sigma}(u, w)} = \sup_{(u,w) \in I} |t(u, w)|$$

and each sample size n, we show that there exists a random variable $\bar{V} = \bar{V}_n$ that is conditional on $(Z_i)_{i=1}^n$ equal in distribution to

$$\sup_{(u,w) \in I} \frac{|\ell(w)' J^{-1}(u) G(u)/\sqrt{n}|}{\sigma(u, w)} \tag{F.95}$$

and is such that

$$|V - \bar{V}| = o_P(n^{-\varepsilon'}),$$

where $G(\cdot) = G_n(\cdot)$ is a process on $\mathcal{U}$ that is conditionally on $(Z_i)_{i=1}^n$ zero-mean Gaussian with a.s. continuous sample paths and the covariance function

$$E\left[G(u_1) G(u_2)' \mid (Z_i)_{i=1}^n \right] = \mathbb{E}_n[Z_i Z_i'](u_1 \wedge u_2 - u_1 u_2), \quad \text{for all } u_1 \text{ and } u_2 \text{ in } \mathcal{U}.$$

To prove this claim, first, note that by the first part of Theorem 11, we have

$$|V - \widetilde{V}| = o_P(n^{-\varepsilon'}), \tag{F.96}$$

where

$$\widetilde{V} = \sup_{(u,w) \in I} \frac{|\ell(w)' J^{-1}(u) \mathbb{U}(u)/\sqrt{n}|}{\sigma(u, w)}.$$

Second, consider the function classes

$$\mathcal{W}_1 = \left\{ (Z, U) \mapsto u - 1\{U \le u\} \colon (u, w) \in I \right\},$$

$$\mathcal{W}_2 = \left\{ (Z, U) \mapsto \|\ell(w)\|/(\sqrt{n}\sigma(u, w)) \colon (u, w) \in I \right\},$$

$$\mathcal{W}_3 = \left\{ (Z, U) \mapsto (\ell(w)/\|\ell(w)\|)' J^{-1}(u) Z \colon (u, w) \in I \right\},$$

mapping $B_m(0, \zeta_m) \times [0, 1]$ into $\mathbb{R}$. The function class $\mathcal{W}_1$ has an envelope $F_1(Z, U) = 1$ and its uniform entropy numbers satisfy

$$\sup_Q \log N(\epsilon \|F_1\|_{Q,2}, \mathcal{W}_1, L_2(Q)) \lesssim \log(1/\epsilon), \quad \text{uniformly over } 0 < \epsilon \le 1.$$

Also, by Condition S, $\|\ell(w)\|/(\sqrt{n}\sigma(u, w)) \lesssim 1$ uniformly over $(u, w) \in I$, and so the function class $\mathcal{W}_2$ has an envelope $F_2(Z, U) = C$ for sufficiently large constant C and its uniform entropy numbers satisfy

$$\sup_Q \log N(\epsilon \|F_2\|_{Q,2}, \mathcal{W}_2, L_2(Q)) \lesssim \log(1/\epsilon), \quad \text{uniformly over } 0 < \epsilon \le 1.$$

In addition, the function class $\mathcal{W}_3$ has an envelope $F_3(Z, U) = C\zeta_m$ for sufficiently large constant C and by Lemma 8, the uniform entropy numbers of $\mathcal{W}_3$ satisfy

$$\sup_Q \log N(\epsilon \|F_3\|_{Q,2}, \mathcal{W}_3, L_2(Q)) \lesssim \log(n/\epsilon), \quad \text{uniformly over } 0 < \epsilon \le 1.$$

Hence, by Lemma 24, the uniform entropy numbers of the function class

$$\widetilde{\mathcal{F}} = \left\{ (Z, U) \mapsto \frac{(\ell(w)/\|\ell(w)\|)' J^{-1}(u) Z (u - 1\{U \le u\})}{\sqrt{n}\sigma(u, w)/\|\ell(w)\|} \colon (u, w) \in I \right\} \subset \mathcal{W}_1 \mathcal{W}_2 \mathcal{W}_3$$

satisfy

$$\sup_Q \log N(\epsilon \|F\|_{Q,2}, \widetilde{\mathcal{F}}, L_2(Q)) \lesssim \log(n/\epsilon), \quad \text{uniformly over } 0 < \epsilon \le 1,$$

where $F(Z, U) = C\zeta_m$ for sufficiently large constant C is its envelope. Thus, the uniform entropy numbers of the function class $\mathcal{F} = \widetilde{\mathcal{F}} \cup (-\widetilde{\mathcal{F}})$ satisfy

$$\sup_Q \log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \lesssim \log(n/\epsilon), \quad \text{uniformly over } 0 < \epsilon \le 1, \tag{F.97}$$

where the function F is its envelope. Moreover,

$$\sup_{f \in \mathcal{F}} E[f(Z, U)^2] = 1, \quad \sup_{f \in \mathcal{F}} E[|f(Z, U)|^3] \lesssim \zeta_m, \quad \sup_{f \in \mathcal{F}} E[|f(Z, u)|^4] \lesssim \zeta_m^2 \tag{F.98}$$

Now, observe that

$$\widetilde{V} = \sup_{f \in \mathcal{F}} \mathbb{G}_n f$$

and that by (F.97) and (F.98), Conditions (A), (B), and (C) of Chernozhukov, Chetverikov and Kato [28] hold for the function class $\mathcal{F}$ with $A = n$, $v = C$, $\sigma = 1$, $b = C\zeta_m$, and arbitrarily large q. Hence, given that $\zeta_m^2 = o(n^{1-\varepsilon})$, Theorem 2.1 of Chernozhukov, Chetverikov and Kato [28] implies that there exists a random variable $\bar{V}$ with the required distribution and such that

$$|\widetilde{V} - \bar{V}| = o_P(n^{-\varepsilon'}).$$

Combining this bound with (F.96) gives the claim of this step.

Step 2 (Coupling for V^*). Here for the random variable

$$V^* = \sup_{(u,w) \in I} |t^*(u,w)|$$

and for each sample size n, we show that there exists a random variable $\bar{V}^* = \bar{V}_n^*$ that (i) is conditional on $(Z_i)_{i=1}^n$ equal in distribution to the random variable in (F.95), (ii) depends on the data only via $(Z_i)_{i=1}^n$, and (iii) is such that

$$|V^* - \bar{V}^*| = o_P(n^{-\varepsilon'}). \tag{F.99}$$

To prove this claim, we consider four resampling methods separately. We start with the pivotal method case. By Theorem 13, we have

$$|V^* - \widetilde{V}^*| = o_P(n^{-\varepsilon'}), \tag{F.100}$$

where

$$\widetilde{V}^* = \sup_{(u,w) \in I} \frac{|\ell(w)' J^{-1}(u) \mathbb{U}^*(u)/\sqrt{n}|}{\sigma(u,w)}.$$

Then given that the process $\mathbb{U}^*(\cdot)$ is a copy of $\mathbb{U}(\cdot)$ conditional on $(Z_i)_{i=1}^n$ and depends on the data only via $(Z_i)_{i=1}^n$, by the same argument as that used in Step 1, it follows that there exists a random variable $\bar{V}^*$ with the required distribution and such that

$$|\widetilde{V}^* - \bar{V}^*| = o_P(n^{-\varepsilon'}).$$

Combining this bound with (F.100), gives the claim of Step 2 in this case.

Next, in the gradient bootstrap method case, construction of a random variable $\bar{V}^*$ with the required distribution follows from the same argument as that used in the pivotal process case.

Next, consider the Gaussian method case. By Theorem 13, we have uniformly over $(u,w) \in I$ that

$$t^*(u,w) = \frac{\ell(w)' J^{-1}(u) G^*(u)/\sqrt{n}}{\sigma(u,w)} + o_P(n^{-\varepsilon'}).$$

Thus, it follows that (F.99) holds for

$$\bar{V}^* = \sup_{(u,w)\in I} \frac{|\ell(w)'J^{-1}(u)G^*(u)/\sqrt{n}|}{\sigma(u,w)},$$

and since the random variable $\bar{V}^*$ has the required distribution, the claim of Step 2 in this case holds.

Finally, consider the weighted bootstrap method case. Like in the proof of Theorem 11,

$$\sup_{(u,w)\in I} \left|\frac{\widehat{\sigma}(u,w)}{\sigma(u,w)} - 1\right| = o_P(n^{-\varepsilon'}), \tag{F.101}$$

$$\widehat{\sigma}(u,w) \gtrsim_P \|\ell(w)\|/\sqrt{n}, \quad \text{uniformly over } (u,w) \in I; \tag{F.102}$$

see (F.76) and (F.78). Also, by the same argument as that used in Lemma 9,

$$\sup_{(u,w)\in I} \left| \frac{(\ell(w)/\|\ell(w)\|)'J^{-1}(u)}{\sqrt{n}} \sum_{i=1}^{n} (\pi_i - 1)Z_i(u - 1\{U_i \le u\}) \right| \lesssim_P (\log n)^{1/2}.$$

In addition, by Condition S, $\|\ell(w)\|/(\sqrt{n}\sigma(u,w)) \lesssim 1$ uniformly over $(u,w) \in I$, and so

$$\sup_{(u,w)\in I} \left| \frac{\ell(w)'J^{-1}(u)}{\sqrt{n}\sigma(u,w)} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\pi_i - 1)Z_i(u - 1\{U_i \le u\}) \right| \lesssim_P (\log n)^{1/2}.$$

Combining this inequality with (F.101) gives

$$\sup_{(u,w)\in I} \left| \left(\frac{1}{\sigma(u,w)} - \frac{1}{\widehat{\sigma}(u,w)}\right) \frac{\ell(w)'J^{-1}(u)}{\sqrt{n}} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\pi_i - 1)Z_i(u - 1\{U_i \le u\}) \right| = o_P(n^{-\varepsilon'}).$$

Hence, by Lemma 7, we have uniformly over $(u,w) \in I$ that

$$\begin{aligned}
t^*(u,w) &= \frac{\ell(w)'(\widehat{\beta}^b(u) - \widehat{\beta}(u))}{\widehat{\sigma}(u,w)} \\
&= \frac{\ell(w)'J^{-1}(u)}{\sqrt{n}\widehat{\sigma}(u,w)} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\pi_i - 1)Z_i(u - 1\{U_i \le u\}) + o_P(n^{-\varepsilon'}) \\
&= \frac{\ell(w)'J^{-1}(u)}{\sqrt{n}\sigma(u,w)} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\pi_i - 1)Z_i(u - 1\{U_i \le u\}) + o_P(n^{-\varepsilon'}),
\end{aligned}$$

where we used (F.102) in the second line. Thus, denoting

$$\widetilde{V}^* = \sup_{(u,w)\in I} \left| \frac{\ell(w)'J^{-1}(u)}{\sqrt{n}\sigma(u,w)} \cdot \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (\pi_i - 1)Z_i(u - 1\{U_i \le u\}) \right|,$$

we obtain

$$|V^* - \widetilde{V}^*| = o_P(n^{-\varepsilon'}). \tag{F.103}$$

Now we apply Lemma 41 with

$$\ell(u, w) = \left(\frac{\ell(w)' J^{-1}(u)}{\sqrt{n}\sigma(u, w)} \right)' \tag{F.104}$$

and $v_i = \pi_i - 1$ to couple $\widetilde{V}^*$ with a random variable $\bar{V}^*$ that has the required distribution. Observe that

$$\sup_{(u,w) \in I} \|\ell(u, w)\| \lesssim 1 \tag{F.105}$$

and

$$\|\ell(u, w) - \ell(\tilde{u}, \tilde{w})\| \lesssim (1 + \zeta_{m,\theta}^L)\|(u, w) - (\tilde{u}, \tilde{w})\|, \tag{F.106}$$

uniformly over (u, w) and $(\tilde{u}, \tilde{w})$ in I by Lemma 10. Note also that $\log \zeta_{m,\theta}^L \lesssim \log n$ by Condition U. Thus, conditions of Lemma 41 on $\ell(u, w)$ are satisfied. Moreover, by Corollary 4 in Appendix H, there exists a constant $C > 0$ such that $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \leq C$ with probability $1 - o(1)$. Thus, applying Lemma 41 conditional on $(Z_i)_{i=1}^n$ on the event $\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2] \leq C$ shows that there exists a random variable $\bar{V}^*$ with the required distribution such that

$$|\widetilde{V}^* - \bar{V}^*| = o_P(n^{-\varepsilon}).$$

Combining this bound with (F.103) gives (F.99), so that the claim of Step 2 in this case holds. This completes Step 2.

Step 3 (Approximation of Critical Values). For all $\eta \in (0, 1)$, let $\bar{k}^*(\eta)$ denote the η-th quantile of the conditional distribution of $\bar{V}^*$ given the data. Here we show that

$$P\left(k^*(1 - \alpha) \leq \bar{k}^*(1 - \alpha + \nu_{n,1}) + n^{-\varepsilon'} \right) = 1 - o(1),$$

$$P\left(k^*(1 - \alpha) \geq \bar{k}^*(1 - \alpha - \nu_{n,1}) - n^{-\varepsilon'} \right) = 1 - o(1),$$

for some sequence of positive numbers $(\nu_{n,1})_{n \geq 1}$ converging to zero. Indeed, recall that $k^*(\eta)$ denotes the η-th quantile of the conditional distribution of V^* given the data and also that $|V^* - \bar{V}^*| = o_P(n^{-\varepsilon'})$ by Step 2. Hence, the claim of this step follows from Lemma 13.

Step 4 (Anti-Concentration). Here we show that

$$P\left(\bar{k}^*(1 - \alpha - \nu_{n,1}) - \bar{k}^*(1 - \alpha - \nu_{n,1} - \nu_{n,2}) > 2n^{-\varepsilon'} \right) = 1 - o(1),$$

$$P\left(\bar{k}^*(1 - \alpha + \nu_{n,1} + \nu_{n,2}) - \bar{k}^*(1 - \alpha + \nu_{n,1}) > 2n^{-\varepsilon'} \right) = 1 - o(1),$$

for some sequence of positive numbers $(\nu_{n,2})_{n\geq 1}$ converging to zero. By Lemma 11, there exists a constant $C_V > 0$ such that with probability $1 - o(1)$,

$$E\left[\sup_{(u,w)\in I}\left|(\ell(w)/\|\ell(w)\|)'J^{-1}(u)G(u)\right| \mid (Z_i)_{i=1}^n\right] \leq C_V\sqrt{\log n}.$$

Also, by Condition S, $\|\ell(w)\|/(\sqrt{n}\sigma(u,w)) \lesssim 1$ uniformly over $(u,w) \in I$, and so with probability $1 - o(1)$,

$$E\left[\sup_{(u,w)\in I}\frac{|\ell(w)'J^{-1}(u)G(u)/\sqrt{n}|}{\sigma(u,w)} \mid (Z_i)_{i=1}^n\right] \leq C_V\sqrt{\log n}. \tag{F.107}$$

Hence, existence of the required sequence $(\nu_{n,2})_{n\geq 1}$ follows from recalling that $\bar{V}^*$ is equal in distribution conditional on $(Z_i)_{i=1}^n$ to the random variable under the expectation sign in (F.107) and applying Lemma 12 conditional $(Z_i)_{i=1}^n$ on the event in (F.107), since $n^{-\varepsilon'}\sqrt{\log n} \to 0$ as $n \to \infty$. This completes Step 4.

Step 5 (Main Argument). Here we complete the proof of the theorem. Observe that

$$\begin{aligned}
P\Big(V > k^*(1-\alpha)\Big) &\leq P\Big(V > \bar{k}^*(1-\alpha-\nu_{n,1}) - n^{-\varepsilon'}\Big) + o(1)\\
&\leq P\Big(\bar{V} > \bar{k}^*(1-\alpha-\nu_{n,1}) - 2n^{-\varepsilon'}\Big) + o(1)\\
&\leq P\Big(\bar{V} > \bar{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2})\Big) + o(1)\\
&\leq E\Big[P\Big(\bar{V} > \bar{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2})|(Z_i)_{i=1}^n\Big)\Big] + o(1)\\
&\leq \alpha + \nu_{n,1} + \nu_{n,2} + o(1) = \alpha + o(1),
\end{aligned}$$

where the first line follows from Step 3, the second from Step 1, the third from Step 4, the fourth from the law of iterated expectations, and the fifth from the definition of $\bar{V}$ and $\bar{k}^*(1-\alpha-\nu_{n,1}-\nu_{n,2})$. Similarly,

$$\begin{aligned}
P\Big(V > k^*(1-\alpha)\Big) &\geq P\Big(V > \bar{k}^*(1-\alpha+\nu_{n,1}) + n^{-\varepsilon'}\Big) + o(1)\\
&\geq P\Big(\bar{V} > \bar{k}^*(1-\alpha+\nu_{n,1}) + 2n^{-\varepsilon'}\Big) + o(1)\\
&\geq P\Big(\bar{V} > \bar{k}^*(1-\alpha+\nu_{n,1}+\nu_{n,2})\Big) + o(1)\\
&\geq E\Big[P\Big(\bar{V} > \bar{k}^*(1-\alpha+\nu_{n,1}+\nu_{n,2})|(Z_i)_{i=1}^n\Big)\Big] + o(1)\\
&\geq \alpha - \nu_{n,1} - \nu_{n,2} + o(1) = \alpha + o(1),
\end{aligned}$$

This gives the first asserted claim. The second asserted claim follows immediately from the first one. The last asserted claim follows since $\sup_{(u,w)\in I} \widehat{\sigma}(u,w) \lesssim_P \sqrt{\zeta^2_{m,\theta}/n}$ and $k^*(1-\alpha) \lesssim_P \sqrt{\log n}$ by Steps 3 and 4. This completes the proof of the theorem. $\qquad\square$

We use the following lemmas in the proof of Theorem 15.

Lemma 12 (Anti-Concentration for Separable Gaussian Processes, Chernozhukov, Chetverikov and Kato [25])**.** *Let* $Y = (Y_t)_{t\in T}$ *be a separable Gaussian process indexed by a semimetric space* T *such that* $E[Y_t] = 0$ *and* $E[Y_t^2] = 1$ *for all* $t \in T$. *Assume that* $\sup_{t\in T} |Y_t| < \infty$ *a.s. Then for all* $\varepsilon \geq 0$ *and some absolute constant* A *we have*

$$\sup_{x\in\mathbb{R}} P\left(\left|\sup_{t\in T} |Y_t| - x\right| \leq \varepsilon\right) \leq A\varepsilon E\left[\sup_{t\in T} |Y_t|\right].$$

Lemma 13 (Closeness in Probability Implies Closeness of Conditional Quantiles)**.** *Let* X_n *and* Y_n *be random variables and* $\mathcal{D}_n$ *be a random vector. Let* $F_{X_n}(x|\mathcal{D}_n)$ *and* $F_{Y_n}(x|\mathcal{D}_n)$ *denote the conditional distribution functions, and* $F_{X_n}^{-1}(p|\mathcal{D}_n)$ *and* $F_{Y_n}^{-1}(p|\mathcal{D}_n)$ *denote the corresponding conditional quantile functions. If* $|X_n - Y_n| = o_P(\gamma_n)$, *then there exists a sequence of positive numbers* $(\nu_n)_{n\geq 1}$ *converging to zero such that with probability* $1 - o(1)$,

$$F_{X_n}^{-1}(p|\mathcal{D}_n) \leq F_{Y_n}^{-1}(p+\nu_n|\mathcal{D}_n) + \gamma_n \ \text{ and } \ F_{Y_n}^{-1}(p|\mathcal{D}_n) \leq F_{X_n}^{-1}(p+\nu_n|\mathcal{D}_n) + \gamma_n$$

for all $p \in (\nu_n, 1-\nu_n)$.

Proof. Since $|X_n - Y_n| = o_P(\gamma_n)$, there exists a sequence of positive numbers $(\nu_n)_{n\geq 1}$ converging to zero such that $P(|X_n - Y_n| > \gamma_n) = o(\nu_n)$. Hence,

$$P\left(P(|X_n - Y_n| > \gamma_n|\mathcal{D}_n) \leq \nu_n\right) \to 1,$$

that is, there is a set Ω_n of values of $\mathcal{D}_n$ such that $P(\Omega_n) \to 1$ and $P(|X_n - Y_n| > \gamma_n|\mathcal{D}_n) \leq \nu_n$ for all $\mathcal{D}_n \in \Omega_n$. Now, for all $\mathcal{D}_n \in \Omega_n$,

$$F_{X_n}(x|\mathcal{D}_n) \geq F_{Y_n+\gamma_n}(x|\mathcal{D}_n) - \nu_n \text{ and } F_{Y_n}(x|\mathcal{D}_n) \geq F_{X_n+\gamma_n}(x|\mathcal{D}_n) - \nu_n, \quad \text{for all } x \in \mathbb{R},$$

which implies the inequality stated in the lemma, by definition of the conditional quantile function and equivariance of quantiles to location shifts. $\qquad\square$

Proof of Theorem 16. The first asserted claim follows from the same argument as that used in the proof of Theorem 15: simply replace $\sup_{(u,w)\in I} |t(u,w)|$ and $\sup_{(u,w)\in I} |t^*(u,w)|$ by $\sup_{(u,w)\in I} t(u,w)$ and $\sup_{(u,w)\in I} t^*(u,w)$. To prove the second asserted claim, suppose to

the contrary that there exists a set of data-generating processes $\mathcal{M} = \mathcal{M}_n$ satisfying H_0 and such that the conditions of Theorem 15 hold uniformly over this set but the inequality

$$\sup_{M \in \mathcal{M}} P_M\left(T > \tilde{k}^*(1 - \alpha)\right) \leq \alpha + o(1)$$

does not hold. Then there exists $\epsilon > 0$ and a sequence of integers $(n_k)_{k \geq 1}$ such that for each $k \geq 1$, we can find $M_{n_k} \in \mathcal{M}_{n_k}$ such that

$$P_{M_{n_k}}\left(T > \tilde{k}^*(1 - \alpha)\right) > \alpha + \epsilon, \quad \text{for all } k \geq 1.$$

However, this contradicts the first asserted claim because we allow the data-generating process to depend on n and the sequence of data-generating processes $(M_{n_k})_{k \geq 1}$ satisfy conditions of Theorem 15. The second asserted claim follows. This completes the proof of the theorem. $\qquad\square$

APPENDIX G. A LEMMA ON STRONG APPROXIMATION OF AN EMPIRICAL PROCESS OF AN INCREASING DIMENSION BY A GAUSSIAN PROCESS

Lemma 14. *(Approximation of a Sequence of Empirical Processes of Increasing Dimension by a Sequence of Gaussian Processes) Let $(Z_i)_{i=1}^n$ be a sequence of non-stochastic vectors in $\mathbb{R}^m$ and consider the empirical process $\mathbb{U}_n$ in $[\ell^\infty(\mathcal{U})]^m$, $\mathcal{U} \subseteq (0, 1)$, defined by*

$$\mathbb{U}_n(u) = \mathbb{G}_n\left(v_i Z_i \psi_i(u)\right), \quad \psi_i(u) = u - 1\{U_i \leq u\}, \quad u \in \mathcal{U},$$

where $(U_i, v_i)_{i=1}^n$ is an i.i.d. sequence of pairs of independent random variables where $U_i \sim$ Uniform$(0, 1)$, $E[v_i^2] = 1$, $E[|v_i|^4] \lesssim 1$, and $\max_{1 \leq i \leq n} |v_i| \lesssim_P \log n$. Suppose that the vectors Z_i are such that

$$\sup_{\alpha \in S^{m-1}} \mathbb{E}_n\left[(\alpha' Z_i)^2\right] \lesssim 1, \quad \max_{1 \leq i \leq n} \|Z_i\| \lesssim \zeta_m, \quad \text{and} \quad m^7 \zeta_m^6 = o(n^{1-\varepsilon})$$

where ζ_m satisfies $1/\zeta_m \lesssim 1$ and $\varepsilon > 0$ is some constant. Then there exists a sequence of zero-mean Gaussian processes $(G_n)_{n \geq 1}$ with a.s. continuous paths such that (i) the covariance functions of G_n coincide with those of $\mathbb{U}_n$, namely,

$$E[G_n(u)G_n(\tilde{u})'] = E[\mathbb{U}_n(u)\mathbb{U}_n(\tilde{u})'] = \mathbb{E}_n[Z_i Z_i'](u \wedge \tilde{u} - u\tilde{u}), \quad \text{for all } u \text{ and } \tilde{u} \in \mathcal{U},$$

and (ii) G_n approximates $\mathbb{U}_n$, namely,

$$\sup_{u \in \mathcal{U}} \|\mathbb{U}_n(u) - G_n(u)\| \lesssim_P o(n^{-\varepsilon'}),$$

where $\varepsilon' > 0$ is some constant.

Proof. The proof is based on the use of maximal inequalities and Yurinskii's coupling. We define the sequence of projections $\pi_j \colon \mathcal{U} \to \mathcal{U}$, $j = 0, 1, 2, \ldots, \infty$ by $\pi_j(u) = u_{kj} = k/2^j$ if $u \in ((k-1)/2^j, k/2^j]$, $k = 1, \ldots, 2^j$. In what follows, given a process G in $[\ell^\infty(\mathcal{U})]^m$ and its projection $G \circ \pi_j$, whose paths are step functions with at most 2^j steps, we shall identify the process $G \circ \pi_j$ with a random vector $G \circ \pi_j$ in $\mathbb{R}^{m2^j}$, when convenient. Analogously, given a random vector W in $\mathbb{R}^{m2^j}$, we identify it with a process W in $[\ell^\infty(\mathcal{U})]^m$, whose paths are step functions with at most 2^j steps.

The following relations will be proven below for some $\varepsilon' > 0$ and some $j = j_n \to \infty$:

(1) (Finite-Dimensional Approximation)

$$r_1 = \sup_{u \in \mathcal{U}} \|\mathbb{U}_n(u) - \mathbb{U}_n \circ \pi_j(u)\| \lesssim_P o(n^{-\varepsilon'});$$

(2) (Coupling with a Normal Vector) there exists $\mathcal{N}_{nj} =_d N(0, \mathrm{var}[\mathbb{U}_n \circ \pi_j])$ such that

$$r_2 = \|\mathcal{N}_{nj} - \mathbb{U}_n \circ \pi_j\| \lesssim_P o(n^{-\varepsilon'});$$

(3) (Embedding a Normal Vector into a Gaussian Process) there exists a Gaussian process G_n with properties stated in the lemma such that $\mathcal{N}_{nj} = G_n \circ \pi_j$ a.s.;

(4) (Infinite-Dimensional Approximation)

$$r_3 = \sup_{u \in \mathcal{U}} \|G_n(u) - G_n \circ \pi_j(u)\| \lesssim_P o(n^{-\varepsilon'}).$$

The result then follows from the triangle inequality:

$$\sup_{u \in \mathcal{U}} \|\mathbb{U}_n(u) - G_n(u)\| \leq r_1 + r_2 + r_3.$$

We now prove relations (1)-(4). Relation (1) follows from

$$r_1 = \sup_{u \in \mathcal{U}} \|\mathbb{U}_n(u) - \mathbb{U}_n \circ \pi_j(u)\| \leq \sup_{|u - \tilde{u}| \leq 2^{-j}} \|\mathbb{U}_n(u) - \mathbb{U}_n(\tilde{u})\|$$

$$\lesssim_P \sqrt{2^{-j} m \log n} + \sqrt{\frac{m^2 \zeta_m^2 \log^4 n}{n}} \lesssim_P o(n^{-\varepsilon'}),$$

where the first inequality in the second line follows from Lemma 15 and second holds by setting $2^j = mn^{\tilde{\varepsilon}}$ for sufficiently small $\tilde{\varepsilon}$ and recalling that $m^7 \zeta_m^6/n = o(n^{-\varepsilon})$ and $1/\zeta_m \lesssim 1$.

Relation (2) follows from the use of Yurinskii's coupling (Pollard [69], Chapter 10, Theorem 10): Let $\xi_1, \ldots, \xi_n$ be independent zero-mean p-vectors such that $\kappa := \sum_{i=1}^n E\left[\|\xi_i\|^3\right]$ is finite. Let $S = \xi_1 + \cdots + \xi_n$. Then for each $\delta > 0$, there exists a random vector T with a

$N(0, \text{var}(S))$ distribution such that

$$P\{\|S - T\| > 3\delta\} \le C_0 B \left(1 + \frac{|\log(1/B)|}{p} \right) \quad \text{where } B := \kappa p \delta^{-3},$$

for some universal constant C_0.

In order to apply the coupling, we collapse $v_i Z_i \psi_i \circ \pi_j$ to a p-vector, and let

$$\xi_i = v_i Z_i \psi_i \circ \pi_j \in \mathbb{R}^p, \quad p = 2^j m$$

so that $\mathbb{U}_n \circ \pi_j = \sum_{i=1}^n \xi_i / \sqrt{n}$. Then

$$\mathbb{E}_n E[\|\xi_i\|^3] = \mathbb{E}_n E \left[\left(\sum_{k=1}^{2^j} \sum_{w=1}^m \psi_i(u_{kj})^2 v_i^2 Z_{iw}^2 \right)^{3/2} \right] \le 2^{3j/2} E[|v_i|^3] \mathbb{E}_n[\|Z_i\|^3] \lesssim 2^{3j/2} \zeta_m^3.$$

Therefore, by Yurinskii's coupling, since $\log n \lesssim 2^j m$, by the choice $2^j = mn^{\tilde{\varepsilon}}$,

$$P \left\{ \left\| \frac{\sum_{i=1}^n \xi_i}{\sqrt{n}} - \mathcal{N}_{nj} \right\| \ge 3\delta \right\} \lesssim \frac{n 2^{3j/2} \zeta_m^3 2^j m}{(\delta \sqrt{n})^3} = \frac{2^{5j/2} m \zeta_m^3}{\delta^3 n^{1/2}} \to 0$$

by setting $\delta = (2^{5j} m^2 \zeta_m^6 \log n / n)^{1/6}$. This verifies relation (2) with

$$r_2 \lesssim_P \left(\frac{2^{5j} m^2 \zeta_m^6 \log n}{n} \right)^{1/6} = \left(\frac{n^{5\tilde{\varepsilon}} m^7 \zeta_m^6 \log n}{n} \right)^{1/6} = o(n^{-\varepsilon'}),$$

provided that $5\tilde{\varepsilon} + 6\varepsilon' < \varepsilon$.

Relation (3) follows from the a.s. embedding of a finite-dimensional random normal vector into a path of a continuous Gaussian process, which is possible by Lemma 17.

Relation (4) follows from

$$r_3 = \sup_{u \in \mathcal{U}} \|G_n(u) - G_n \circ \pi_j(u)\| \le \sup_{|u - \tilde{u}| \le 2^{-j}} \|G_n(u) - G_n(\tilde{u})\|$$

$$\lesssim_P \sqrt{2^{-j} m \log n} \lesssim_P o(n^{-\varepsilon'}),$$

where the first inequality in the second line follows from Lemma 16 and the second holds by the choice of j. This completes the proof of the lemma. $\qquad \square$

Next, we establish auxiliary lemmas that were used in the preceding proof.

Lemma 15 (Finite-Dimensional Approximation). *Consider the setting of Lemma 14 and denote $\varphi = \sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2]$. Then for any $\gamma > 0$, the process $\mathbb{U}_n$ satisfies*

$$\sup_{|u - \tilde{u}| \le \gamma} \|\mathbb{U}_n(u) - \mathbb{U}_n(\tilde{u})\| \lesssim_P \sqrt{\gamma \varphi m \log n} + \sqrt{\frac{m^2 \zeta_m^2 \log^4 n}{n}}.$$

Proof. Consider the function class

$$\mathcal{G}_{m,n} = \Big\{ (Z, U, v) \mapsto g_{\alpha, u}(Z, U, v) = (\alpha' Z) \cdot (u - \mathbf{1}\{U \leq u\}) \colon u \in \mathcal{U}, \alpha \in S^{m-1} \Big\},$$

mapping $B_m(0, \zeta_m) \times [0, 1] \times \mathbb{R}$ into $\mathbb{R}$. Note that $G_{m,n}(Z, U, v) = \zeta_m$ is its envelope. By Lemma 27, the uniform entropy numbers of $\mathcal{G}_{m,n}$ satisfy

$$\sup_Q \log N(\epsilon \| G_{m,n} \|_{Q,2}, \mathcal{G}_{m,n}, L_2(Q)) \lesssim O(m) \log(1/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1. \tag{G.108}$$

Next, consider the function class

$$\widetilde{\mathcal{G}}_{m,n} = \Big\{ (Z, U, v) \mapsto v \cdot g(Z, U, v) \colon g \in \mathcal{G}_{m,n} \Big\}.$$

The function $\widetilde{G}_{m,n}(Z, U, v) = |v| \cdot G_{m,n}(Z, U, v)$ is an envelope of $\widetilde{\mathcal{G}}_{m,n}$. By (G.108) and Lemma 24, the uniform entropy numbers of $\widetilde{\mathcal{G}}_{m,n}$ satisfy

$$\sup_Q \log N(\epsilon \| \widetilde{G}_{m,n} \|_{Q,2}, \widetilde{\mathcal{G}}_{m,n}, L_2(Q)) \lesssim O(m) \log(1/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1. \tag{G.109}$$

Further, consider the function class

$$\mathcal{G}_{m,n,\gamma} = \Big\{ (Z, U, v) \mapsto v \cdot (g_{\alpha, u}(Z, U, v) - g_{\alpha, \tilde{u}}(Z, U, v)) \colon u, \tilde{u} \in \mathcal{U}, \alpha \in S^{m-1}, |u - \tilde{u}| \leq \gamma \Big\}.$$

The function $2\widetilde{G}_{m,n}$ is its envelope. By (G.109) and Lemma 24, the uniform entropy numbers of $\mathcal{G}_{m,n,\gamma}$ satisfy

$$\sup_Q \log N(\epsilon \| 2\widetilde{G}_{m,n} \|_{Q,2}, \mathcal{G}_{m,n,\gamma}, L_2(Q)) \lesssim O(m) \log(1/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1. \tag{G.110}$$

With this notation, we have

$$\sup_{|u - \tilde{u}| \leq \gamma} \| \mathbb{U}_n(u) - \mathbb{U}_n(\tilde{u}) \| = \sup_{g \in \mathcal{G}_{m,n,\gamma}} |\mathbb{G}_n g|,$$

and so to prove the asserted claim, we can apply the second part of Lemma 22 using the sequence of independent observations $(Z_i, U_i, v_i)_{i=1}^n$. Note that Lemma 22 does not require i.i.d. observations, and so it can be applied even though Z_i's are non-stochastic.

By (G.110), (H.119) is satisfied with $\omega = 1$, $J(m) = O(\sqrt{m})$, and $F_m = 2\widetilde{G}_{m,n}$. Note that

$$\max_{1 \leq i \leq n} F_m(Z_i, U_i, v_i) \lesssim_P M_m = \zeta_m \log n$$

since $\max_{1 \leq i \leq n} v_i \lesssim_P \log n$. Further,

$$
\sup_{g \in \mathcal{G}_{m,n,\gamma}} \frac{1}{n} \sum_{i=1}^{n} E[g(Z_i, U_i, v_i)^2] = \sup_{\alpha \in S^{m-1}, |u-\tilde{u}| \leq \gamma} \frac{1}{n} \sum_{i=1}^{n} E\Big[v_i^2 (\alpha' Z_i)^2 (\psi_i(u) - \psi_i(\tilde{u}))^2\Big]
$$

$$
= \sup_{\alpha \in S^{m-1}, |u-\tilde{u}| \leq \gamma} \frac{1}{n} \sum_{i=1}^{n} (\alpha' Z_i)^2 E\Big[(\psi_i(u) - \psi_i(\tilde{u}))^2\Big]
$$

$$
\leq \sup_{\alpha \in S^{m-1}} \frac{1}{n} \sum_{i=1}^{n} (\alpha' Z_i)^2 \gamma (1 - \gamma) \leq \varphi \gamma,
$$

where the second line holds because Z_i is non-stochastic, v_i is independent of U_i, and $E[v_i^2] = 1$, and the third line holds because $(\psi_i(u) - \psi_i(\tilde{u}))^2 =_d (|u - \tilde{u}| - 1\{U_i \leq |u - \tilde{u}|\})^2$. Moreover,

$$
\sup_{g \in \mathcal{G}_{m,n,\gamma}} \frac{1}{n} \sum_{i=1}^{n} E[g(Z_i, U_i, v_i)^4] = \sup_{\alpha \in S^{m-1}, |u-\tilde{u}| \leq \gamma} \frac{1}{n} \sum_{i=1}^{n} E\Big[v_i^4 (\alpha' Z_i)^4 (\psi_i(u) - \psi_i(\tilde{u}))^4\Big]
$$

$$
\lesssim \sup_{\alpha \in S^{m-1}, |u-\tilde{u}| \leq \gamma} \frac{1}{n} \sum_{i=1}^{n} (\alpha' Z_i)^4 E\Big[(\psi_i(u) - \psi_i(\tilde{u}))^4\Big]
$$

$$
\lesssim \sup_{\alpha \in S^{m-1}, |u-\tilde{u}| \leq \gamma} \frac{1}{n} \sum_{i=1}^{n} \zeta_m^2 (\alpha' Z_i)^2 E\Big[(\psi_i(u) - \psi_i(\tilde{u}))^2\Big]
$$

$$
\leq \sup_{\alpha \in S^{m-1}} \frac{1}{n} \sum_{i=1}^{n} \zeta_m^2 (\alpha' Z_i)^2 \gamma (1 - \gamma) \leq \zeta_m^2 \varphi \gamma,
$$

where we used the same arguments as above in addition to the facts that $E[v_i^4] \lesssim 1$, $\max_{1 \leq i \leq n} \|Z_i\| \leq \zeta_m$, and $|\psi_i(u) - \psi_i(\tilde{u})| \leq 1$. Substituting these bounds into the second part of Lemma 22 gives

$$
\sup_{g \in \mathcal{G}_{m,n,\gamma}} |\mathbb{G}_n g| \lesssim_P \sqrt{\gamma \varphi m \log n} + \sqrt{\frac{m^2 \zeta_m^2 \log^4 n}{n}}.
$$

This completes the proof of the lemma. $\qquad\square$

Lemma 16 (Infinite-Dimensional Approximation). *Let $G_n \colon \mathcal{U} \to \mathbb{R}^m$ be a zero-mean Gaussian process whose covariance structure is given by*

$$
E\left[G_n(u) G_n(\tilde{u})'\right] = \mathbb{E}_n[Z_i Z_i'](u \wedge \tilde{u} - u\tilde{u}), \text{ for all } u \text{ and } \tilde{u} \text{ in } \mathcal{U},
$$

where $(Z_i)_{i=1}^{n}$ is a non-stochastic sequence in $\mathbb{R}^m$. Then for any $\gamma \in (0, 1/2)$,

$$
\sup_{|u-\tilde{u}| \leq \gamma} \|G_n(u) - G_n(\tilde{u})\| \lesssim_P \sqrt{\varphi \gamma m \log(m/\gamma)},
$$

where $\varphi = \sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2]$.

Proof. We will use the following maximal inequality for Gaussian processes (Proposition A.2.7 in van der Vaart and Wellner [80]). Let $X = (X_t)_{t \in T}$ be a separable zero-mean Gaussian process indexed by a set T. Suppose that for some $K > \sigma(X) = \sup_{t \in T} \sigma(X_t)$ and $0 < \epsilon_0 \leq \sigma(X)$, we have

$$N(\epsilon, T, \rho) \leq \left(\frac{K}{\epsilon}\right)^V, \text{ for all } 0 < \epsilon < \epsilon_0, \tag{G.111}$$

where $N(\epsilon, T, \rho)$ is the covering number of T by ϵ-balls with respect to the standard deviation metric $\rho(t, t') = \sigma(X_t - X_{t'})$. Then there exists a universal constant D such that for every $\lambda \geq \sigma^2(X)(1 + \sqrt{V})/\epsilon_0$,

$$P\left(\sup_{t \in T} X_t > \lambda\right) \leq \left(\frac{DK\lambda}{\sqrt{V}\sigma^2(X)}\right)^V \bar{\Phi}(\lambda/\sigma(X)), \tag{G.112}$$

where $\bar{\Phi} = 1 - \Phi$, and Φ is the cumulative distribution function of a standard Gaussian random variable.

We apply this result to the zero-mean Gaussian process $X_n \colon S^{m-1} \times \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ defined as

$$X_{n,t} = \alpha'(G_n(u) - G_n(\tilde{u})), \quad t = (\alpha, u, \tilde{u}), \ \alpha \in S^{m-1}, \ |u - \tilde{u}| \leq \gamma,$$

since this process is such that $\sup_{t \in T} X_{n,t} = \sup_{|u - \tilde{u}| \leq \gamma} \|G_n(u) - G_n(\tilde{u})\|$. For this process, we have

$$\sigma(X_n) \leq \sqrt{\gamma \sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2]},$$

and (G.111) holds with

$$\epsilon_0 = \sigma(X_n), \quad K \lesssim \sqrt{\sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2]}, \quad \text{and } V \lesssim m.$$

Therefore, the result follows by setting

$$\lambda = C\sqrt{\gamma m \log(m/\gamma) \sup_{\alpha \in S^{m-1}} \mathbb{E}_n[(\alpha' Z_i)^2]},$$

where C is a sufficiently large constant, and using (G.112). This completes the proof of the lemma. $\qquad \square$

In what follows, as before, given a process G in $[\ell^\infty(\mathcal{U})]^m$ and its projection $G \circ \pi_j$, whose paths are step functions with at most 2^j steps, we shall identify the process $G \circ \pi_j$ with a random vector $G \circ \pi_j$ in $\mathbb{R}^{m2^j}$, when convenient. Analogously, given a random vector W in $\mathbb{R}^{m2^j}$ we identify it with a process W in $[\ell^\infty(\mathcal{U})]^m$, whose paths are step functions with at most 2^j steps.

Lemma 17. *(Construction of a Gaussian Process with a Pre-scribed Projection) Let $\mathcal{N}_j$ be a given random vector such that*

$$\mathcal{N}_j =_d \tilde{G} \circ \pi_j =: N(0, \Sigma_j),$$

where $\Sigma_j := \mathrm{Var}(\mathcal{N}_j)$ and $\tilde{G}$ is a zero-mean Gaussian process in $[\ell^\infty(\mathcal{U})]^m$ whose paths are a.s. uniformly continuous with respect to the Euclidian metric $|\cdot|$ on $\mathcal{U}$. There exists a zero-mean Gaussian process in $[\ell^\infty(\mathcal{U})]^m$, whose paths are a.s. uniformly continuous with respect to the Euclidian metric $|\cdot|$ on $\mathcal{U}$, such that

$$\mathcal{N}_j = G \circ \pi_j \ \text{and} \ G =_d \tilde{G} \ \text{in} \ [\ell^\infty(\mathcal{U})]^m.$$

Proof. Consider a vector $\tilde{G} \circ \pi_\ell$ for $\ell = j+1$. Then $\tilde{\mathcal{N}}_j = \tilde{G} \circ \pi_j$ is a subvector of $\tilde{G} \circ \pi_\ell = \tilde{\mathcal{N}}_\ell$. Denote the remaining components of $\tilde{\mathcal{N}}_\ell$ as $\tilde{\mathcal{N}}_{\ell \backslash j}$. We can construct an identically distributed copy $\mathcal{N}_\ell$ of $\tilde{\mathcal{N}}_\ell$ such that $\mathcal{N}_j$ is a subvector of $\mathcal{N}_\ell$. Indeed, we set $\mathcal{N}_\ell$ as a vector with components

$$\mathcal{N}_j \ \text{and} \ \mathcal{N}_{\ell \backslash j},$$

arranged in appropriate order, namely that $\mathcal{N}_\ell \circ \pi_j = \mathcal{N}_j$, where

$$\mathcal{N}_{\ell \backslash j} = \Sigma_{\ell \backslash j, j} \Sigma_{j,j}^{-1} \mathcal{N}_j + \eta_j,$$

where $\eta_j \perp \mathcal{N}_j$ and $\eta_j =_d N(0, \Sigma_{\ell \backslash j, \ell \backslash j} - \Sigma_{\ell \backslash j, j} \Sigma_{j,j}^{-1} \Sigma_{j, \ell \backslash j})$, where

$$\begin{pmatrix} \Sigma_{j,j} & \Sigma_{\ell \backslash j, j} \\ \Sigma_{j, \ell \backslash j} & \Sigma_{\ell \backslash j, \ell \backslash j} \end{pmatrix} := \mathrm{var} \begin{pmatrix} \tilde{N}_j \\ \tilde{N}_{\ell \backslash j} \end{pmatrix}.$$

Having constructed $\mathcal{N}_l = \mathcal{N}_{j+1}$, we can proceed using the same procedure to construct $\mathcal{N}_{j+2}$ from $\mathcal{N}_{j+1}$. Repeating this procedure, we obtain the whole sequence $(\mathcal{N}_l)_{l \geq j}$.

For each $l \geq j$, we then identify the vector $\mathcal{N}_\ell$ with a process $\mathcal{N}_\ell$ in $\ell^\infty(\mathcal{U})$, and define G as the pointwise limit of this process:

$$G(u) := \lim_{\ell \to \infty} \mathcal{N}_\ell(u) \ \text{for each} \ u \in \mathcal{U}_0,$$

where $\mathcal{U}_0 = (\cup_{j=1}^\infty \cup_{k=1}^{2^j} u_{kj}) \cap \mathcal{U}$ is a countable dense subset of $\mathcal{U}$. Note that the pointwise limit exists since by construction of $\{\pi_\ell\}$ and $\mathcal{U}_0$, for each $u \in \mathcal{U}_0$, we have that $\pi_\ell(u) = u$ for all $\ell \geq \ell(u)$, where $\ell(u)$ is a sufficiently large constant.

Next, we extend the process G from $\mathcal{U}_0$ to $\mathcal{U}$ as follows. By construction, $G_\ell = G \circ \pi_\ell =_d \tilde{G} \circ \pi_\ell$. Therefore, for each $\epsilon > 0$, there exists $\eta(\epsilon) > 0$ small enough such that

$$
\begin{aligned}
P\Big(\sup_{u,\tilde{u}\in\mathcal{U}_0:|u-\tilde{u}|\leq\eta(\epsilon)} \|G(u) - G(\tilde{u})\| \geq \epsilon \Big) &\leq P\Big(\sup_{|u-\tilde{u}|\leq\eta(\epsilon)} \sup_k \|G \circ \pi_k(u) - G \circ \pi_k(\tilde{u})\| \geq \epsilon \Big) \\
&\leq P\Big(\sup_{|u-\tilde{u}|\leq\eta(\epsilon)} \sup_k \|\tilde{G} \circ \pi_k(u) - \tilde{G} \circ \pi_k(\tilde{u})\| \geq \epsilon \Big) \\
&\leq P\Big(\sup_{|u-\tilde{u}|\leq\eta(\epsilon)} \|\tilde{G}(u) - \tilde{G}(\tilde{u})\| \geq \epsilon \Big) \leq \epsilon,
\end{aligned}
$$

where the last line holds because $\sup_{|u-\tilde{u}|\leq\eta} \|\tilde{G}(u) - \tilde{G}(\tilde{u})\| \to 0$ as $\eta \to 0$ almost surely and thus also in probability, by a.s. continuity of sample paths of $\tilde{G}$. Setting $\epsilon = 2^{-m}$ for each $m \in \mathbb{N}$ in the above display, and summing the resulting inequalities over m, we get a finite number on the right-hand side. Hence, by the Borel-Cantelli lemma, $|G(u) - G(\tilde{u})| \leq 2^{-m}$ for all $|u - \tilde{u}| \leq \eta(2^{-m})$ for all sufficiently large m almost surely. This implies that the sample path of the process G is uniformly continuous on $\mathcal{U}_0$ almost surely, and so we can extend the process by continuity to a process $\{G(u) \colon u \in \mathcal{U}\}$ such that its sample path is uniformly continuous almost surely.

Finally, in order to show that the law of G is equal to the law of $\tilde{G}$ in $[\ell^\infty(\mathcal{U})]^m$, it suffices to demonstrate that

$$
E[g(G)] = E[g(\tilde{G})] \text{ for all } g\colon [\ell^\infty(\mathcal{U})]^m \to \mathbb{R} \text{ with } |g(z) - g(\tilde{z})| \leq \sup_{u\in\mathcal{U}} \|z(u) - \tilde{z}(u)\| \wedge 1.
$$

To do so, note that

$$
\begin{aligned}
\Big| E[g(G)] - E[g(\tilde{G})] \Big| &\leq \Big| E[g(G \circ \pi_\ell)] - E[g(\tilde{G} \circ \pi_\ell)] \Big| \\
&\quad + E\Big[\sup_{u\in\mathcal{U}} \|G \circ \pi_\ell(u) - G(u)\| \wedge 1 \Big] \\
&\quad + E\Big[\sup_{u\in\mathcal{U}} \|\tilde{G} \circ \pi_\ell(u) - \tilde{G}(u)\| \wedge 1 \Big] \to 0 \text{ as } \ell \to \infty.
\end{aligned}
$$

Indeed, the first term on the right-hand side converges to zero by construction, and the second and third terms converge to zero by the dominated convergence theorem since

$$
G \circ \pi_\ell \to G \text{ and } \tilde{G} \circ \pi_\ell \to \tilde{G} \text{ in } [\ell^\infty(\mathcal{U})]^m \text{ as } \ell \to \infty \text{ a.s.,}
$$

holding due to a.s. uniform continuity of sample paths of G and $\tilde{G}$. This completes the proof of the lemma. $\qquad\square$

APPENDIX H. TECHNICAL LEMMAS ON BOUNDING EMPIRICAL ERRORS

H.1. **Some Preliminary Lemmas.**

Lemma 18. *Under Condition S, we have*

$$\|J(u) - \widetilde{J}(u)\| \lesssim m^{-\kappa} = o(1),$$

uniformly over $u \in \mathcal{U}$, *where* $\widetilde{J}(u) = E[f_{Y|X}(Z'\beta(u)|X)ZZ']$ *and* $J(u)$ *is defined in* (2.10).

Proof. Note that $\|J(u) - \widetilde{J}(u)\| = \sup_{\alpha \in S^{m-1}} |\alpha'(J(u) - \widetilde{J}(u))\alpha|$. In addition,

$$|\alpha'(J(u) - \widetilde{J}(u))\alpha| = E\left[\left|f_{Y|X}(Z'\beta(u) + R(u, X)|X) - f_{Y|X}(Z'\beta(u)|X)\right| \cdot (Z'\alpha)^2\right]$$
$$\lesssim E\left[|R(u, X)| \cdot (Z'\alpha)^2\right] \lesssim m^{-\kappa}$$

uniformly over $u \in \mathcal{U}$ and $\alpha \in S^{m-1}$, where we used Condition S in the second line. The asserted claim follows. $\square$

Lemma 19 (Auxiliary Matrix). *Suppose that Condition S holds. Then*

$$\|J(u_2) - J(u_1)\| \lesssim |u_2 - u_1|, \quad \text{uniformly over } u_1 \text{ and } u_2 \text{ in } \mathcal{U}.$$

In addition,

$$\left|\frac{z'(J^{-1}(u_2) - J^{-1}(u_1))\mathbb{U}(u_2)}{\sqrt{u_1(1-u_1)z'J^{-1}(u_1)\Sigma J^{-1}(u_1)z}}\right| \lesssim_P |u_2 - u_1|\sqrt{m}$$

uniformly over $z \in \{Z(x) \colon x \in \mathcal{X}\}$ *and* $u_1, u_2 \in \mathcal{U}$ *for* $\mathbb{U}(u)$ *defined in* (3.16).

Proof. Recall that $J(u) = E\left[f_{Y|X}(Q(u, X)|X)ZZ'\right]$ for any $u \in \mathcal{U}$. Moreover, we have $\sup_{u \in \mathcal{U}} \|J^{-1}(u)\mathbb{U}(u)\| \lesssim_P \sqrt{m}$ by Lemma 32 since all eigenvalues of $J(u)$ are bounded below from zero uniformly over $u \in \mathcal{U}$. In addition, using the matrix identity $A^{-1} - B^{-1} = B^{-1}(B - A)A^{-1}$ with $A = J(u_2)$ and $B = J(u_1)$ gives

$$J^{-1}(u_2) - J^{-1}(u_1) = J^{-1}(u_1)(J(u_1) - J(u_2))J^{-1}(u_2).$$

Also, since $|f_{Y|X}(Q(u_2, x)|x) - f_{Y|X}(Q(u_1, x)|x)| \lesssim |u_2 - u_1|$ uniformly over $u_1, u_2 \in \mathcal{U}$ and $x \in \mathcal{X}$ by Lemma 20, it follows from the same argument as that used in the proof of Lemma 18 that

$$\|J(u_2) - J(u_1)\| \lesssim |u_2 - u_1|, \quad \text{uniformly over } u_1, u_2 \in \mathcal{U},$$

which gives the first asserted claim. Further,

$$\left| \frac{z'(J^{-1}(u_2) - J^{-1}(u_1))\mathbb{U}(u_2)}{\sqrt{u_1(1-u_1)z'J^{-1}(u_1)\Sigma J^{-1}(u_1)z}} \right|$$

$$= \left| \frac{z'J^{-1}(u_1)}{\sqrt{u_1(1-u_1)z'J^{-1}(u_1)\Sigma J^{-1}(u_1)z}}(J(u_1) - J(u_2))J^{-1}(u_2)\mathbb{U}(u_2) \right|$$

$$\lesssim \frac{\|z'J^{-1}(u_1)\|}{\sqrt{z'J^{-1}(u_1)\Sigma J^{-1}(u_1)z}} \cdot \left\| J(u_1) - J(u_2) \right\| \cdot \left\| J^{-1}(u_2)\mathbb{U}(u_2) \right\| \lesssim_P |u_2 - u_1|\sqrt{m},$$

uniformly over $u_1, u_2 \in \mathcal{U}$, where in the third line, we used the fact that $\mathcal{U} \subset (0,1)$ is compact to show that $u_1(1 - u_1)$ is bounded away from zero uniformly over $u_1 \in \mathcal{U}$ and also the fact that all eigenvalues of the matrix Σ are bounded away from zero to show that

$$\frac{\|z'J^{-1}(u_1)\|}{\sqrt{z'J^{-1}(u_1)\Sigma J^{-1}(u_1)z}} \lesssim 1$$

uniformly over $u_1 \in \mathcal{U}$. This gives the second asserted claim and completes the proof of the lemma. $\qquad\square$

Lemma 20. *Suppose that Condition S holds. Then*

$$|u_2 - u_1| \lesssim |Q(u_2, x) - Q(u_1, x)| \lesssim |u_2 - u_1|, \tag{H.113}$$

$$|f_{Y|X}(Q(u_2, x)|x) - f_{Y|X}(Q(u_1, x)|x)| \lesssim |u_2 - u_1|, \tag{H.114}$$

uniformly over $u_1, u_2 \in \mathcal{U}$ and $x \in \mathcal{X}$. In addition,

$$\left| \frac{\partial^2 Q(u, x)}{\partial u^2} \right| \lesssim 1 \tag{H.115}$$

uniformly over $u \in \mathcal{U}$ and $x \in \mathcal{X}$.

Proof. Since $U|X \sim \text{Uniform}(0, 1)$, it follows that for all $u \in \mathcal{U}$ and $x \in \mathcal{X}$, we have

$$u = \int_{-\infty}^{Q(u,x)} f_{Y|X}(y|x)dy,$$

and so

$$\frac{\partial Q(u, x)}{\partial u} = \frac{1}{f_{Y|X}(Q(u, x)|x)}. \tag{H.116}$$

Hence,

$$\frac{\partial^2 Q(u, x)}{\partial u^2} = -\frac{1}{f_{Y|X}^2(Q(u, x)|x)} \cdot f'_{Y|X}(Q(u, x)|x) \cdot \frac{\partial Q(u, x)}{\partial u} = -\frac{f'_{Y|X}(Q(u, x)|x)}{f_{Y|X}^3(Q(u, x)|x)},$$

where $f'_{Y|X}(y|x)$ denotes the derivative of the function $y \mapsto f_{Y|X}(y|x)$. Hence, it follows from Condition S that (H.115) holds uniformly over $u \in \mathcal{U}$.

Also, combining Condition S and (H.116) shows that $1 \lesssim \partial Q(u,x)/\partial u \lesssim 1$ uniformly over $u \in \mathcal{U}$ and $x \in \mathcal{X}$, and since $\mathcal{U} \subset (0,1)$ is a connected compact set (that is, closed interval), this implies that (H.113) holds uniformly over $u_1, u_2 \in \mathcal{U}$ and $x \in \mathcal{X}$.

Finally, combining Condition S and (H.113) shows that (H.114) hold uniformly over $u \in \mathcal{U}$ and $x \in \mathcal{X}$. This completes the proof of the lemma. $\qquad\square$

H.2. **Maximal Inequalities.** In this section we derive some maximal inequalities that are useful to prove main results of the paper. Let $Z_1, \ldots, Z_n$ be a sequence of independent random variables taking values in some set $\mathcal{Z}$. Let $\mathcal{F}$ be a class of functions defined on $\mathcal{Z}$, and let

$$F(z) \geq \sup_{f \in \mathcal{F}} |f(z)|, \quad z \in \mathcal{Z},$$

be an envelope of $\mathcal{F}$. Let $\mathbb{P}_n$ be the empirical measure corresponding to the sequence $Z_1, \ldots, Z_n$, and let

$$\mathbb{G}_n(f) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \Big(f(Z_i) - E[f(Z_i)] \Big), \quad f \in \mathcal{F},$$

be the corresponding empirical process on $\mathcal{F}$. For a probability measure Q and a constant $p > 1$, such that $\|F\|_{Q,p} > 0$, we use $N(\varepsilon\|F\|_{Q,p}, \mathcal{F}, L_p(Q))$ to denote the minimal number of $L_p(Q)$-balls of radius $\varepsilon\|F\|_{Q,p}$ needed to cover $\mathcal{F}$. Following literature, we refer to $\sup_Q \log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q))$, where the supremum is taken over all finitely-discrete probability measures Q, as a uniform entropy number of $\mathcal{F}$; see Dudley [36] for details of the definitions.

Below we derive some bounds on $\sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)|$. Importantly, our bounds do not require that the random variables Z_i are identically distributed, and it will be sufficient to assume only that these random variables are independent. In addition, the function class $\mathcal{F}$ will actually be allowed to depend on the sample size n via the sequence $(m_n)_{n \geq 1}$, that is, we consider the case where $\mathcal{F} = \mathcal{F}_m$ and $m = m_n$.

Since Z_i's are not necessarily identically distributed, we will use $E[f^2]$ and $E[f^4]$ to denote $n^{-1} \sum_{i=1}^{n} E[f(Z_i)^2]$ and $n^{-1} \sum_{i=1}^{n} E[f(Z_i)^4]$, respectively. Further, we will say that the covering numbers $N(\epsilon\|F\|_{\mathbb{P}_n,2}, \mathcal{F}, L_2(\mathbb{P}_n))$, corresponding to the empirical measure $\mathbb{P}_n$, satisfy the monotonicity hypotheses if

$$N(\epsilon\|F\|_{\mathbb{P}_n,2}, \mathcal{F}, L_2(\mathbb{P}_n)) \leq n(\epsilon, \mathcal{F}, \mathbb{P}_n) \text{ for all } 0 < \epsilon \leq 1,$$

where $n(\epsilon, \mathcal{F}, \mathbb{P}_n)$ is such that (i) $\epsilon \mapsto n(\epsilon, \mathcal{F}, \mathbb{P}_n)$ is decreasing, (ii) $\epsilon \mapsto \epsilon\sqrt{\log n(\epsilon, \mathcal{F}, \mathbb{P}_n)}$ is increasing, and (iii) $\epsilon\sqrt{\log n(\epsilon, \mathcal{F}, \mathbb{P}_n)} \to 0$ as $\epsilon \to 0$. Also, we define $\rho(\mathcal{F}, \mathbb{P}_n) =$

$\sup_{f \in \mathcal{F}}(\|f\|_{\mathbb{P}_n,2}/\|F\|_{\mathbb{P}_n,2})$. In what follows, we refer to any function $x \colon \mathcal{Z}^n \mapsto \mathbb{R}$ as k-sub-exchangeable if for any $v, w \in \mathcal{Z}^n$ and any vectors $\tilde{v}, \tilde{w}$ created by the pairwise exchange of some components in v with the corresponding components in w, we have that $x(\tilde{v}) \vee x(\tilde{w}) \geq [x(v) \vee x(w)]/k$.

Lemma 21 (Exponential inequality for separable empirical process)**.** *Consider the setting specified above. Suppose that the empirical process $\{\mathbb{G}_n(f), f \in \mathcal{F}\}$ is separable. Also, suppose that the covering numbers $N(\epsilon \|F\|_{\mathbb{P}_n,2}, \mathcal{F}, L_2(\mathbb{P}_n))$ satisfy the monotonicity hypotheses. Further, let $K > 1$ and $\tau \in (0,1)$ be constants, and $e_n(\mathcal{F}, \mathbb{P}_n) = e_n(\mathcal{F}, Z_1, \ldots, Z_n)$ be a k-sub-exchangeable random variable, such that*

$$\|F\|_{\mathbb{P}_n,2} \int_0^{\rho(\mathcal{F},\mathbb{P}_n)/4} \sqrt{\log n(\epsilon, \mathcal{F}, \mathbb{P}_n)} d\epsilon \leq e_n(\mathcal{F}, \mathbb{P}_n) \qquad \text{(H.117)}$$

and

$$\sup_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^n \left(E[f(Z_i)^2] - (E[f(Z_i)])^2 \right) \leq \frac{\tau}{2} (4kcKe_n(\mathcal{F}, \mathbb{P}_n))^2 \qquad \text{(H.118)}$$

for some universal constant $c > 1$. Then

$$\mathbb{P}\left\{ \sup_{f \in \mathcal{F}} |\mathbb{G}_n(f)| \geq 4kcKe_n(\mathcal{F}, \mathbb{P}_n) \right\}$$

$$\leq \frac{4}{\tau} E_{\mathbb{P}} \left(\left[\int_0^{\rho(\mathcal{F},\mathbb{P}_n)/2} \epsilon^{-1} n(\epsilon, \mathcal{F}, \mathbb{P}_n)^{-\{K^2-1\}} d\epsilon \right] \wedge 1 \right) + \tau.$$

Proof. See Lemma 18 in Belloni and Chernozhukov [7] and note that the proof there does not require identically distributed Z_i's since only independence among Z_i's is used. $\qquad\square$

The next lemma establishes new maximal inequalities, which are most useful for the purposes of this paper.

Lemma 22. *Suppose that for all $n \geq 1$ and $0 < \epsilon \leq 1$, we have*

$$N(\epsilon \|F_m\|_{\mathbb{P}_n,2}, \mathcal{F}_m, L_2(\mathbb{P}_n)) \leq (\omega/\epsilon)^{J(m)^2} \qquad \text{(H.119)}$$

for some $\omega = \omega_n$ such that $\log \omega \lesssim \log n$ and some $J(m) = J(m_n)$ such that $J(m) \geq 1$, where $F_m \geq \sup_{f \in \mathcal{F}_m} |f|$ is an envelope of $\mathcal{F}_m$. Let M_m be a random variable that satisfies $\max_{1 \leq i \leq n} F_m(Z_i) \lesssim_P M_m$. Then we have the following results.

1. A Maximal Inequality Based on Entropy and Moments:

$$\sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f)|$$

$$\lesssim_P J(m) \sup_{f \in \mathcal{F}_m} \left(E[f^2] + J(m)\left(\mathbb{E}_n[f^4] + E[f^4] + n^{-1}\mathbb{E}_n[F_m^4]\right)^{1/2}\left(\frac{\log n}{n}\right)^{1/2}\right)^{1/2} \log^{1/2} n.$$

2. A Maximal Inequality Based on Entropy, Moments, and Random Extremum:

$$\sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f)| \lesssim_P J(m) \left(\sup_{f \in \mathcal{F}_m} \left(E[f^2] + J(m)\left(\frac{E[f^4]\log n}{n}\right)^{1/2}\right) + J(m)^2 \frac{M_m^2 \log n}{n}\right)^{1/2} \log^{1/2} n.$$

3. A Maximal Inequality Based on Entropy, Moments, and Non-Random Extremum:

$$\sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f)| \lesssim_P J(m) \left(\sup_{f \in \mathcal{F}_m} E[f^2] + J(m)^2 \frac{\bar{F}_m^2 \log n}{n}\right)^{1/2} \log^{1/2} n,$$

if there exists a non-stochastic constant $\bar{F}_m = \bar{F}_{m,n} \in \mathbb{R}$ such that $\sup_{z \in \mathcal{Z}} F_m(z) \leq \bar{F}_m$.

Proof. Note that if the condition (H.119) holds for some ω, it also holds with ω replaced by any $\omega' > \omega$. Therefore, it is without loss of generality to assume that $\omega \geq 1$, which we do. Also, since $\log \omega \lesssim \log n$, there exists a constant C such that $\log \omega \leq C \log n$. This constant will be used later in the proof.

We divide the proof into three steps. Step 1 consists of the main argument, Step 2 is an application of Lemma 21, and Step 3 contains some auxiliary calculations.

Step 1 (Main Argument). We first prove the first asserted claim. By Step 2 below, we have

$$\sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f)| \lesssim_P J(m) \sup_{f \in \mathcal{F}_m} \left(\mathbb{E}_n[f^2] + E[f^2] + n^{-1}\mathbb{E}_n[F_m^2]\right)^{1/2} \log^{1/2} n. \tag{H.120}$$

Also by Step 2,

$$\sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f^2)| \lesssim_P J(m) \sup_{f \in \mathcal{F}_m} \left(\mathbb{E}_n[f^4] + E[f^4] + n^{-1}\mathbb{E}_n[F_m^4]\right)^{1/2} \log^{1/2} n. \tag{H.121}$$

Now, by the triangle inequality and (H.121),

$$\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f^2] \leq \sup_{f \in \mathcal{F}_m} E[f^2] + n^{-1/2} \sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f^2)| \tag{H.122}$$

$$\lesssim_P \sup_{f \in \mathcal{F}_m} E[f^2] + n^{-1/2}J(m) \sup_{f \in \mathcal{F}_m} \left(\mathbb{E}_n[f^4] + E[f^4] + n^{-1}\mathbb{E}_n[F_m^4]\right)^{1/2} \log^{1/2} n. \tag{H.123}$$

Substituting this bound into (H.120) and using inequalities $\mathbb{E}_n[F_m^2] \leq (\mathbb{E}_n[F_m^4])^{1/2}$, $J(m) \geq 1$, and $\log^{1/2} n \geq 1$ gives the first asserted claim.

Next, we prove the second asserted claim. Using (H.122)-(H.123) and noting that $\mathbb{E}_n[f^4] \lesssim_P \mathbb{E}_n[f^2] M_m^2$ uniformly over $f \in \mathcal{F}_m$ and that $\mathbb{E}_n[F_m^4] \leq M_n^4$ gives

$$\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f^2] \lesssim_P \sup_{f \in \mathcal{F}_m} \left(E[f^2] + J(m) \left(\frac{(E[f^4] + M_n^4/n) \log n}{n} \right)^{1/2} \right)$$

$$+ J(m) \left(\frac{M_m^2 \log n}{n} \right)^{1/2} \sup_{f \in \mathcal{F}_m} \left(\mathbb{E}_n[f^2] \right)^{1/2}.$$

Since for positive numbers a, c, and x, $x \leq a + c|x|^{1/2}$ implies that $x \leq 4a + 4c^2$, this inequality gives

$$\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f^2] \lesssim_P \sup_{f \in \mathcal{F}_m} \left(E[f^2] + J(m) \left(\frac{(E[f^4] + M_n^4/n) \log n}{n} \right)^{1/2} \right) + J(m)^2 \frac{M_m^2 \log n}{n}$$

$$\lesssim_P \sup_{f \in \mathcal{F}_m} \left(E[f^2] + J(m) \left(\frac{E[f^4] \log n}{n} \right)^{1/2} \right) + J(m)^2 \frac{M_m^2 \log n}{n}.$$

Substituting this bound into (H.120) and noting that $\mathbb{E}_n[F_m^2] \leq M_n^2$ gives the second asserted claim.

Finally, the third asserted claim follows from the second one by substituting $\bar{F}_m$ instead of M_m and using the inequality $E[f^4] \leq \bar{F}_m E[f^2]$, which holds for any $f \in \mathcal{F}_m$.

Step 2 (Applying Lemma 21). Here we prove (H.120) and (H.121). In fact, note that by Lemma 24, (H.119) implies that

$$N(\epsilon \|F_m^2\|_{\mathbb{P}_n,2}, \mathcal{F}_m^2, L_2(\mathbb{P}_n)) \leq (2\omega/\epsilon)^{2J(m)^2},$$

so that (H.121) follows from the same argument as that used for (H.120), so we only prove (H.120). To do so, we apply Lemma 21 to $\mathcal{F} = \mathcal{F}_m$ with

$$\tau = \tau_m = \frac{4}{J(m)^2(K^2 - 1)}$$

for some large constant K to be set later, and

$$e_n(\mathcal{F}_m, \mathbb{P}_n) = \sqrt{C + 2} J(m) \sup_{f \in \mathcal{F}_m} \left((\mathbb{E}_n[f^2])^{1/2} + (E[f^2])^{1/2} + (n^{-1} \mathbb{E}_n[F_m^2])^{1/2} \right) \log^{1/2} n,$$

where C is the constant appearing in the beginning of the proof.

We first verify that conditions of Lemma 21 hold for all n large enough. Note that by (H.119), the covering numbers $N(\epsilon \|F_m\|_{\mathbb{P}_n,2}, \mathcal{F}_m, L_2(\mathbb{P}_n))$ satisfy the monotonicity hypotheses for $n(\epsilon, \mathcal{F}_m, \mathbb{P}_n) = (\omega/\epsilon)^{J(m)^2}$. Next, by Step 3 below, the function $(Z_1, \ldots, Z_n) \mapsto \sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P}_n,2}$ is $\sqrt{2}$-sub-exchangeable, and so is $e_n(\mathcal{F}_m, \mathbb{P}_n)$. Also, denoting $\tilde{\rho} = n^{-1/2} \vee$

$\rho(\mathcal{F}_m, \mathbb{P}_n)/4$, we have

$$\|F_m\|_{\mathbb{P}_n,2} \int_0^{\rho(\mathcal{F}_m,\mathbb{P}_n)/4} \sqrt{\log n(\epsilon, \mathcal{F}_m, \mathbb{P}_n)} d\epsilon \leq \|F_m\|_{\mathbb{P}_n,2} \int_0^{\rho(\mathcal{F}_m,\mathbb{P}_n)/4} J(m)\sqrt{\log(\omega/\epsilon)} d\epsilon$$

$$\leq \|F_m\|_{\mathbb{P}_n,2} \int_0^{\tilde{\rho}} J(m)\sqrt{\log(\omega/\epsilon)} d\epsilon$$

$$\leq \|F_m\|_{\mathbb{P}_n,2} J(m)\tilde{\rho}\sqrt{(C+2)\log n}$$

$$\leq e_n(\mathcal{F}_m, \mathbb{P}_n),$$

where the third line follows from

$$\int_0^{\tilde{\rho}} \sqrt{\log(\omega/\epsilon)} d\epsilon \leq \left(\int_0^{\tilde{\rho}} 1 d\epsilon\right)^{1/2} \left(\int_0^{\tilde{\rho}} \log(\omega/\epsilon) d\epsilon\right)^{1/2}$$

$$= \tilde{\rho}^{1/2} \left(\int_0^{\tilde{\rho}} (\log \omega + \log(1/\epsilon)) d\epsilon\right)^{1/2}$$

$$= \tilde{\rho}^{1/2} \left(\tilde{\rho}\log \omega + \tilde{\rho}\log(1/\tilde{\rho}) + \tilde{\rho}\right)^{1/2} \leq \tilde{\rho}\sqrt{(C+2)\log n},$$

since $\tilde{\rho} \geq n^{-1/2}$ and $\log \omega \leq C \log n$. Hence, condition (H.117) is satisfied. Moreover, the condition (H.118) is satisfies by our choice of $e_n(\mathcal{F}_m, \mathbb{P}_n)$ for sufficiently large n.

Therefore, all conditions of Lemma 21 are satisfied and its application gives

$$\mathbb{P}\left(\sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f)| > 4\sqrt{2}cKe_n(\mathcal{F}_m, \mathbb{P}_n)\right) \leq \frac{4}{\tau_m} \int_0^{1/2} \frac{\omega^{-J(m)^2(K^2-1)}}{\epsilon^{1-J(m)^2(K^2-1)}} d\epsilon + \tau_m$$

$$\leq \frac{4}{\tau_m} \frac{(1/(2\omega))^{J(m)^2[K^2-1]}}{J(m)^2[K^2-1]} + \tau_m$$

$$= \left(\frac{1}{2\omega}\right)^{J(m)^2[K^2-1]} + \frac{4}{J(m)^2(K^2-1)}, \quad \text{(H.124)}$$

where we used the inequality $\rho(\mathcal{F}_m, \mathbb{P}_n) \leq 1$. Since $\omega \geq 1$ and $J(m) \geq 1$, the expression in (H.124) can be made arbitrarily small by setting K sufficiently large. Hence, (H.120) follows.

Step 3 (Auxiliary calculations). Here we establish that the function $(Z_1, \ldots, Z_n) \mapsto \sup_{f \in \mathcal{F}_m} (\mathbb{E}_n[f(Z_i)^2 + F_m(Z_i)^2/n])^{1/2}$ mapping $\mathcal{Z}^n$ into $\mathbb{R}$ is $\sqrt{2}$-sub-exchangeable. Indeed, let $Z = (Z_1, \ldots, Z_n)$ and $Y = (Y_1, \ldots, Y_n)$ be two elements of $\mathcal{Z}^n$ and define $\tilde{Z}$ and $\tilde{Y}$ by

exchanging some components in Z with corresponding components in Y. Then

$$2\Big(\sup_{f\in\mathcal{F}_m} (\mathbb{E}_n[f(\tilde{Z}_i)^2 + F_m(\tilde{Z}_i)^2/n]) \vee \sup_{f\in\mathcal{F}_m} (\mathbb{E}_n[f(\tilde{Y}_i)^2 + F_m(\tilde{Y}_i)^2/n])\Big)$$

$$\geq \sup_{f\in\mathcal{F}_m} (\mathbb{E}_n[f(\tilde{Z}_i)^2 + F_m(\tilde{Z}_i)^2/n]) + \sup_{f\in\mathcal{F}_m} (\mathbb{E}_n[f(\tilde{Y}_i)^2 + F_m(\tilde{Y}_i)^2/n])$$

$$\geq \sup_{f\in\mathcal{F}_m} \Big((\mathbb{E}_n[f(\tilde{Z}_i)^2 + F_m(\tilde{Z}_i)^2/n]) + (\mathbb{E}_n[f(\tilde{Y}_i)^2 + F_m(\tilde{Y}_i)^2/n]) \Big)$$

$$\geq \sup_{f\in\mathcal{F}_m} \Big((\mathbb{E}_n[f(Z_i)^2 + F_m(Z_i)^2/n]) + (\mathbb{E}_n[f(Y_i)^2 + F_m(Y_i)^2/n]) \Big)$$

$$\geq \sup_{f\in\mathcal{F}_m} (\mathbb{E}_n[f(Z_i)^2 + F_m(Z_i)^2/n]) \vee \sup_{f\in\mathcal{F}_m} (\mathbb{E}_n[f(Y_i)^2 + F_m(Y_i)^2/n]).$$

This gives the asserted claim and completes the proof of the lemma. $\qquad\square$

H.3. **Uniform Entropy Numbers.**

Lemma 23 (Uniform Entropy of VC classes). *Suppose that the class of functions $\mathcal{F}$ has VC index V and an envelope F. Then for some absolute constants c and C,*

$$\sup_Q N(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq (C/\epsilon)^{cV}, \quad \text{for all } 0 < \epsilon \leq 1,$$

where Q ranges over all finitely-discrete probabilities measures.

Proof. The bound follows from Theorem 2.6.7 in van der Vaart and Wellner [80]. $\qquad\square$

Lemma 24 (Uniform entropy for products and sums). *Let $\mathcal{F}$ and $\mathcal{G}$ be two classes of functions with envelopes F and G respectively. Then the uniform entropy numbers of $\mathcal{F}\mathcal{G} = \{fg\colon f\in\mathcal{F},\ g\in\mathcal{G}\}$ satisfy*

$$\sup_Q \log N(\epsilon\|FG\|_{Q,2}, \mathcal{F}\mathcal{G}, L_2(Q))$$

$$\leq \sup_Q \log N\Big(\frac{\epsilon\|F\|_{Q,2}}{2}, \mathcal{F}, L_2(Q)\Big) + \sup_Q \log N\Big(\frac{\epsilon\|G\|_{Q,2}}{2}, \mathcal{G}, L_2(Q)\Big)$$

for all $\epsilon > 0$. Also, the uniform entropy numbers of $\mathcal{F}+\mathcal{G} = \{f+g\colon f\in\mathcal{F},\ g\in\mathcal{G}\}$ satisfy

$$\sup_Q \log N(\epsilon\|F+G\|_{Q,2}, \mathcal{F}+\mathcal{G}, L_2(Q))$$

$$\leq \sup_Q \log N\Big(\frac{\epsilon\|F\|_{Q,2}}{2}, \mathcal{F}, L_2(Q)\Big) + \sup_Q \log N\Big(\frac{\epsilon\|G\|_{Q,2}}{2}, \mathcal{G}, L_2(Q)\Big)$$

for all $\epsilon > 0$. In both cases, Q ranges over all finitely-discrete probability measures.

Proof. The result is proven in the proof of Theorem 3 of Andrews [1]. $\qquad\square$

Lemma 25. *For any $r > 0$, define the class of functions*

$$\mathcal{F}_{m,n} = \Big\{(Z,Y) \mapsto (\alpha'Z) \cdot \Big(1\{Y \leq Z'\beta\} - 1\{Y \leq Z'\beta(u)\}\Big):$$

$$u \in \mathcal{U}, \ \alpha \in S^{m-1}, \ \|\beta - \beta(u)\| \leq r\Big\},$$

mapping $\mathbb{R}^m \times \mathbb{R}$ into $\mathbb{R}$, and let $F_{m,n}(Z,Y) = \|Z\|$ be its envelope. The uniform entropy numbers of $\mathcal{F}_{m,n}$ satisfy

$$\sup_Q \log N(\epsilon\|F_{m,n}\|_{Q,2}, \mathcal{F}_{m,n}, L_2(Q)) \lesssim m\log(1/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1.$$

Proof. Consider function classes

$$\mathcal{W}_1 = \Big\{(Z,Y) \mapsto \alpha'Z \colon \alpha \in \mathbb{R}^m\Big\} \text{ and } \mathcal{V}_1 = \Big\{(Z,Y) \mapsto 1\{Y \leq Z'\beta\} \colon \beta \in \mathbb{R}^m\Big\}.$$

Their VC indices are bounded by $m+2$ by Lemmas 2.6.15 and 2.6.18 in van der Vaart and Wellner [80]. Hence, since any $f \in \mathcal{F}_{m,n}$ can be written as $f = g \cdot (v - p)$ for $g \in \mathcal{W}_1$, $v \in \mathcal{V}_1$, and $p \in \mathcal{V}_1$, the asserted claim follows from Lemmas 23 and 24. $\square$

Lemma 26. *For any $r, h > 0$, define the class of functions*

$$\mathcal{H}_{m,n} = \Big\{(Z,Y) \mapsto (\alpha'Z)^2 \cdot 1\{|Y - Z'\beta| \leq h\} \colon u \in \mathcal{U}, \ \alpha \in S^{m-1}, \ \|\beta - \beta(u)\| \leq r\Big\},$$

mapping $\mathbb{R}^m \times \mathbb{R}$ into $\mathbb{R}$, and let $H_{m,n}(Z,Y) = \|Z\|$ be its envelope. The uniform entropy numbers of $\mathcal{H}_{m,n}$ satisfy

$$\sup_Q \log N(\epsilon\|H_{m,n}\|_{Q,2}, \mathcal{H}_{m,n}, L_2(Q)) \lesssim m\log(1/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1.$$

Proof. Consider function classess

$$\mathcal{W}_2 = \Big\{(Z,Y) \mapsto 1\{Y - Z'\beta - a \leq 0\} \colon \beta \in \mathbb{R}^m, \ a \in \mathbb{R}\Big\},$$

$$\mathcal{V}_2 = \Big\{(Z,Y) \mapsto 1\{Y - Z'\beta - a < 0\} \colon \beta \in \mathbb{R}^m, \ a \in \mathbb{R}\Big\}.$$

Their VC indices are $m+3$ by Lemmas 2.6.15 and 2.6.18 in van der Vaart and Wellner [80]. Hence, since any $f \in \mathcal{H}_{m,n}$ can be written as $f = g^2 \cdot (v - p)$ for $g \in \mathcal{W}_1$, $v \in \mathcal{W}_2$, and $p \in \mathcal{V}_2$, the asserted claim follows from Lemmas 23 and 24. $\square$

Lemma 27. *Define the class of functions*

$$\mathcal{G}_{m,n} = \Big\{(Z,Y) \mapsto (\alpha'Z) \cdot \Big(1\{Y \leq Z'\beta(u)\} - u\Big) \colon u \in \mathcal{U}, \ \alpha \in S^{m-1}\Big\},$$

mapping $\mathbb{R}^m \times \mathbb{R}$ *into* $\mathbb{R}$, *and let* $G_{m,n}(Z,Y) = \|Z\|$ *be its envelope. The uniform entropy numbers of* $\mathcal{G}_{m,n}$ *satisfy*

$$\sup_Q \log N(\epsilon \|G_{m,n}\|_{Q,2}, \mathcal{G}_{m,n}, L_2(Q)) \lesssim m \log(1/\epsilon), \quad \textit{uniformly over } 0 < \epsilon \leq 1.$$

Proof. Consider the function class

$$\mathcal{W}_3 = \Big\{ (Z,Y) \mapsto u \colon u \in \mathcal{U} \Big\}.$$

Its VC index is $O(1)$. Hence, since any function $f \in \mathcal{G}_{m,n}$ can be written as $f = g \cdot (v - p)$ for $g \in \mathcal{W}_1$, $v \in \mathcal{V}_1$, and $p \in \mathcal{W}_3$, the asserted claim follows from Lemmas 23 and 24. $\qquad\square$

Lemma 28. *Define the class of functions*

$$\mathcal{A}_{m,n} = \Big\{ (X,Y) \mapsto (\alpha' Z(X)) \cdot \Big(\mathbf{1}\{Y \leq Q(u,X)\} - \mathbf{1}\{Y \leq Z(X)'\beta(u)\} \Big) \colon u \in \mathcal{U}, \ \alpha \in S^{m-1} \Big\},$$

mapping $\mathcal{X} \times \mathbb{R}$ *into* $\mathbb{R}$, *and let* $A_{m,n}(X,Y) = \|Z(X)\|$ *be its envelope. The uniform entropy numbers of* $\mathcal{A}_{m,n}$ *satisfy*

$$\sup_Q \log N(\epsilon \|A_{m,n}\|_{Q,2}, \mathcal{A}_{m,n}, L_2(Q)) \lesssim m \log(1/\epsilon), \quad \textit{uniformly over } 0 < \epsilon \leq 1.$$

Proof. Consider the function class

$$\mathcal{V}_3 = \Big\{ (X,Y) \mapsto \mathbf{1}\{Y \leq Q(u,X)\} \colon u \in \mathcal{U} \Big\}.$$

Note that since $u \mapsto Q(u,x)$ is increases for all $x \in \mathcal{X}$, it follows that

$$\Big\{ (X,Y) \colon Y \leq Q(u_1, X) \Big\} \subset \Big\{ (X,Y) \colon Y \leq Q(u_2, X) \Big\}$$

for all $u_1, u_2 \in \mathcal{U}$ with $u_1 < u_2$. Therefore, VC index of $\mathcal{V}_3$ is $O(1)$. Hence, the asserted claim follows from the same argument as that used in the proof of Lemma 25. $\qquad\square$

H.4. **Eigenvalues of Gram Matrices.** Consider the maximum between the maximum eigenvalues associated with the empirical Gram matrix and the population Gram matrix:

$$\phi_n = \max_{\alpha \in S^{m-1}} \mathbb{E}_n \left[(\alpha' Z_i)^2 \right] \vee E \left[(\alpha' Z_i)^2 \right]. \tag{H.125}$$

The factor ϕ_n will be used to bound the quantities $\epsilon_0(m,n)$ and $\epsilon_1(m,n)$ in the analysis for the rates of convergence. To bound ϕ_n, we use the following result due to Guédon and Rudelson [39] specialized to our framework:

Lemma 29 (Guédon and Rudelson [39])**.** *Let $(Z_i)_{i=1}^n$ be i.i.d. random vectors in $\mathbb{R}^m$. Suppose that*

$$\delta^2 := \frac{\log n}{n} \cdot \frac{E[\max_{1 \leq i \leq n} \|Z_i\|^2]}{\max_{\alpha \in S^{m-1}} E[(Z_1'\alpha)^2]} < 1.$$

Then we have

$$E\left[\max_{\alpha \in S^{m-1}} \left|\frac{1}{n}\sum_{i=1}^n (Z_i'\alpha)^2 - E[(Z_i'\alpha)^2]\right|\right] \leq 2\delta \cdot \max_{\alpha \in S^{m-1}} E[(Z_1'\alpha)^2].$$

Corollary 4. *Denote $\lambda_{max} = \max_{\alpha \in S^{m-1}} E[(Z_1'\alpha)^2]$. Suppose that Condition S holds. In addition, suppose that $\zeta_m^2 \log n = o(n)$. Then for all sufficiently large n, we have for ϕ_n defined in (H.125) that*

$$E[\phi_n] \leq \left(1 + 2\sqrt{\frac{\zeta_m^2 \log n}{n\lambda_{max}}}\right)\lambda_{max} \quad and \quad P(\phi_n > 2\lambda_{max}) \leq 2\sqrt{\frac{\zeta_m^2 \log n}{n\lambda_{max}}}.$$

Proof. Let δ be defined as in Lemma 29. Note that $\|Z_i\|^2 \leq \zeta_m^2$ for all $i = 1, \ldots, n$ and $1 \lesssim \lambda_{\max} \lesssim 1$ by Condition S. Hence, $\delta^2 \lesssim \zeta_m^2 \log n/n$, and so $\delta^2 < 1$ for all n large enough under our assumption that $\zeta_m^2 \log n = o(n)$. Therefore, the first result follows by applying Lemma 29 and the triangle inequality.

To show the second result, note that the event $\{\phi_n > 2\lambda_{max}\}$ cannot occur if $\phi_n = \max_{\alpha \in S^{m-1}} E[(Z_1'\alpha)^2] = \lambda_{max}$. Thus,

$$\begin{aligned}
P(\phi_n > 2\lambda_{max}) &= P\left(\max_{\alpha \in S^{m-1}} \mathbb{E}_n[(Z_i'\alpha)^2] > 2\lambda_{max}\right) \\
&\leq P\left(\max_{\alpha \in S^{m-1}} \left|\mathbb{E}_n[(Z_i'\alpha)^2] - E\left[(Z_i'\alpha)^2\right]\right| > \lambda_{max}\right) \\
&\leq E\left[\max_{\alpha \in S^{m-1}} \left|\mathbb{E}_n[(Z_i'\alpha)^2] - E\left[(Z_i'\alpha)^2\right]\right|\right]/\lambda_{max} \leq 2\delta,
\end{aligned}$$

by the triangle inequality, Markov's inequality, and Lemma 29. This completes the proof of the corollary. $\qquad \square$

H.5. **Estimation of Matrices.** Some of the procedures proposed in this paper rely on estimators of the Gram and Jacobian matrices defined in (2.12) and (2.11), respectively. The following result establishes rates of convergence of these estimators.

Lemma 30 (Estimation of Gram and Jacobian Matrices)**.** *Suppose that Condition S holds. Then*

$$\|\widehat{\Sigma} - \Sigma\| \lesssim_P \sqrt{\frac{\zeta_m^2 \log n}{n}} \tag{H.126}$$

for $\widehat{\Sigma}$ defined in (2.12) as long as $\zeta_m^2 \log n = o(n)$. In addition,

$$\sup_{u \in \mathcal{U}} \|\widehat{J}(u) - J(u)\| \lesssim_P \sqrt{\frac{m \zeta_m^2 \log n}{nh}} + m^{-\kappa} + h \qquad \text{(H.127)}$$

for $\widehat{J}(u)$ defined in (2.11) as long as $h = o(1)$, $m \zeta_m^2 \log^2 n = o(nh)$, and $m^{-\kappa} \log n = o(1)$.

Proof. Note that (H.126) follows from Lemma 29. Hence, it suffices to show (H.127). To do so, observe that conditions of the lemma imply Conditions of Theorem 1, and so for any $\varphi \in (0,1)$, there exists $B > 0$ such that with probability at least $1 - \varphi$, we have $(u, \widehat{\beta}(u)) \in R_{n,m}$ for all $u \in \mathcal{U}$ for $R_{n,m}$ defined in (D.58) with $r_n = B\sqrt{m/n} = o(1)$. For this $R_{n,m}$, define

$$\epsilon_4(m,n) \;=\; \frac{1}{2\sqrt{n}h} \sup_{\substack{\alpha \in S^{m-1}, \\ (u,\beta) \in R_{m,n}}} \left| \mathbb{G}_n \Big(1\{|Y_i - Z_i'\beta| \le h\}(\alpha' Z_i)^2 \Big) \right|, \qquad \text{(H.128)}$$

$$\epsilon_5(m,n) \;=\; \sup_{\substack{\alpha \in S^{m-1}, \\ (u,\beta) \in R_{n,m}}} \left| \frac{1}{2h} E\Big[1\{|Y_i - Z_i'\beta| \le h\}(\alpha' Z_i)^2 \Big] - \alpha' J(u)\alpha \right|. \qquad \text{(H.129)}$$

Then on the event that $(u, \widehat{\beta}(u)) \in R_{n,m}$ for all $u \in \mathcal{U}$,

$$\sup_{u \in \mathcal{U}} \|\widehat{J}(u) - J(u)\| \le \epsilon_4(m,n) + \epsilon_5(m,n).$$

But by Lemma 35,

$$\epsilon_4(m,n) + \epsilon_5(m,n) \lesssim_P \sqrt{\frac{m \zeta_m^2 \log n}{nh}} + \frac{m \zeta_m^2 \log n}{nh} + m^{-\kappa} + \zeta_m \sqrt{m/n} + h$$

$$\lesssim \sqrt{\frac{m \zeta_m^2 \log n}{nh}} + m^{-\kappa} + h.$$

Hence, (H.129) follows. This completes the proof of the lemma. $\qquad\qquad \square$

H.6. **Bounds on Various Empirical Errors.** Here we provide probabilistic bounds for the error terms $\epsilon_0(m,n)$–$\epsilon_5(m,n)$ along with some auxiliary bounds. Our results rely on empirical processes techniques; in particular, they rely on the maximal inequalities derived in Section H.2.

Lemma 31 (Bounds on Approximation Error for Uniform Linear Approximation)**.** *Under Condition S,*

$$\tilde{r}_u := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \Big(1\{Y_i \le Q(u, X_i)\} - 1\{Y_i \le Z_i'\beta(u)\} \Big), \quad u \in \mathcal{U}, \qquad \text{(H.130)}$$

satisfies

$$\sup_{u \in \mathcal{U}} \|\tilde{r}_u\| \lesssim_P \sqrt{m^{1-\kappa}\log n} + \frac{\zeta_m m \log n}{\sqrt{n}}. \tag{H.131}$$

In addition, $E[\tilde{r}_u] = 0$ for all $u \in \mathcal{U}$.

Proof. We first prove the second asserted claim. Fix $u \in \mathcal{U}$. Note that

$$E\Big[Z \cdot 1\{Y \le Q(u,X)\}\Big] = E\Big[Z \cdot 1\{Q(U,X) \le Q(u,X)\}\Big] = E\Big[Z \cdot 1\{U \le u\}\Big] = uE[Z].$$

Also, since $\beta(u)$ is the solution of the optimization problem (2.2), the first-order conditions imply that

$$E\Big[Z \cdot 1\{Y \le Z'\beta(u)\}\Big] = uE[Z].$$

Thus,

$$E\Big[Z \cdot (1\{Y \le Q(u,X)\} - 1\{Y \le Z'\beta(u)\})\Big] = 0,$$

and so $E[\tilde{r}_u] = 0$, which is the second asserted claim.

To prove the first asserted claim, recall the class of functions $\mathcal{A}_{m,n}$ defined in Lemma 28. Since $E[\widetilde{r}_u] = 0$ for all $u \in \mathcal{U}$, it follows that

$$\sup_{u \in \mathcal{U}} \|\tilde{r}_u\| = \sup_{f \in \mathcal{A}_{m,n}} |\mathbb{G}_n(f)|$$

$$\lesssim_P J(m)\Big(\sup_{f \in \mathcal{A}_{m,n}} E\left[f^2\right] + n^{-1}J(m)^2\bar{F}_m^2\log n\Big)^{1/2}\log^{1/2}n$$

by the third maximal inequality in Lemma 22, where $\bar{F}_m = \zeta_m$ by Condition S and $J(m) \lesssim \sqrt{m}$ by Lemma 28. Hence, (H.131) holds provided we can show that

$$\sup_{f \in \mathcal{A}_{m,n}} E\left[f^2\right] \lesssim m^{-\kappa}. \tag{H.132}$$

In turn, to show (H.132), note that for any $f \in \mathcal{A}_{m,n}$, there exist $\alpha \in S^{m-1}$ and $u \in \mathcal{U}$ such that

$$|f(X,Y)| = |\alpha'Z| \cdot \Big|1\{Y \le Q(u,X)\} - 1\{Y \le Z'\beta(u)\}\Big|$$

$$\le |\alpha'Z| \cdot 1\Big\{|Y - Z'\beta(u)| \le |R(u,X)|\Big\},$$

where the second line holds because $R(u,X) = Q(u,X) - Z'\beta(u)$. Thus,

$$E[|f(X,Y)|^2] \lesssim m^{-\kappa}E[|\alpha'Z|^2] \lesssim m^{-\kappa},$$

since $Z = Z(X)$ and we have by Condition S that (i) the conditional pdf of Y given X is bounded from above, (ii) $|R(u,X)| \lesssim m^{-\kappa}$, and (iii) all eigenvalues of the matrix $E[ZZ']$ are bounded from above. This gives (H.132) and completes the proof of the lemma. $\qquad\square$

Lemma 32 (Bounds on $\epsilon_0(m,n)$ and $\sup_{u\in\mathcal{U}}\|U(u)\|$). *Under Condition S, we have*

$$\epsilon_0(m,n) \lesssim_P \sqrt{m}\left(1 + \sqrt{m^{-\kappa}\log n} + \frac{\sqrt{m}\zeta_m\log n}{\sqrt{n}}\right) \quad and \quad \sup_{u\in\mathcal{U}}\|\mathbb{U}(u)\| \lesssim_P \sqrt{m},$$

for $\epsilon_0(m,n)$ and $\mathbb{U}(u)$ defined in (D.59) and (3.16), respectively.

Proof. First, we establish the bound on $\sup_{u\in\mathcal{U}}\|\mathbb{U}(u)\|$. We have

$$\sup_{u\in\mathcal{U}}\|\mathbb{U}(u)\|^2 = \sup_{u\in\mathcal{U}}\frac{1}{n}\sum_{j=1}^{m}\left(\sum_{i=1}^{n}Z_{ij}(u - 1\{U_i \le u\})\right)^2.$$

Therefore, by the triangle inequality,

$$E\left[\sup_{u\in\mathcal{U}}\|\mathbb{U}(u)\|^2\right] \le \frac{1}{n}\sum_{j=1}^{m}E\left[\sup_{u\in\mathcal{U}}\left(\sum_{i=1}^{n}Z_{ij}(u - 1\{U_i \le u\})\right)^2\right].$$

In addition, since $E[Z_{ij}(u - 1\{U_i \le u\})] = 0$, by Theorem 2.14.1 in van der Vaart and Wellner [80],

$$E\left[\sup_{u\in\mathcal{U}}\left(\sum_{i=1}^{n}Z_{ij}(u - 1\{U_i \le u\})\right)^2\right] \lesssim nE[|Z_{1j}|^2]$$

uniformly over $j = 1,\ldots,m$ since the function class

$$\mathcal{G}_n = \left\{(Z,U) \mapsto Z\cdot(u - 1\{U \le u\}): u \in \mathcal{U}\right\},$$

mapping $\mathbb{R}\times[0,1]$ into $\mathbb{R}$, has an envelope $F(Z,U) = |Z|$, and its uniform entropy numbers satisfy

$$\sup_{Q} N(\epsilon\|G_n\|_{Q,2}, \mathcal{G}_n, L_2(Q)) \lesssim (1/\epsilon)^{O(1)}, \quad \text{uniformly over } 0 < \epsilon \le 1,$$

by Lemmas 23 and 24 (here we used the fact that the function class $\{U \mapsto 1\{U \le u\}: u \in \mathcal{U}\}$, mapping $[0,1]$ into $\mathbb{R}$, has VC index $O(1)$). In addition, $E[|Z_{1j}|^2] \le \|E[Z_1 Z_1']\| \lesssim 1$ uniformly over $j = 1,\ldots,m$ by Condition S. Combining these inequalities shows that $E[\sup_{u\in\mathcal{U}}\|\mathbb{U}(u)\|^2] \lesssim m$, and so $\sup_{u\in\mathcal{U}}\|\mathbb{U}(u)\| \lesssim_P \sqrt{m}$ by Markov's inequality, which is the asserted claim for $\sup_{u\in\mathcal{U}}\|\mathbb{U}(u)\|$.

Second, to bound $\epsilon_0(m,n)$, observe that $\epsilon_0(m,n)$ is equal to

$$\sup_{u\in\mathcal{U}}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Z_i\cdot(1\{Y_i \le Z_i\beta(u)\} - u) - E\left[Z_i\cdot(1\{Y_i \le Z_i\beta(u)\} - u)\right]\right)\right\|$$

$$= \sup_{u\in\mathcal{U}}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(Z_i\cdot(1\{Y_i \le Q(u,X_i)\} - u) - E\left[Z_i\cdot(1\{Y_i \le Q(u,X_i)\} - u)\right]\right) - \tilde{r}_u\right\|$$

$$= \sup_{u\in\mathcal{U}}\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Z_i\cdot(1\{U_i \le u\} - u) - \tilde{r}_u\right\| = \sup_{u\in\mathcal{U}}\left\|\mathbb{U}(u) + \tilde{r}_u\right\| \le \sup_{u\in\mathcal{U}}\|\mathbb{U}(u)\| + \sup_{u\in\mathcal{U}}\|\tilde{r}_u\|,$$

where the second line follows from the definition of $\tilde{r}_u$ in (H.130) and the fact that $E[\tilde{r}_u] = 0$ for all $u \in \mathcal{U}$, which is proven in Lemma 31, and the third line follows from the definition of $\mathbb{U}(u)$ and the fact that $Y_i = Q(U_i, X_i)$ with the function $u \mapsto Q(u, X_i)$ being increasing for all $i = 1, \ldots, n$. Hence,

$$\epsilon_0(m, n) \leq \sup_{u \in \mathcal{U}} \|\mathbb{U}(u)\| + \sup_{u \in \mathcal{U}} \|\tilde{r}_u\|$$

$$\lesssim_P \sqrt{m} + \sqrt{m^{1-\kappa} \log n} + \frac{\zeta_m m \log n}{\sqrt{n}}$$

by the bound on $\sup_{u \in \mathcal{U}} \|\mathbb{U}(u)\|$ derived above and the bound on $\sup_{u \in \mathcal{U}} \|\tilde{r}_u\|$ derived in Lemma 31. This completes the proof of the lemma. $\qquad\square$

Lemma 33 (Bounds on $\epsilon_1(m, n)$ and $\epsilon_2(m, n)$). *Under Condition S, we have*

$$\epsilon_1(m, n) \lesssim_P \left(m r_n \zeta_m \log n \right)^{1/2} + \frac{m \zeta_m}{\sqrt{n}} \log n, \tag{H.133}$$

$$\epsilon_2(m, n) \lesssim_P \sqrt{n} r_n^2 \zeta_m + \sqrt{n} m^{-\kappa} r_n \tag{H.134}$$

for $\epsilon_1(m, n)$ and $\epsilon_2(m, n)$ defined in (D.59).

Proof. We first prove (H.133). Recall the class of functions $\mathcal{F}_{m,n}$ defined in Lemma 25. It follows that

$$\epsilon_1(m, n) = \sup_{f \in \mathcal{F}_{m,n}} |\mathbb{G}_n(f)|$$

$$\lesssim_P J(m) \left(\sup_{f \in \mathcal{F}_{m,n}} E[f^2] + n^{-1} J(m)^2 \bar{F}_m^2 \log n \right)^{1/2} \log^{1/2} n$$

by the third maximal inequality in Lemma 22, where $\bar{F}_m = \zeta_m$ by Condition S and $J(m) \lesssim \sqrt{m}$ by Lemma 25. Hence, (H.133) holds provided we can show that

$$\sup_{f \in \mathcal{F}_{m,n}} E[f^2] \lesssim r_n \zeta_m. \tag{H.135}$$

In turn, to show (H.135), note that for any $f \in \mathcal{F}_{m,n}$, there exist $\alpha \in S^{m-1}$, $u \in \mathcal{U}$, and $\beta \in \mathbb{R}^m$ such that $\|\beta - \beta(u)\| \leq r_n$ and

$$|f(Z, Y)| = |\alpha' Z| \cdot \left| 1\{Y \leq Z'\beta\} - 1\{Y \leq Z'\beta(u)\} \right|$$

$$\leq |\alpha' Z| \cdot \left| 1\{|Y - Z'\beta(u)| \leq |Z'(\beta - \beta(u))|\} \right|$$

$$\leq |\alpha' Z| \cdot 1\{|Y - Z'\beta(u)| \leq r_n \zeta_m\},$$

where the third line follows from the Cauchy-Schwarz inequality and the observation that $\|Z\| = \|Z(X)\| \leq \zeta_m$, almost surely. Thus,

$$E[|f(Z,Y)|^2] \lesssim r_n \zeta_m E[|\alpha'Z|^2] \lesssim r_n \zeta_m,$$

like in the proof of Lemma 31. This gives (H.135) and completes the proof of (H.133).

Second, we prove (H.134). By Lemma 18, the matrix $\widetilde{J}(u) = E[f_{Y|X}(Z'\beta(u)|X)ZZ']$ satisfies

$$\sqrt{n}\left|\alpha'(J(u) - \widetilde{J}(u))(\beta - \beta(u))\right| \leq \sqrt{n}\|\alpha\| \cdot \|J(u) - \widetilde{J}(u)\| \cdot \|\beta - \beta(u)\| \lesssim \sqrt{n}m^{-\kappa}r_n$$

uniformly over $\alpha \in S^{m-1}$ and $(u, \beta) \in R_{n,m}$. Thus, if we define

$$\epsilon_2(m, n, \alpha, u, \beta) = \sqrt{n}\left|\alpha'\Big(E[\psi_i(\beta, u)] - E[\psi_i(\beta(u), u)]\Big) - \alpha'\widetilde{J}(u)(\beta - \beta(u))\right|,$$

then

$$\epsilon_2(m, n) \lesssim \sup_{\alpha \in S^{m-1}, (u,\beta) \in R_{n,m}} \epsilon_2(m, n, \alpha, u, \beta) + \sqrt{n}m^{-\kappa}r_n.$$

Next, note that by the law of iterated expectations and the mean-value theorem,

$$\alpha'\Big(E[\psi_i(\beta, u)] - E[\psi_i(\beta(u), u)]\Big) = E\Big[f_{Y|X}(Z'\widetilde{\beta}|X) \cdot (\alpha'Z) \cdot (Z'(\beta - \beta(u)))\Big]$$

for some $\widetilde{\beta}$ on the line segment between $\beta(u)$ and β. In addition, it follows from Condition S that

$$\left|f_{Y|X}(Z'\widetilde{\beta}|X) - f_{Y|X}(Z'\beta(u)|X)\right| \lesssim |Z'(\widetilde{\beta} - \beta(u))| \leq |Z'(\beta - \beta(u))| \lesssim \zeta_m r_n$$

uniformly over $X \in \mathcal{X}$, $\alpha \in S^{m-1}$, and $(u, \beta) \in R_{n,m}$. Hence,

$$\epsilon_2(m, n, \alpha, u, \beta) = \sqrt{n}\left|E\Big[(f_{Y|X}(Z'\widetilde{\beta}|X) - f_{Y|X}(Z'\beta(u)|X)) \cdot (\alpha'Z) \cdot (Z'(\beta - \beta(u)))\Big]\right|$$

$$\lesssim \sqrt{n}\zeta_m r_n \cdot E\Big[\alpha'ZZ'(\beta - \beta(u))\Big] \lesssim \sqrt{n}\zeta_m r_n^2$$

uniformly over $\alpha \in S^{m-1}$ and $(u, \beta) \in R_{n,m}$ by Condition S. Combining presented inequalities gives (H.134) and completes the proof of the lemma. $\square$

Lemma 34 (Bound on $\epsilon_3(m, n)$). *Let $\widehat{\beta}(u)$ be a solution to the perturbed QR problem (D.61). If the data are in general position so that (D.60) holds, we have*

$$\epsilon_3(m, n) \leq \min\left(\frac{\zeta_m m}{\sqrt{n}}, \phi_n\sqrt{m}\right)$$

holds with probability 1 for $\epsilon_3(m, n)$ and ϕ_n defined in (D.65) and (H.125), respectively.

Proof. Note that the dual problem associated with the perturbed QR problem (D.61) is

$$\max_{(u-1)\leq a_i \leq u} \mathbb{E}_n[Y_i a_i] : \mathbb{E}_n[Z_i a_i] = -\mathcal{A}_n(u).$$

Letting $\widehat{a}(u)$ denote the solution for the dual problem above, and letting $a_i(\widehat{\beta}(u)) := (u - 1\{Y_i \leq Z_i'\widehat{\beta}(u)\})$, we have by the triangle inequality that

$$\epsilon_3(m,n) \leq \sup_{\alpha \in S^{m-1}, u \in \mathcal{U}} \sqrt{n}\left|\mathbb{E}_n\left[(Z_i'\alpha)(a_i(\widehat{\beta}(u)) - \widehat{a}_i(u))\right]\right|$$

$$+ \sup_{u \in \mathcal{U}} \sqrt{n}\|\mathbb{E}_n\left[Z_i\widehat{a}_i(u)\right] + \mathcal{A}_n(u)\|.$$

By dual feasibility $\mathbb{E}_n\left[Z_i\widehat{a}_i(u)\right] = -\mathcal{A}_n(u)$, and the second term is identically equal to zero.

Also, note that $a_i(\widehat{\beta}(u)) \neq \widehat{a}_i(u)$ only if the ith point is interpolated. Since the data are in general position, with probability one the quantile regression interpolates m points ($Z_i'\widehat{\beta}(u) = Y_i$ for m points for every $u \in \mathcal{U}$).

Therefore, noting that $|a_i(\widehat{\beta}(u)) - \widehat{a}_i(u)| \leq 1$, we have

$$\epsilon_3(m,n) \leq \sup_{\|\alpha\|\leq 1, u \in \mathcal{U}} \sqrt{n}\sqrt{\mathbb{E}_n\left[(Z_i'\alpha)^2\right]}\sqrt{\mathbb{E}_n\left[\{a_i(\widehat{\beta}(u)) - \widehat{a}_i(u)\}^2\right]} \leq \phi_n\sqrt{m}$$

and, with probability 1,

$$\epsilon_3(m,n) \leq \sup_{\|\alpha\|\leq 1, u \in \mathcal{U}} \sqrt{n}\mathbb{E}_n\left[1\{a_i(\widehat{\beta}(u)) \neq \widehat{a}_i(u)\} \max_{1\leq i \leq n} \|Z_i\|\right] \leq \frac{m}{\sqrt{n}}\max_{1\leq i \leq n} \|Z_i\|.$$

This completes the proof of the lemma. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Lemma 35 (Bounds on $\epsilon_4(m,n)$ and $\epsilon_5(m,n)$). *Suppose that Condition S holds. In addition, suppose that $h = h_n = o(1)$ and $r_n = o(1)$. Then*

$$\epsilon_4(m,n) \lesssim_P \sqrt{\frac{\zeta_m^2 m \log n}{nh}} + \frac{m\zeta_m^2}{nh}\log n \quad \text{and} \quad \epsilon_5(m,n) \lesssim m^{-\kappa} + r_n\zeta_m + h$$

for $\epsilon_4(m,n)$ and $\epsilon_5(m,n)$ defined in (H.128) and (H.129), respectively.

Proof. To bound $\epsilon_4(m,n)$, consider the function class

$$\mathcal{H}_{m,n} = \left\{(Z,Y) \mapsto (\alpha'Z)^2 \cdot 1\{|Y - Z'\beta| \leq h\} : \alpha \in S^{m-1}, (u,\beta) \in R_{n,m}\right\}$$

for $R_{n,m}$ defined in (D.58). Then

$$\epsilon_4(m,n) = \frac{1}{2\sqrt{nh}} \sup_{f \in \mathcal{H}_{n,m}} |\mathbb{G}_n(f)|$$

$$\lesssim_P \frac{1}{\sqrt{nh}}J(m)\left(\sup_{f \in \mathcal{H}_{m,n}} E[f^2] + n^{-1}J(m)^2\bar{F}^2 \log n\right)^{1/2} \log^{1/2} n$$

by the third maximal inequality in Lemma 22, where $\bar{F}_m = \zeta_m^2$ by Condition S and $J(m) \lesssim \sqrt{m}$ by Lemma 26. In addition,

$$\sup_{f \in \mathcal{H}_{m,n}} E[f^2] = \sup_{\alpha \in S^{m-1}} \sup_{(u,\beta) \in R_{n,m}} E\Big[(\alpha'Z)^4 \cdot 1^2\{|Y - Z'\beta| \le h\}\Big]$$

$$= \sup_{\alpha \in S^{m-1}} \sup_{(u,\beta) \in R_{n,m}} E\Big[(\alpha'Z)^4 \cdot 1\{|Y - Z'\beta| \le h\}\Big]$$

$$\lesssim h \sup_{\alpha \in S^{m-1}} E\Big[(\alpha'Z)^4\Big] \lesssim h\zeta_m^2 \sup_{\alpha \in S^{m-1}} E\Big[(\alpha'Z)^2\Big] \lesssim h\zeta_m^2$$

by Condition S. The bound for $\epsilon_4(m,n)$ follows from combining the inequalities above.

To show the bound on $\epsilon_5(m,n)$, let $f'_{Y|X}(y|x)$ denote the derivative of the function $y \mapsto f_{Y|X}(y|x)$. By Condition S, we have for some finite constant $\bar{f}'$ that $|f'_{Y|X}(y|x)| \le \bar{f}'$ for all $y \in \mathcal{Y}_x$ and $x \in \mathcal{X}$. Therefore,

$$E\Big[1\{|Y - Z'\beta| \le h\}(\alpha'Z)^2\Big] = E\Big[(\alpha'Z)^2 \int_{-h}^{h} f_{Y|X}(Z'\beta + t|X)dt\Big]$$

$$= E\Big[(\alpha'Z)^2 \int_{-h}^{h} \Big(f_{Y|X}(Z'\beta|X) + t f'_{Y|X}(Z'\beta + \tilde{t}|X)\Big)dt\Big]$$

$$= 2hE\Big[f_{Y|X}(Z'\beta|X)(\alpha'Z)^2\Big] + O\Big(h^2 \bar{f}' E[(Z'\alpha)^2]\Big)$$

$$= 2hE\Big[f_{Y|X}(Z'\beta|X)(\alpha'Z)^2\Big] + O(h^2)$$

for some $\tilde{t}$ between 0 and t by the mean-value theorem. Moreover, for any $(u,\beta) \in R_{m,n}$,

$$E\Big[f_{Y|X}(Z'\beta|X)(\alpha'Z)^2\Big] = E\Big[f_{Y|X}(Z'\beta(u)|X)(\alpha'Z)^2\Big]$$

$$+ E\Big[(f_{Y|X}(Z'\beta|X) - f_{Y|X}(Z'\beta(u)|X))(\alpha'Z)^2\Big]$$

$$= E\Big[f_{Y|X}(Z'\beta(u)|X)(\alpha'Z)^2\Big] + O\Big(E[\bar{f}'Z'(\beta - \beta(u))(\alpha'Z)^2]\Big)$$

$$= \alpha'\widetilde{J}(u)\alpha + O\Big(r_n\zeta_m E[(\alpha'Z)^2]\Big)$$

$$= \alpha'J(u)\alpha + O(m^{-\kappa}) + O(r_n\zeta_m),$$

by Condition S, where $\widetilde{J}(u) = E[f_{Y|X}(Z'\beta(u)|X)ZZ']$ and where the last line follows from Lemma 19. Thus, $\epsilon_5(m,n) \lesssim m^{-\kappa} + r_n\zeta_m + h$. This completes the proof of the lemma. $\square$

Appendix I. A Lemma on Strong Approximation for a Process in the Sup-Norm

In this section, we develop a novel technique to construct a Gaussian coupling for certain empirical processes in the sup-norm. The technique is based on an extension of the

Yurinskii's coupling in the Euclidean norm to arbitrary norm (including the sup-norm). Although the technique can be applied to general empirical processes and may be of independent interest, we apply it to a particular empirical process, which is needed to establish Theorems 12 and 14, for brevity of the paper.

Lemma 36. *(Approximation of Linear Functionals of a Sequence of Empirical Processes of Increasing Dimension by a Sequence of Gaussian Processes) Let $(Z_i)_{i=1}^n$ be a sequence of non-stochastic vectors in $\mathbb{R}^m$ and consider the empirical process $\mathbb{U}_n$ in $[\ell^\infty(\mathcal{U})]^m$, $\mathcal{U} \subseteq (0,1)$, defined by*

$$\mathbb{U}_n(u) = \mathbb{G}_n \left(v_i Z_i \psi_i(u) \right), \quad \psi_i(u) = u - 1\{U_i \le u\}, \quad u \in \mathcal{U},$$

where $(U_i, v_i)_{i=1}^n$ is an i.i.d. sequence of pairs of independent random variables where $U_i \sim Uniform(0,1)$, $E[v_i^2] = 1$, $E[|v_i|^4] \lesssim 1$, and $\max_{1 \le i \le n} |v_i| \lesssim_P \log n$. Suppose that the vectors Z_i are such that

$$\sup_{\alpha \in S^{m-1}} \mathbb{E}_n \left[(\alpha' Z_i)^2 \right] \lesssim 1 \ and \ \max_{1 \le i \le n} \|Z_i\| \lesssim \zeta_m,$$

where ζ_m satisfies $1/\zeta_m \lesssim 1$. Also, let $\mathcal{W}$ be a set in $\mathbb{R}^d$ and let I be a subset of $\mathcal{U} \times \mathcal{W}$ whose dimension d_I is independent of n and whose diameter is bounded uniformly over n. Moreover, let $\ell \colon I \to \mathbb{R}^m$ be a function such that

$$\|\ell(u,w)\| \le \xi_\ell \ and \ \|\ell(u,w) - \ell(\tilde{u}, \tilde{w})\| \le L_\ell \|(u,w) - (\tilde{u}, \tilde{w})\|$$

for all (u,w) and $(\tilde{u}, \tilde{w})$ in I, where ξ_ℓ and L_ℓ satisfy $\xi_\ell \lesssim 1$ and $L_\ell \gtrsim 1$, and consider the process $\mathbb{L}_n$ in $\ell^\infty(I)$ defined by

$$\mathbb{L}_n(u,w) = \ell(u,w)' \mathbb{U}_n(u), \quad (u,w) \in I.$$

Finally, suppose that

$$L_\ell^{2d_I} \zeta_m^2 = o(n^{1-\varepsilon}) \tag{I.136}$$

for some constant $\varepsilon > 0$. Then there exists a sequence of zero-mean Gaussian processes $(G_n)_{n \ge 1}$ with a.s. continuous paths such that (i) the covariance functions of G_n coincide with those of $\mathbb{L}_n$, namely,

$$E[G_n(u,w)G_n(\tilde{u},\tilde{w})] = \mathbb{E}_n[\ell(u,w)' Z_i Z_i' \ell(\tilde{u},\tilde{w})](u \wedge \tilde{u} - u\tilde{u}),$$

for all (u,w) and $(\tilde{u}, \tilde{w})$ in I, and (ii) G_n approximates $\mathbb{L}_n$, namely,

$$\sup_{(u,w) \in I} \left| \mathbb{L}_n(u,w) - G_n(u,w) \right| \lesssim_P o(n^{-\varepsilon'}),$$

where ε' is some constant.

Proof. The proof is based on the use of maximal inequalities and our extension of Yurinskii's coupling to the sup norm; see Lemma 39 below. Although the proof is closely related to that of Lemma 14, the details of calculations are rather different.

For each $j \geq 1$, we define a sequence of projections $\pi_j \colon I \to I$ associated with a covering of the set I by balls of radius 2^{-j}, where each element of the set I is mapped to the center of the ball containing this element (if an element is contained by several balls, choose one according to some predetermined rule). We assume that all balls used to construct the projection π_j are also used to construct projections $\pi_{j'}$ for all $j' > j$. Since the set I is such that its dimension d_I is independent of n and its diameter is bounded uniformly over n, it is standard to show that the projection π_j can be constructed using k_j balls with $k_j \lesssim 2^{jd_I}$.

In what follows, given a process G in $\ell^\infty(I)$ and its projection $G \circ \pi_j$, we shall identify the process $G \circ \pi_j$ with a random vector $G \circ \pi_j$ in $\mathbb{R}^{k_j}$, when convenient. Analogously, given a random vector W in $\mathbb{R}^{k_j}$, we shall identify it with a process W in $\ell^\infty(I)$ that is piece-wise constant.

The following relations will be proven below for some $\varepsilon' > 0$ and $j = j_n \to \infty$:

(1) (Finite-Dimensional Approximation)

$$r_1 = \sup_{(u,w) \in I} |\mathbb{L}_n(u,w) - \mathbb{L}_n \circ \pi_j(u,w)| \lesssim_P o(n^{-\varepsilon'});$$

(2) (Coupling with a Normal Vector) there exists $\mathcal{N}_{nj} =_d N(0, \mathrm{var}[\mathbb{L}_n \circ \pi_j])$ such that

$$r_2 = \|\mathcal{N}_{nj} - \mathbb{L}_n \circ \pi_j\|_\infty \lesssim_P o(n^{-\varepsilon'});$$

(3) (Embedding a Normal Vector into a Gaussian Process) there exists a Gaussian process G_n with properties stated in the lemma such that $\mathcal{N}_{nj} = G_n \circ \pi_j$ a.s.;

(4) (Infinite-Dimensional Approximation)

$$r_3 = \sup_{(u,w) \in I} |G_n(u,w) - G_n \circ \pi_j(u,w)| \lesssim_P o(n^{-\varepsilon'}).$$

The result then follows from the triangle inequality:

$$\sup_{(u,w) \in I} |\mathbb{L}_n(u,w) - G_n(u,w)| \leq r_1 + r_2 + r_3.$$

We now prove relations (1)-(4). Relation (1) follows from

$$
\begin{aligned}
r_1 &= \sup_{(u,w)\in I} |\mathbb{L}_n(u,w) - \mathbb{L}_n \circ \pi_j(u,w)| \\
&\le \sup_{\|(u,w)-(\tilde{u},\tilde{w})\|\le 2^{-j}} |\mathbb{L}_n(u,w) - \mathbb{L}_n(\tilde{u},\tilde{w})| \\
&\lesssim_P \sqrt{(L_\ell^2 2^{-2j} + 2^{-j})\log n} + \sqrt{\frac{\zeta_m^2 \log^4 n}{n}} \le o(n^{-\varepsilon'}),
\end{aligned}
$$

where the first inequality in the third line holds by Lemma 37 and the second by substituting $2^j = L_\ell n^{\tilde{\varepsilon}}$ for an appropriate $\tilde{\varepsilon} > 0$ and using (I.136).

Relation (2) follows from the use of Yurinskii's coupling extended to the sup norm; see Lemma 39 with $\|\cdot\|_d = \|\cdot\|_\infty$. In order to apply the coupling, set $\widehat{I}_j = \{\pi_j(u,w) \colon (u,w) \in I\}$ and for all $i = 1, \ldots, n$, let

$$
\xi_i = \left(v_i \cdot (\ell(u,w)'Z_i) \cdot \psi_i(u)/\sqrt{n} \right)_{(u,w)\in\widehat{I}_j},
$$

so that ξ_i is a zero-mean random vector in $\mathbb{R}^{k_j}$, and we have $\mathbb{L}_n \circ \pi_j =_d \sum_{i=1}^n \xi_i$. Then

$$
\begin{aligned}
&\sum_{i=1}^n E\left[\|\xi_i\|^2 \|\xi_i\|_\infty\right] \\
&= \frac{1}{n^{3/2}} \sum_{i=1}^n E\left[|v_i|^3 \left(\sum_{(u,w)\in\widehat{I}_j} |\ell(u,w)'Z_i|^2 |\psi_i(u)|^2 \right) \left(\max_{(u,w)\in\widehat{I}_j} |\ell(u,w)'Z_i| \cdot |\psi_i(u)| \right)\right] \\
&\lesssim \frac{1}{n^{3/2}} \sum_{i=1}^n E\left[\left(\sum_{(u,w)\in\widehat{I}_j} |\ell(u,w)'Z_i|^2 \right) \left(\max_{(u,w)\in\widehat{I}_j} |\ell(u,w)'Z_i| \right)\right] \\
&\le \frac{1}{n^{3/2}} \sum_{i=1}^n \xi_\ell \zeta_m \sum_{(u,w)\in\widehat{I}_j} \ell(u,w)'Z_i Z_i' \ell(u,w) = \frac{k_j \xi_\ell^3 \zeta_m}{\sqrt{n}}.
\end{aligned}
$$

Further, for $i = 1, \ldots, n$, let $g_i \sim N(0, \Sigma_i)$, where $\Sigma_i = \mathrm{var}(\xi_i)$. Note that the elements of the matrix Σ_j are given by

$$
\ell(u,w)'Z_i Z_i' \ell(\tilde{u},\tilde{w})(u \wedge \tilde{u} - u\tilde{u})/n, \quad \text{for } (u,w) \text{ and } (\tilde{u},\tilde{w}) \text{ in } \widehat{I}_j.
$$

Hence,

$$
\left(E[\|g_i\|_\infty^2]\right)^{1/2} \lesssim \frac{\xi_\ell \zeta_m \sqrt{\log k_j}}{\sqrt{n}}.
$$

Also,

$$\sum_{i=1}^{n}\left(E[\|g_i\|^4]\right)^{1/2} = \sum_{i=1}^{n}\left(E\left[\left(\sum_{l=1}^{k_j}g_{il}^2\right)^2\right]\right)^{1/2} \le \sum_{i=1}^{n}\sum_{l=1}^{k_j}\left(E[g_{il}^4]\right)^{1/2}$$

$$\lesssim \sum_{i=1}^{n}\sum_{l=1}^{k_j}E[g_{il}^2] = \sum_{i=1}^{n}E[\|g_i\|^2] = \sum_{i=1}^{n}E[\|\xi_i\|^2]$$

$$\lesssim \frac{1}{n}\sum_{i=1}^{n}\sum_{(u,w)\in\widehat{I}_j}|\ell(u,w)'Z_i|^2 \lesssim k_j\xi_\ell^2.$$

Therefore,

$$\sum_{i=1}^{n}E\left[\|g_i\|^2\|g_i\|_\infty\right] \le \sum_{i=1}^{n}\left(E[\|g_i\|^4]\right)^{1/2}\left(E[\|g_i\|_\infty^2]\right)^{1/2} \lesssim \frac{k_j\xi_\ell^3\zeta_m\sqrt{\log k_j}}{\sqrt{n}}.$$

So, applying Lemma 39 with

$$\beta = \sum_{i=1}^{n}E\left[\|\xi_i\|^2\|\xi_i\|_\infty\right] + \sum_{i=1}^{n}E\left[\|g_i\|^2\|g_i\|_\infty\right] \lesssim \frac{k_j\xi_\ell^3\zeta_m\sqrt{\log k_j}}{\sqrt{n}}$$

shows that for any $\delta > 0$, there exists a vector $\mathcal{N}_{nj} =_d N(0, \mathrm{var}[\mathbb{L}_n \circ \pi_j])$ such that

$$P\left(\|\mathbb{L}_n \circ \pi_j - \mathcal{N}_{nj}\|_\infty > 3\delta\right) \le \min_{t\ge 0}\left(2P(\|Z\|_\infty > t) + \frac{\beta}{\delta^3}t^2\right)$$

$$= o(1) + O\left(\frac{k_j\xi_\ell^3\zeta_m(\log k_j)^{3/2}}{\delta^3\sqrt{n}}\right), \qquad (\text{I.137})$$

where Z is a standard Gaussian k_j-dimensional random vector and where in the second line we set $t = C(\log k_j)^{1/2}$ for a sufficiently large constant C. Since $\xi_\ell \lesssim 1$ and $k_j \lesssim 2^{jd_I}$, substituting $2^j = L_\ell n^{\tilde{\varepsilon}}$ and using (I.136) shows that there exists $\delta = o(n^{-\varepsilon'})$ such that the expression in (I.137) is $o(1)$ leading to relation (2).

Relation (3) follows from the a.s. embedding of a finite-dimensional random normal vector into a path of a continuous Gaussian process, which is possible by Lemma 17 applied with $m = 1$ and $\mathcal{U}$ replaced by I.

Finally, relation (4) follows from

$$r_3 = \sup_{(u,w)\in I}|G_n(u,w) - G_n \circ \pi_j(u,w)|$$

$$\le \sup_{\|(u,w)-(\tilde{u},\tilde{w})\|\le 2^{-j}}|G_n(u,w) - G_n(\tilde{u},\tilde{w})|$$

$$\lesssim_P \sqrt{L_\ell 2^{-j}\log(2^j)} \le o(n^{-\varepsilon'}),$$

where the first inequality in the third line holds by Lemma 38 and the second by substituting $2^j = L_\ell n^{\tilde{\varepsilon}}$ and using (I.136). This completes the proof of the lemma. $\qquad\square$

Lemma 37 (Finite-Dimensional Approximation)**.** *Consider the setting of Lemma 36. Then for any $\gamma > 0$, the process $\mathbb{L}_n$ satisfies*

$$\sup \left| \mathbb{L}_n(u, w) - \mathbb{L}_n(\tilde{u}, \tilde{w}) \right| \lesssim_P \sqrt{(L_\ell^2 \gamma^2 + \gamma) \log n} + \sqrt{\frac{\zeta_m^2 \log^4 n}{n}},$$

where the supremum is over all (u, w) and $(\tilde{u}, \tilde{w})$ in I such that $\|(u, w) - (\tilde{u}, \tilde{w})\| \leq \gamma$.

Proof. Consider the function class

$$\mathcal{L}_{m,n} = \left\{ (Z, U, v) \mapsto f_{u,w}(Z, U, v) = \ell(u, w)'Z : (u, w) \in I \right\},$$

mapping $B_m(0, \zeta_m) \times [0, 1] \times \mathbb{R}$. Note that $L_{m,n}(Z, U, v) = (\xi_\ell \vee 1)\zeta_m$ is its envelope. Also, since (i) for all (u, w) and $(\tilde{u}, \tilde{w})$ in I we have

$$\left| \ell(u, w)'Z - \ell(\tilde{u}, \tilde{w})'Z \right| \leq \|\ell(u, w) - \ell(\tilde{u}, \tilde{w})\| \cdot \|Z\| \leq L_\ell \|(u, w) - (\tilde{u}, \tilde{w})\| \zeta_m$$

and (ii) the set I is such that its dimension is independent of n and its diameter is bounded uniformly over n, it follows that the uniform entropy numbers of $\mathcal{L}_{m,n}$ satisfy

$$\sup_Q \log N(\epsilon \|L_{m,n}\|_{Q,2}, \mathcal{L}_{m,n}, L_2(Q)) \lesssim \log(L_\ell/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1. \qquad \text{(I.138)}$$

Next, consider the function class

$$\widetilde{\mathcal{L}}_{m,n} = \left\{ (Z, U, v) \mapsto v \cdot (\ell(u, w)'Z) \cdot (u - 1\{U \leq u\}) : (u, w) \in I \right\}.$$

The function $\widetilde{L}_{m,n}(Z, U, v) = |v| \cdot L_{m,n}(Z, U, v)$ is an envelope of $\widetilde{\mathcal{L}}_{m,n}$. By (I.138) and Lemma 24, the uniform entropy numbers of $\widetilde{\mathcal{L}}_{m,n}$ satisfy

$$\sup_Q \log N(\epsilon \|\widetilde{L}_{m,n}\|_{Q,2}, \widetilde{\mathcal{L}}_{m,n}, L_2(Q)) \lesssim \log(L_\ell/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1. \qquad \text{(I.139)}$$

Further, consider the function class

$$\overline{\mathcal{L}}_{m,n} = \Big\{ (Z, U, v) \mapsto v \cdot (\ell(u, w)'Z) \cdot (u - 1\{U \leq u\})$$
$$- v \cdot (\ell(\tilde{u}, \tilde{w})'Z) \cdot (\tilde{u} - 1\{U \leq \tilde{u}\}) : (u, w) \in I, \ (\tilde{u}, \tilde{w}) \in I, \ \|(u, w) - (\tilde{u}, \tilde{w})\| \leq \gamma \Big\}.$$

The function $2\widetilde{L}_{m,n}$ is its envelope. By (I.139) and Lemma 24, the uniform entropy numbers of $\overline{\mathcal{L}}_{m,n}$ satisfy

$$\sup_Q \log N(\epsilon \|2\widetilde{L}_{m,n}\|_{Q,2}, \overline{\mathcal{L}}_{m,n}, L_2(Q)) \lesssim \log(L_\ell/\epsilon), \quad \text{uniformly over } 0 < \epsilon \leq 1. \qquad \text{(I.140)}$$

With this notation, we have

$$\sup_{\|(u,w)-(\tilde{u},\tilde{w})\|\leq\gamma}\left|\mathbb{L}_n(u,w)-\mathbb{L}_n(\tilde{u},\tilde{w})\right| = \sup_{f\in\overline{\mathcal{L}}_{m,n}}|\mathbb{G}_nf|,$$

and so to prove the asserted claim, we can apply the second part of Lemma 22 using the sequence of independent observations $(Z_i, U_i, v_i)_{i=1}^n$.

By (I.140), (H.119) is satisfied with $\omega = L_\ell$, $J(m) = O(1)$, and $F_m = 2\widetilde{\mathcal{L}}_{m,n}$. Also, note that

$$\max_{1\leq i\leq n} F_m(Z_i, U_i, v_i) \lesssim_P M = \xi_\ell\zeta_m\log n$$

since $\max_{1\leq i\leq n} v_i \lesssim_P \log n$. Moreover, $\log\omega \lesssim \log n$ since $\log L_\ell \lesssim \log n$. Further,

$$\sup_{f\in\overline{\mathcal{L}}_{m,n}}\frac{1}{n}\sum_{i=1}^n E[f(Z_i, U_i, v_i)^2]$$

$$= \sup_{\|(u,w)-(\tilde{u},\tilde{w})\|\leq\gamma}\frac{1}{n}\sum_{i=1}^n E\left[v_i^2\left((\ell(u,w)'Z_i)\cdot\psi_i(u) - (\ell(\tilde{u},\tilde{w})'Z_i)\cdot\psi_i(\tilde{u})\right)^2\right]$$

$$= \sup_{\|(u,w)-(\tilde{u},\tilde{w})\|\leq\gamma}\frac{1}{n}\sum_{i=1}^n E\left[\left((\ell(u,w)'Z_i)\cdot\psi_i(u) - (\ell(\tilde{u},\tilde{w})'Z_i)\cdot\psi_i(\tilde{u})\right)^2\right]$$

$$\lesssim \sup_{\|(u,w)-(\tilde{u},\tilde{w})\|\leq\gamma}\frac{1}{n}\sum_{i=1}^n \left((\ell(u,w) - \ell(\tilde{u},\tilde{w}))'Z_i\right)^2$$

$$+ \sup_{\|(u,w)-(\tilde{u},\tilde{w})\|\leq\gamma}\frac{1}{n}\sum_{i=1}^n E\left[(\ell(\tilde{u},\tilde{w})'Z_i)^2\cdot(\psi_i(u) - \psi_i(\tilde{u}))^2\right]$$

$$\lesssim L_\ell^2\gamma^2\varphi + \gamma(1-\gamma)\xi_\ell^2\varphi \lesssim L_\ell^2\gamma^2 + \xi_\ell^2\gamma,$$

where $\varphi = \sup_{\alpha\in S^{m-1}}\mathbb{E}_n[(\alpha'Z_i)^2] \lesssim 1$ and where we used arguments similar to those used in the proof of Lemma 15. Moreover,

$$\sup_{f\in\overline{\mathcal{L}}_{m,n}}\frac{1}{n}\sum_{i=1}^n E[f(Z_i, U_i, v_i)^4]$$

$$= \sup_{\|(u,w)-(\tilde{u},\tilde{w})\|\leq\gamma}\frac{1}{n}\sum_{i=1}^n E\left[v_i^4\left((\ell(u,w)'Z_i)\cdot\psi_i(u) - (\ell(\tilde{u},\tilde{w})'Z_i)\cdot\psi_i(\tilde{u})\right)^4\right]$$

$$= \sup_{\|(u,w)-(\tilde{u},\tilde{w})\|\leq\gamma}\frac{1}{n}\sum_{i=1}^n E\left[\left((\ell(u,w)'Z_i)\cdot\psi_i(u) - (\ell(\tilde{u},\tilde{w})'Z_i)\cdot\psi_i(\tilde{u})\right)^4\right]$$

$$\lesssim \sup_{\|(u,w)-(\tilde{u},\tilde{w})\|\leq\gamma}\frac{1}{n}\sum_{i=1}^n \xi_\ell^2\zeta_m^2 E\left[\left((\ell(u,w)'Z_i)\cdot\psi_i(u) - (\ell(\tilde{u},\tilde{w})'Z_i)\cdot\psi_i(\tilde{u})\right)^2\right]$$

$$\lesssim \xi_\ell^2\zeta_m^2(L_\ell^2\gamma^2 + \xi_\ell^2\gamma),$$

where we used the bound above and the inequalities $\|\ell(u, w)\| \lesssim \xi_\ell$, $\max_{1 \leq i \leq n} E[v_i^4] \lesssim 1$ and $\max_{1 \leq i \leq n} \|Z_i\| \lesssim \zeta_m$. Substituting these bounds into the second part of Lemma 22 gives

$$\sup_{f \in \overline{\mathcal{L}}_{m,n}} |\mathbb{G}_n f| \lesssim_P \sqrt{(L_\ell^2 \gamma^2 + \xi_\ell^2 \gamma) \log n} + \sqrt{\frac{\xi_\ell^2 \zeta_m^2 \log^4 n}{n}}$$

$$\lesssim \sqrt{(L_\ell^2 \gamma^2 + \gamma) \log n} + \sqrt{\frac{\zeta_m^2 \log^4 n}{n}}.$$

This completes the proof of the lemma. $\qquad\square$

Lemma 38 (Infinite-Dimensional Approximation)**.** *Consider the setting of Lemma 36, and let $G_n \colon I \to \mathbb{R}$ be a zero-mean Gaussian process whose covariance structure is given by*

$$E[G_n(u, w) G_n(\tilde{u}, \tilde{w})] = \mathbb{E}_n[\ell(u, w)' Z_i Z_i' \ell(\tilde{u}, \tilde{w})](u \wedge \tilde{u} - u\tilde{u}),$$

for all (u, w) and $(\tilde{u}, \tilde{w})$ in I. Then for any $\gamma \in (0, 1/2)$,

$$\sup_{\|(u,w)-(\tilde{u},\tilde{w})\| \leq \gamma} |G_n(u, w) - G_n(\tilde{u}, \tilde{w})| \lesssim_P \sqrt{L_\ell \gamma \log(1/\gamma)}.$$

Proof. To prove the asserted claim, we apply the maximal inequality (G.112) from Lemma 16 to the zero-mean Gaussian process $X_n \colon I \times I \to \mathbb{R}$ defined by

$$X_{n,t} = G_n(u, w) - G_n(\tilde{u}, \tilde{w}), \ \ t = (u, w, \tilde{u}, \tilde{w}), \ \ \|(u, w) - (\tilde{u}, \tilde{w})\| \leq \gamma.$$

Note that for any (u, w) and $(\tilde{u}, \tilde{w})$ in I with $\|(u, w) - (\tilde{u}, \tilde{w})\| \leq \gamma$ and $u \leq \tilde{u}$, we have

$$\begin{aligned} E\left[\left(G_n(u, w) - G_n(\tilde{u}, \tilde{w})\right)^2\right] &= \mathbb{E}_n\left[\ell(u, w)' Z_i Z_i' \ell(u, w)\right](u - u^2) \\ &\quad + \mathbb{E}_n\left[\ell(\tilde{u}, \tilde{w})' Z_i Z_i' \ell(\tilde{u}, \tilde{w})\right](\tilde{u} - \tilde{u}^2) \\ &\quad - 2\mathbb{E}_n\left[\ell(u, w)' Z_i Z_i' \ell(\tilde{u}, \tilde{w})\right](u - u\tilde{u}) \\ &\leq I_1 + I_2 \lesssim \xi_\ell(\xi_\ell + L_\ell)\gamma, \end{aligned}$$

where

$$\begin{aligned} I_1 &= \left|\mathbb{E}_n\left[\ell(u, w)' Z_i Z_i' \ell(u, w)\right](u - u^2) - \mathbb{E}_n\left[\ell(u, w)' Z_i Z_i' \ell(\tilde{u}, \tilde{w})\right](u - u\tilde{u})\right| \\ &\leq 2\left|\mathbb{E}_n\left[\ell(u, w)' Z_i Z_i'(\ell(u, w) - \ell(\tilde{u}, \tilde{w}))\right]\right| + \left|\mathbb{E}_n\left[\ell(u, w)' Z_i Z_i' \ell(\tilde{u}, \tilde{w})\right](u - \tilde{u})\right| \\ &\lesssim \xi_\ell L_\ell \gamma + \xi_\ell^2 \gamma \end{aligned}$$

and

$$I_2 = \left| \mathbb{E}_n \Big[\ell(\tilde{u}, \tilde{w})' Z_i Z_i' \ell(\tilde{u}, \tilde{w}) \Big] (\tilde{u} - \tilde{u}^2) - \mathbb{E}_n \Big[\ell(u, w)' Z_i Z_i' \ell(\tilde{u}, \tilde{w}) \Big] (u - u\tilde{u}) \right|$$

$$\leq 2 \left| \mathbb{E}_n \Big[(\ell(\tilde{u}, \tilde{w}) - \ell(u, w))' Z_i Z_i' \ell(\tilde{u}, \tilde{w}) \Big] \right| + 2 \left| \mathbb{E}_n \Big[\ell(u, w)' Z_i Z_i' \ell(\tilde{u}, \tilde{w}) \Big] (u - \tilde{u}) \right|$$

$$\lesssim \xi_\ell L_\ell \gamma + \xi_\ell^2 \gamma.$$

Therefore,

$$\sigma(X_n) = \left(\sup_{t \in I \times I} E[X_{n,t}^2] \right)^{1/2} \lesssim \Big(\xi_\ell (\xi_\ell + L_\ell) \gamma \Big)^{1/2}.$$

The calculation above also implies that (G.111) holds for the process X_n with

$$\epsilon_0 = \sigma(X_n), \quad K \lesssim \sqrt{\xi_\ell (\xi_\ell + L_\ell)}, \quad \text{and } V = d_I,$$

since the set I has dimension d_I, which is independent of n, and its diameter is bounded from above uniformly over n. Therefore, the result follows by setting

$$\lambda = C \sqrt{\xi_\ell (\xi_\ell + L_\ell) \gamma \log(1/\gamma)} \lesssim \sqrt{L_\ell \gamma \log(1/\gamma)},$$

where C is a sufficiently large constant, and using (G.112). This completes the proof of the lemma. □

I.1. **Yurinskii's coupling for d-norm.** In this section, we extend the original Yurinskii's coupling, used in Appendix G, from the Euclidean norm $\| \cdot \|$ to arbitrary norm $\| \cdot \|_d$ on $\mathbb{R}^k$. The new coupling is used in Section 4 with the sup-norm $\| \cdot \|_\infty$ to construct Gaussian couplings for functionals. Replacing the Euclidean norm by the sup-norm is important because it allows us to construct couplings under substantially weakened side conditions.

Lemma 39 (Yurinskii's coupling, d-norm). *Let $\xi_1, \ldots, \xi_n$ be independent random zero-mean k-vectors, and let*

$$\beta = \sum_{i=1}^n E[\|\xi_i\|^2 \|\xi_i\|_d] + \sum_{i=1}^n E[\|g_i\|^2 \|g_i\|_d]$$

be finite, where $g_1, \ldots, g_n$ is a sequence of independent random k-vectors such that $g_i \sim N(0, \mathrm{var}(\xi_i))$ for all $i = 1, \ldots, n$. Let $S = \xi_1 + \cdots + \xi_n$. Then for each $\delta > 0$, there exists a random vector T with a $N(0, \mathrm{var}(S))$ distribution such that

$$P(\|S - T\|_d > 3\delta) \leq \min_{t \geq 0} \left(2P(\|Z\|_d > t) + \frac{\beta}{\delta^3} t^2 \right),$$

where $Z \sim N(0, I_k)$.

Proof of Lemma 39. Fix $\delta > 0$. By Strassen's theorem (see, for example, Theorem 8 in Section 10 of Pollard [69]), one can construct T such that $P(\|S-T\| > 3\delta) \leq \epsilon'$ if and only if $P(S \in A) \leq P(T \in A^{3\delta}) + \epsilon'$ for all Borel subsets A of $\mathbb{R}^k$, where $A^\delta = \{x \in \mathbb{R}^k \colon d(x, A) \leq \delta\}$ and the metric $d(\cdot, \cdot)$ is induced by the norm $\|\cdot\|$. Below we will apply this theorem.

Fix a Borel subset A of $\mathbb{R}^k$. Using Lemma 40, construct a smooth function $f \colon \mathbb{R}^k \to \mathbb{R}$ that approximates the indicator function of A, namely, a function f such that for all $x, y \in \mathbb{R}^k$,

$$\left| f(x+y) - f(x) - y'\nabla f(x) - \frac{1}{2}y'\nabla^2 f(x)y \right| \leq \frac{\|y\|^2\|y\|_d}{\sigma^2\delta},$$

$$(1-\epsilon)\mathbf{1}\{x \in A\} \leq f(x) \leq \epsilon + (1-\epsilon)\mathbf{1}\{x \in A^{3\delta}\},$$

where $\epsilon = P(\|Z\|_d > \delta/\sigma)$ and σ is to be chosen below. Then we have

$$P(S \in A) = E[\mathbf{1}\{S \in A\} - f(S)] + E[f(S) - f(T)] + E[f(T)]$$

$$\leq \epsilon E[\mathbf{1}\{S \in A\}] + (\sigma^2\delta)^{-1}\sum_{i=1}^{n} E\left[\|\xi_i\|^2\|\xi_i\|_d + \|g_i\|^2\|g_i\|_d\right]$$

$$+ \epsilon + (1-\epsilon)E[\mathbf{1}\{T \in A^{3\delta}\}]$$

$$\leq P(T \in A^{3\delta}) + 2\epsilon + (\sigma^2\delta)^{-1}\sum_{i=1}^{n} E\left[\|\xi_i\|^2\|\xi_i\|_d + \|g_i\|^2\|g_i\|_d\right]$$

$$= P(T \in A^{3\delta}) + 2\epsilon + (\sigma^2\delta)^{-1}\beta,$$

where in the first inequality we used

$$E[f(S) - f(T)] = \sum_{i=1}^{n} E[f(X_i + Y_i) - f(X_i + W_i)]$$

$$\leq \sum_{i=1}^{n} E\left[f(X_i) + Y_i'\nabla f(X_i) + \frac{1}{2}Y_i'\nabla^2 f(X_i)Y_i + \frac{\|Y_i\|^2\|Y_i\|_d}{\sigma^2\delta}\right]$$

$$- \sum_{i=1}^{n} E\left[f(X_i) + W_i'\nabla f(X_i) + \frac{1}{2}W_i'\nabla^2 f(X_i)W_i - \frac{\|W_i\|^2\|W_i\|_d}{\sigma^2\delta}\right]$$

$$= \sum_{i=1}^{n} \frac{\|Y_i\|^2\|Y_i\|_d + \|W_i\|^2\|W_i\|_d}{\sigma^2\delta}$$

for $X_i = \xi_1 + \cdots + \xi_{i-1} + g_{i+1} + \cdots + g_n$, $Y_i = \xi_i$, $W_i = g_i$, and we assumed (without loss of generality) that the sequences $(\xi_i)_{i=1}^n$ and $(g_i)_{i=1}^n$ are independent. Therefore, setting $\sigma = \delta/t$ for $t > 0$, we obtain

$$P(S \in A) \leq P(T \in A^{3\delta}) + 2P(\|Z\|_d > t) + \frac{\beta}{\delta^3}t^2.$$

The asserted claim now follows by minimizing the right-hand of this inequality with respect to $t > 0$ and applying Strassen's theorem. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

In the next lemma, we construct a function f used in the previous lemma.

Lemma 40 (Smooth Approximation, d-norm)**.** *Let A be a Borel subset of $\mathbb{R}^k$, Z be a $N(0, I_k)$ random vector, and let $d(\cdot, \cdot)$ be the metric on $\mathbb{R}^k$ induced by a norm $\| \cdot \|_d$. Also, for positive constants σ and δ, define*

$$g(x) = \left(1 - \frac{d(x, A^\delta)}{\delta}\right)_+ \quad and \quad f(x) = E[g(x + \sigma Z)].$$

Then for all $x, y \in \mathbb{R}^k$, the function f satisfies

$$\left| f(x + y) - f(x) - y'\nabla f(x) - \frac{1}{2} y'\nabla^2 f(x) y \right| \le \frac{\|y\|^2 \|y\|_d}{\sigma^2 \delta},$$

$$(1 - \epsilon) 1\{x \in A\} \le f(x) \le \epsilon + (1 - \epsilon) 1\{x \in A^{3\delta}\},$$

where $\epsilon = P(\|Z\|_d > \delta/\sigma)$.

Proof. The proof follows closely the proof of Lemma 18 in Section 10 of Pollard [69], with the difference that we allow for a general metric $d(\cdot, \cdot)$. Let ϕ_σ denote the pdf of a $N(0, \sigma^2 I_k)$ random vector. We have

$$\frac{\partial}{\partial z}\phi_\sigma(z) = -\frac{z}{\sigma^2}\phi_\sigma(z) \quad \text{and} \quad \frac{\partial^2}{\partial z z'}\phi_\sigma(z) = \left(\frac{zz'}{\sigma^4} - \frac{I_k}{\sigma^2}\right)\phi_\sigma(z).$$

Further, for fixed x and y, consider the function $h(t) := f(x + ty)$, $0 \le t \le 1$. Its second derivative is

$$\ddot{h}(t) = \sigma^{-2} E[g(x + ty + \sigma Z)\left((y'Z)^2 - \|y\|^2\right)].$$

In addition, the function g has the following Lipschitz property with respect to the metric d:

$$|g(x + ty + \sigma Z) - g(x + \sigma Z)| \le \left| \frac{d(x + ty + \sigma Z, A^\delta) - d(x + \sigma Z, A^\delta)}{\delta} \right|$$

$$\le d(x + ty + \sigma Z, x + \sigma Z)/\delta = t d(y, 0)/\delta.$$

Therefore,

$$|\ddot{h}(t) - \ddot{h}(0)| \le \frac{t d(y, 0)}{\sigma^2 \delta} E\left[(y'Z)^2 + \|y\|^2\right] = \frac{2t\|y\|^2 d(y, 0)}{\sigma^2 \delta}.$$

The first asserted claim now follows from a Taylor expansion for the function h, namely, we have for some $t^* \in (0, 1)$ that

$$|h(1) - h(0) - \dot{h}(0) - \frac{1}{2}\ddot{h}(0)| = \frac{1}{2}|\ddot{h}(t^*) - \ddot{h}(0)|.$$

To establish the second asserted claim, we proceed as follows. By construction, we have for any $x \in A$ that $g(x + \sigma Z) = 1$ if $d(Z, 0) \leq \delta/\sigma$ since $x + \sigma Z \in A^\delta$ in this case. Therefore,

$$f(x) \geq E\Big[g(x + \sigma Z) \cdot 1\{d(Z, 0) \leq \delta/\sigma\}\Big] \geq P(d(Z, 0) \leq \delta/\sigma) = 1 - P(d(Z, 0) > \delta/\sigma).$$

Also by construction, we have for any $x \notin A^{3\delta}$ that $g(x + \sigma Z) = 0$ if $d(Z, 0) \leq \delta/\sigma$ since $d(x + \sigma Z, A^\delta) > \delta$ in this case. Therefore

$$f(x) = E\Big[g(x+\sigma Z)\cdot 1\{d(Z, 0) \leq \delta/\sigma\}\Big] + E\Big[g(x+\sigma Z)\cdot 1\{d(Z, 0) > \delta/\sigma\}\Big] \leq P(d(0, Z) > \delta/\sigma).$$

This completes the proof of the lemma. □

Appendix J. A Lemma on Gaussian Approximation for the Supremum of a Weighted Bootstrap Process

Lemma 41. *(Gaussian Approximation for Supremum of Weighted Bootstrap Process) Let $(Z_i)_{i=1}^n$ be a sequence of non-stochastic vectors in $\mathbb{R}^m$ and consider the empirical process $\mathbb{U}_n$ in $[\ell^\infty(\mathcal{U})]^m$, $\mathcal{U} \subseteq (0, 1)$, defined by*

$$\mathbb{U}_n(u) = \mathbb{G}_n\left(v_i Z_i \psi_i(u)\right), \quad \psi_i(u) = u - 1\{U_i \leq u\}, \quad u \in \mathcal{U},$$

where $(U_i, v_i)_{i=1}^n$ is an i.i.d. sequence of pairs of independent random variables where $U_i \sim Uniform(0, 1)$, $E[v_i] = 0$, $E[v_i^2] = 1$, $E[|v_i|^4] \lesssim 1$, and $\max_{1 \leq i \leq n} |v_i| \lesssim_P \log n$. Suppose that the vectors Z_i are such that

$$\sup_{\alpha \in S^{m-1}} \mathbb{E}_n\left[(\alpha' Z_i)^2\right] \lesssim 1 \quad and \quad \max_{1 \leq i \leq n} \|Z_i\| \lesssim \zeta_m,$$

where ζ_m satisfies $1/\zeta_m \lesssim 1$ and $\zeta_m^2 = o(n^{1-\varepsilon})$ for some $\varepsilon > 0$. Also, let $\mathcal{W}$ be a set in $\mathbb{R}^d$ and let I be a subset of $\mathcal{U} \times \mathcal{W}$ whose dimension d_I is independent of n and whose diameter is bounded uniformly over n. Moreover, let $\ell \colon I \to \mathbb{R}^m$ be a function such that

$$\|\ell(u, w)\| \leq \xi_\ell \quad and \quad \|\ell(u, w) - \ell(\tilde{u}, \tilde{w})\| \leq L_\ell \|(u, w) - (\tilde{u}, \tilde{w})\|$$

for all (u, w) and $(\tilde{u}, \tilde{w})$ in I, where ξ_ℓ and L_ℓ satisfy $\xi_\ell \lesssim 1$ and $\log L_\ell \lesssim \log n$, and define

$$V_n = \sup_{(u,w) \in I} |\ell(u, w)' \mathbb{U}_n(u)|.$$

Then for each n, there exists a random variable $\bar{V}_n$ that is (i) such that

$$|V_n - \bar{V}_n| = o_P(n^{-\varepsilon'}) \tag{J.141}$$

for some $\varepsilon' > 0$, (ii) independent of $(U_i)_{i=1}^n$, and (iii) equal in distribution to the random variable $\sup_{(u,w) \in I} |G_n(u,w)|$, where $G_n(\cdot, \cdot)$ is a zero-mean Gaussian process with a.s. continuous sample paths and the covariance function

$$E[G_n(u,w)G_n(\tilde{u}, \tilde{w})] = \mathbb{E}_n[\ell(u,w)' Z_i Z_i' \ell(\tilde{u}, \tilde{w})](u \wedge \tilde{u} - u\tilde{u}),$$

for all (u,w) and $(\tilde{u}, \tilde{w})$ in I.

Proof. Throughout the proof, we will take the process $G_n(\cdot, \cdot)$ from the statement of the lemma to be independent of $(U_i, v_i)_{i=1}^n$. In addition, we will use $\mathcal{B}$ to denote the class of all Borel sets in $\mathbb{R}$.

To prove the asserted claim, we will show that there exists a sequence of positive numbers $(\eta_n)_{n \geq 1}$ converging to zero $\varepsilon' > 0$ such that for

$$\mathcal{U}_1^n = \left\{ (U_i)_{i=1}^n \colon P(V_n \in A | (U_i)_{i=1}^n) \leq P \left(\sup_{(u,w) \in I} |G_n(u,w)| \in A^{\delta_n} \right) + \eta_n \text{ for all } A \in \mathcal{B} \right\}$$

we have

$$P(\mathcal{U}_1^n) \geq 1 - o(1), \tag{J.142}$$

where $\delta_n = n^{-\varepsilon'}$ and $A^{\delta_n} = \{ x \in \mathbb{R} \colon \inf_{y \in A} |x - y| \leq \delta_n \}$. Then applying Strassen's theorem conditional on $(U_i)_{i=1}^n$ on the event $\mathcal{U}_1^n$ shows that on this event one can construct a random variable $\bar{V}_n$ that is conditional on $(U_i)_{i=1}^n$ equal in distribution to $\sup_{(u,w) \in I} |G_n(u,w)|$ and is such that

$$P \left(|V_n - \bar{V}_n| > \delta_n | (U_i)_{i=1}^n \right) \leq \eta_n \quad \text{on } \mathcal{U}_1^n. \tag{J.143}$$

Outside of the event $\mathcal{U}_1^n$, we define $\bar{V}_n = \sup_{(u,w) \in I} |G_n(u,w)|$. Then combining (J.142) with (J.143) gives (J.141) and since the conditional distribution of $\bar{V}_n$ given $(U_i)_{i=1}^n$ is equal to that of $\sup_{(u,w) \in I} |G_n(u,w)|$, it follows that $\bar{V}_n$ is independent of $(U_i)_{i=1}^n$ and is equal in distribution to $\sup_{(u,w) \in I} |G_n(u,w)|$. Thus, it remains to prove (J.142).

To prove (J.142), recall that $\log L_\ell \lesssim \log n$ by assumption, and so $\log L_\ell \leq C_L \log n$ for some constant C_L. Let $\gamma_n = 1/n^{C_L+1}$ and note that since the set I is such that its dimension is independent of n and its diameter is bounded uniformly over n, there exists a sequence $(u_j, w_j)_{j=1}^{k_n}$ in I with $\log k_n \lesssim \log n$ such that balls with centers at (u_j, w_j) and radius γ_n cover I. Then applying Lemma 37 shows that

$$\sup_{\|(u,w) - (\tilde{u}, \tilde{w})\| \leq \gamma_n} \left| \ell(u,w)' \mathbb{U}_n(u) - \ell(\tilde{u}, \tilde{w})' \mathbb{U}_n(\tilde{u}) \right|$$

$$\lesssim_P \sqrt{(L_\ell^2 \gamma_n^2 + \gamma_n) \log n} + \sqrt{\frac{\zeta_m^2 \log^4 n}{n}} = o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$, and so there exists a sequence of positive numbers $(\eta_{n,1})_{n \geq 1}$ converging to zero such that the event

$$\left| \sup_{(u,w) \in I} |\ell(u,w)' \mathbb{U}_n(u)| - \max_{1 \leq j \leq k_n} |\ell(u_j, w_j) \mathbb{U}_n(u_j)| \right| > \delta_{n,1} = n^{-\varepsilon'}$$

holds with probability at most $\eta_{n,1}$. Hence, there exists a set of values of $(U_i)_{i=1}^n$, say $\mathcal{U}_{1,1}^n$, such that $P(\mathcal{U}_{1,1}^n) \geq 1 - \eta_{n,1}^{1/2} = 1 - o(1)$ and

$$P \left(\left| \sup_{(u,w) \in I} |\ell(u,w)' \mathbb{U}_n(u)| - \max_{1 \leq j \leq k_n} |\ell(u_j, w_j) \mathbb{U}_n(u_j)| \right| > \delta_{n,1} | (U_i)_{i=1}^n \right) \leq \eta_{n,1}^{1/2} \quad \text{on } \mathcal{U}_{1,1}^n. \tag{J.144}$$

Next, let Σ^X and Σ^Y be $k_n \times k_n$-dimensional matrices given by

$$\Sigma_{j,l}^X = \mathbb{E}_n \left[\ell(u_j, w_j)' Z_i Z_i' \ell(u_l, w_l)(u_j - 1\{U_i \leq u_j\})(u_l - 1\{U_i \leq u_l\}) \right], \tag{J.145}$$

$$\Sigma_{j,l}^Y = \mathbb{E}_n \left[\ell(u_j, w_j)' Z_i Z_i' \ell(u_l, w_l) \right] (u_j \wedge u_l - u_j u_l), \tag{J.146}$$

for all $j, l = 1, \ldots, k_n$. Note that $E[\Sigma_{j,l}^X] = \Sigma_{j,l}^Y$ for all $j, l = 1, \ldots, k_n$. In addition,

$$\sigma^2 = \max_{1 \leq j,l \leq k_n} \sum_{i=1}^n E \left[(\ell(u_j, w_j)' Z_i Z_i' \ell(u_l, w_l))^2 (u_j - 1\{U_i \leq u_j\})^2 (u_l - 1\{U_i \leq u_l\})^2 \right]$$

$$\leq \max_{1 \leq j,l \leq k_n} \sum_{i=1}^n (\ell(u_j, w_j)' Z_i Z_i' \ell(u_l, w_l))^2 \lesssim n\zeta_m^2.$$

Moreover,

$$M = \max_{1 \leq i \leq n} \max_{1 \leq j,l \leq k_n} \left| \ell(u_j, w_j)' Z_i Z_i' \ell(u_l, w_l)(u_j - 1\{U_i \leq u_j\})(u_l - 1\{U_i \leq u_l\}) \right| \lesssim \zeta_m^2.$$

Hence, Lemma 42 gives

$$\max_{1 \leq j,l \leq k_n} |\Sigma_{j,l}^X - \Sigma_{j,l}^Y| \lesssim_P \sqrt{\frac{\zeta_m^2 \log n}{n}} + \frac{\zeta_m^2 \log n}{n} = o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$. Hence, there exists a set of values of $(U_i)_{i=1}^n$, say $\mathcal{U}_{1,2}^n$, such that $P(\mathcal{U}_{1,2}^n) = 1 - o(1)$ and

$$\max_{1 \leq j,l \leq k_n} |\Sigma_{j,l}^X - \Sigma_{j,l}^Y| \leq \delta_{n,2} = n^{-\varepsilon'} \quad \text{on } \mathcal{U}_{1,2}^n.$$

In the rest of the proof, we will show that

$$\mathcal{U}_{1,1}^n \cap \mathcal{U}_{1,2}^n \subset \mathcal{U}_1^n \tag{J.147}$$

if ε' is small enough and η_n converges to zero slowly enough in the definition of $\mathcal{U}_1^n$. The desired inequality (J.142) then follows since $P(\mathcal{U}_{1,1}^n) = 1 - o(1)$ and $P(\mathcal{U}_{1,2}^n) = 1 - o(1)$.

To prove (J.147), we proceed in several steps. First, we apply Lemma 43 conditional on $(U_i)_{i=1}^n$ with vectors $X_i = (\ell(u_j, w_j)'Z_i v_i(u_j - 1\{U_i \le u_j\}))_{j=1}^{k_n}$ and $Y_i = (\ell(u_j, w_j)'Z_i e_i(u_j - 1\{U_i \le u_j\}))_{j=1}^{k_n}$, where $(e_i)_{i=1}^n$ is a sequence of n independent $N(0,1)$ random variables that are independent of $(U_i)_{i=1}^n$. Note that

$$
\begin{aligned}
L_n &= \max_{1 \le j \le k_n} \frac{1}{n} \sum_{i=1}^n E[|X_{ij}|^3|(U_i)_{i=1}^n] \\
&= \max_{1 \le j \le k_n} \frac{1}{n} \sum_{i=1}^n E\Big[|\ell(u_j, w_j)'Z_i v_i(u_j - 1\{U_i \le u_j\})|^3|(U_i)_{i=1}^n\Big] \\
&= \max_{1 \le j \le k_n} \frac{1}{n} \sum_{i=1}^n |\ell(u_j, w_j)'Z_i|^3 \le \max_{1 \le j \le k_n} \frac{\zeta_m}{n} \sum_{i=1}^n |\ell(u_j, w_j)'Z_i|^2 \lesssim \zeta_m.
\end{aligned}
$$

Further, uniformly over $\delta > 0$,

$$
\begin{aligned}
M_{n,X}(\delta) &= \frac{1}{n} \sum_{i=1}^n E\Big[\max_{1 \le j \le k_n} |X_{ij}|^3 \cdot 1\Big\{\max_{1 \le j \le k_n} |X_{ij}| > \delta\sqrt{n}/\log k_n\Big\}|(U_i)_{i=1}^n\Big] \\
&\le \frac{\log k_n}{\delta n^{3/2}} \sum_{i=1}^n E\Big[\max_{1 \le j \le k_n} |X_{ij}|^4|(U_i)_{i=1}^n\Big] \\
&\lesssim \frac{\log k_n}{\delta n^{3/2}} \sum_{i=1}^n \max_{1 \le j \le k_n} |\ell(u_j, w_j)'Z_i|^4 \lesssim \frac{\zeta_m^4 \log n}{\delta\sqrt{n}}.
\end{aligned}
$$

In addition, by the same argument, uniformly over $\delta > 0$,

$$
M_{n,Y}(\delta) = \frac{1}{n} \sum_{i=1}^n E\Big[\max_{1 \le j \le k_n} |Y_{ij}|^3 \cdot 1\Big\{\max_{1 \le j \le k_n} |Y_{ij}| > \delta\sqrt{n}/\log k_n\Big\}|(U_i)_{i=1}^n\Big] \lesssim \frac{\zeta_m^4 \log n}{\delta\sqrt{n}}.
$$

Hence, given that $\zeta_m^2 = o(n^{1-\varepsilon})$ and that

$$
\ell(u_j, w_j)'\mathbb{U}_n(u_j) = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_{ij}, \quad j = 1, \ldots, k_n,
$$

and denoting

$$
\bar{G}_{n,j} = \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_{ij}, \quad j = 1, \ldots, k_n,
$$

and $\delta_{n,3} = n^{-\varepsilon'}$ for sufficiently small $\varepsilon' > 0$, we obtain via Lemma 43 that

$$
P\left(\max_{1 \le j \le k_n} |\ell(u_j, w_j)'\mathbb{U}_n(u_j)| \in A|(U_i)_{i=1}^n\right) \le P\left(\max_{1 \le j \le k_n} |\bar{G}_{n,j}| \in A^{\delta_{n,3}}|(U_i)_{i=1}^n\right) + \eta_{n,2}
$$

(J.148)

for all $A \in \mathcal{B}$, where $(\eta_{n,2})_{n \ge 1}$ is a sequence of positive numbers converging to zero.

Second, we apply Lemma 44 conditional on $(U_i)_{i=1}^n$ on $\mathcal{U}_1^n$ with $X = (\bar{G}_{n,j})_{j=1}^{k_n}$ and $Y = (G_n(u_j, w_j))_{j=1}^{k_n}$. Note that conditional covariance matrices of the vectors X and Y are given by

$$E[X_j X_l | (U_i)_{i=1}^n] = \Sigma_{j,l}^X, \quad E[Y_j Y_l | (U_i)_{i=1}^n] = E[Y_j Y_l] = \Sigma_{j,l}^Y, \quad j, l = 1, \ldots, k_n$$

for $\Sigma_{j,l}^X$ and $\Sigma_{j,l}^Y$ defined in (J.145) and (J.146), respectively. Since on the event $\mathcal{U}_{1,2}^n$ we have

$$\max_{1 \leq j, l \leq k_n} |\Sigma_{j,l}^X - \Sigma_{j,l}^Y| \leq \delta_{n,2},$$

it follows from Lemma 44 that on the same event we have

$$P\left(\max_{1 \leq j \leq k_n} |\bar{G}_{n,j}| \in A | (U_i)_{i=1}^n \right) \leq P\left(\max_{1 \leq j \leq k_n} |G_n(u_j, w_j)| \in A^{\delta_{n,2}^{1/3}} \right) + \eta_{n,3} \qquad \text{(J.149)}$$

for all $A \in \mathcal{B}$, where $(\eta_{n,3})_{n \geq 1}$ is a sequence of positive numbers converging to zero.

Third, by Lemma 38,

$$\sup_{\|(u,w)-(\tilde{u},\tilde{w})\| \leq \gamma_n} \left| G_n(u, w) - G_n(\tilde{u}, \tilde{w}) \right| \lesssim_P \sqrt{L_\ell \gamma_n \log(1/\gamma_n)} = o(n^{-\varepsilon'})$$

for some $\varepsilon' > 0$, and so

$$P\left(\left| \sup_{(u,w) \in I} |G_n(u, w)| - \max_{1 \leq j \leq k_n} |G_n(u_j, w_j)| \right| > \delta_{n,4} \right) \leq \eta_{n,4}, \qquad \text{(J.150)}$$

where $\delta_{n,4} = n^{-\varepsilon'}$ and $(\eta_{n,4})_{n \geq 1}$ is a sequence of positive numbers converging to zero.

Finally, combining (J.144), (J.148), (J.149), and (J.150) shows that for all $A \in \mathcal{B}$,

$$P\left(\sup_{(u,w) \in I} |\ell(u, w)' \mathbb{U}_n(u)| \in A | (U_i)_{i=1}^n \right) \leq P\left(\sup_{(u,w) \in I} |G_n(u, w)| \in A^{\delta_{n,1} + \delta_{n,2}^{1/3} + \delta_{n,3} + \delta_{n,4}} \right)$$
$$+ \eta_{n,1}^{1/2} + \eta_{n,2} + \eta_{n,3} + \eta_{n,4} \quad \text{on } \mathcal{U}_{1,1}^n \cap \mathcal{U}_{1,2}^n.$$

Thus, (J.147) follows if in the definition of $\mathcal{U}_1^n$ we set $\eta_n = \eta_{n,1}^{1/2} + \eta_{n,2} + \eta_{n,3} + \eta_{n,4}$ and $\delta_n = n^{-\varepsilon'}$ for sufficiently small $\varepsilon' > 0$ so that $\delta_n \geq \delta_{n,1} + \delta_{n,2}^{1/3} + \delta_{n,3} + \delta_{n,4}$. This completes the proof of the lemma. □

J.1. Technical Lemmas.

Lemma 42. *Let $Z_1, \ldots, Z_n$ be independent random vectors in $\mathbb{R}^p$ with $p \geq 2$. Define $M = \max_{1 \leq i \leq n} \max_{1 \leq j \leq p} |Z_{ij}|$ and $\sigma^2 = \max_{1 \leq j \leq p} \sum_{i=1}^n E[Z_{ij}^2]$. Then*

$$E\left[\max_{1 \leq j \leq p} \left| \sum_{i=1}^n (Z_{ij} - E[Z_{ij}]) \right| \right] \leq C\left(\sigma \sqrt{\log p} + \sqrt{E[M^2]} \log p \right),$$

where C is an absolute constant.

Proof. See Lemma 8 in Chernozhukov, Chetverikov and Kato [27]. ☐

Lemma 43. *Let $X_1, \ldots, X_n$ be independent centered random vectors in $\mathbb{R}^p$, $p \geq 2$, with finite absolute third moments. Consider the statistic $Z = \max_{1 \leq j \leq p} n^{-1/2} \sum_{i=1}^{n} X_{ij}$. Let $Y_1, \ldots, Y_n$ be independent random vectors in $\mathbb{R}^p$ with $Y_i \sim N(0_p, E[X_i X_i'])$, and define $\widetilde{Z} = \max_{1 \leq j \leq p} n^{-1/2} \sum_{i=1}^{n} Y_{ij}$. Then for every $\delta > 0$ and every Borel subset A of $\mathbb{R}$, we have*

$$P(Z \in A) \leq P(\widetilde{Z} \in A^{C_1 \delta}) + \frac{C_2 \log^2 p}{\delta^3 \sqrt{n}} \cdot \Big(L_n + M_{n,X}(\delta) + M_{n,Y}(\delta)\Big),$$

where C_1 and C_2 are absolute constants and

$$L_n = \max_{1 \leq j \leq p} \frac{1}{n} \sum_{i=1}^{n} E[|X_{ij}|^3],$$

$$M_{n,X}(\delta) = \frac{1}{n} \sum_{i=1}^{n} E\Big[\max_{1 \leq j \leq p} |X_{ij}|^3 \cdot 1\Big\{ \max_{1 \leq j \leq p} |X_{ij}| > \delta \sqrt{n}/\log p \Big\} \Big],$$

$$M_{n,Y}(\delta) = \frac{1}{n} \sum_{i=1}^{n} E\Big[\max_{1 \leq j \leq p} |Y_{ij}|^3 \cdot 1\Big\{ \max_{1 \leq j \leq p} |Y_{ij}| > \delta \sqrt{n}/\log p \Big\} \Big].$$

Proof. See Theorem 3.1 in Chernozhukov, Chetverikov and Kato [28]. ☐

Lemma 44. *Let $X = (X_1, \ldots, X_p)'$ and $Y = (Y_1, \ldots, Y_p)'$ be random vectors in $\mathbb{R}^p$, $p \geq 2$, with $X \sim N(0_p, \Sigma^X)$ and $Y \sim N(0_p, \Sigma^Y)$. Let $\Delta = \max_{1 \leq j,k \leq p} |\Sigma_{jk}^X - \Sigma_{jk}^Y|$, where Σ_{jk}^X and Σ_{jk}^Y denote the (j,k)-th elements of Σ^X and Σ^Y, respectively. Define $Z = \max_{1 \leq j \leq p} X_j$ and $\widetilde{Z} = \max_{1 \leq j \leq p} Y_j$. Then for every $\delta > 0$ and every Borel subset A of $\mathbb{R}$,*

$$P(Z \in A) \leq P(\widetilde{Z} \in A^\delta) + C\delta^{-1}\sqrt{\Delta \log p},$$

where C is an absolute constant.

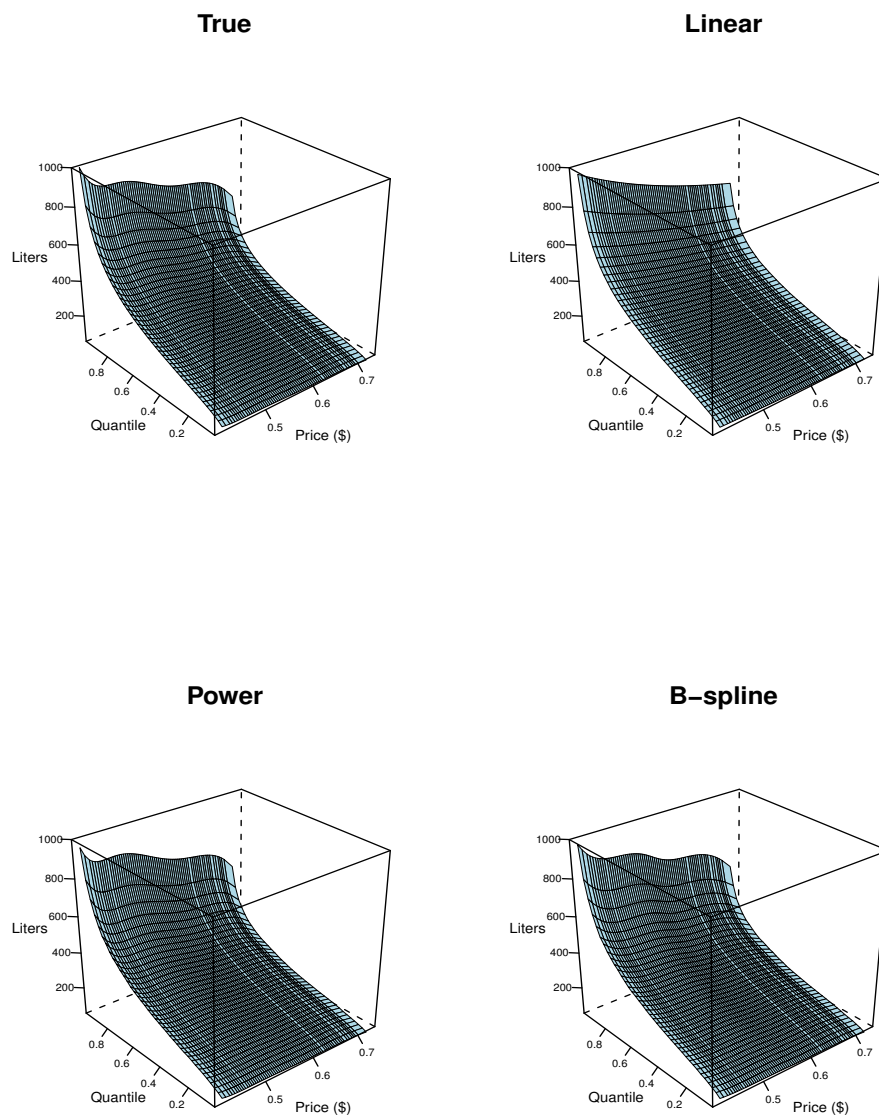*Proof.* See Theorem 3.2 in Chernozhukov, Chetverikov and Kato [28]. ☐

FIGURE 3. Estimands of the quantile demand surface. Estimands for the linear, power and B-spline series estimators are obtained numerically using 500,100 simulations.
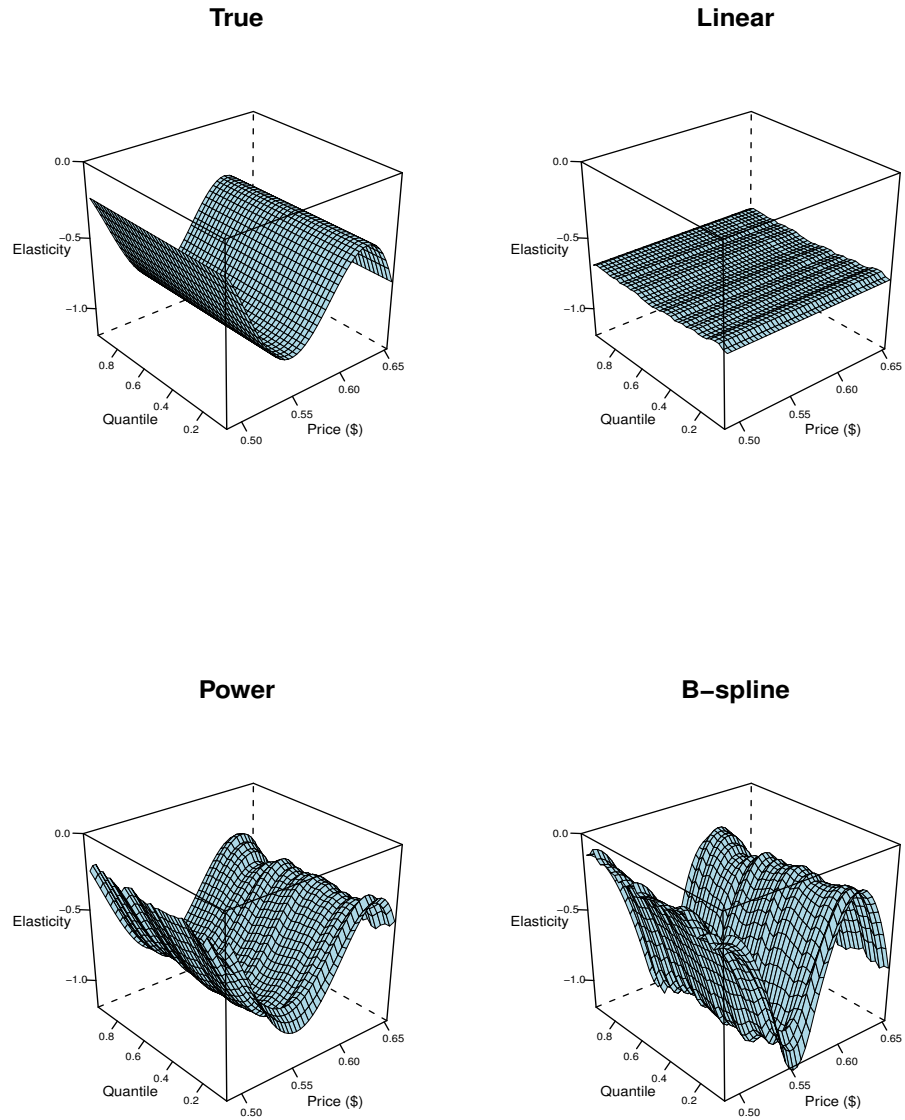
True

Linear

Power

B−spline

FIGURE 4.   Estimands of the quantile price elasticity surface. Estimands for the linear, power and B-spline series estimators are obtained numerically using 500,100 simulations.

References

[1] Andrews, D. W. K. (1994). Empirical process methods in econometrics. *Handbook of Econometrics, Volume IV, Chapter 37,* 2247–2294.

[2] Andrews, D. W. K. (1991). Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica* **59** 307–345.

[3] Andrews, D. W. K. and Soares, G. (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* **78** 119–157.

[4] Angrist, J., Chernozhukov, V., and Fernández-Val, I. (2006). Quantile regression under misspecification, with an application to the u.s. wage structure. *Econometrica* **74** 539–563.

[5] Arias, O., Hallock, K. F., and Sosa-Escudero, W. (2001) Individual heteroneity in the returns to schooling: instrumental variables quantile regression using twins data. *Empirical Economics* **26** 7–40.

[6] Barlow, R., Bartholomew, D., Bremner, J., and Brunk, H. (1972). *Statistical Inference Under Order Restrictions.* John Wiley, New York.

[7] Belloni, A. and Chernozhukov, V. (2011). ℓ_1-penalized quantile regression for high dimensional sparse models. *Ann. Statist.* **39** 82–130.

[8] Belloni, A., Chernozhukov, V., Chetverikov, D., and Fernandez-Val, I. (2016). Supplement to "Conditional Quantile Processes Based on Series or Many Regressors". *Unpublished manuscript.*

[9] Belloni, A., Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). On the asymptotic theory for least squares series: pointwise and uniform results. Journal of Econometrics **186** 345–366.

[10] Belloni, A., Chernozhukov, V., and Kato, K. (2013). Valid Post-Selection Inference in High-Dimensional Approximately Sparse Quantile Regression Models. *arXiv preprint arXiv:1312.7186.*

[11] Bickel, P. and Rosenblatt, M. (1973). On some global measures of the deviations of density function estimates. The Annals of Statistics **1** 1071–1095.

[12] Blundell, R., Horowitz, J., and Parey, M. (2012). Measuring the price responsiveness of gasoline demand: Economic shape restrictions and nonparametric demand estimation. *Quantitative Economics* **3** 29–51.

[13] Buchinsky, M. (1994). Changes in the u.s. wage structure 1963-1987: Application of quantile regression. *Econometrica* **62** 405–458.

[14] Cattaneo, M., Crump, R., and Jansson, M. (2010). Robust data-driven inference for density-weighted average derivatives. *J. Amer. Statist. Assoc.* **105(491)** 1070–1083. With supplementary material available online.

[15] Cattaneo, M. and Farrell, M. (2013). Optimal convergence rates, Bahadur representation, and asymptotic normality of partitioning estimators. *Journal of Econometrics* **174** 127-143.

[16] Cattaneo, M., Jansson, M., and Newey, W. (2015). Treatment effects with many covariates and heteroskedasticity. *Working paper.*

[17] Chamberlain, G. and Imbens, G. (2003). Nonparametric applications of Bayesian inference. *J. Bus. Econom. Statist.* **21** 12–18.

[18] Chaudhuri, P. (1991a). Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.* **19** 760–777.

[19] Chaudhuri, P. (1991b). Global nonparametric estimation of conditional quantile functions and their derivatives. *Journal of Multivariate Analysis,* **39** 246–269.

[20] Chaudhuri, P., Doksum, K., and Samarov, A. (1997). On average derivative quantile regression. *Ann. Statist.* **25** 715–744.

[21] Chen, X. (2006). Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics, Volume VI, Chapter 76,* 5549–5632.

[22] Chen, X. and Christensen, T. (2013). Optimal sup-norm rates, adaptivity and inference in nonparametric instrumental variables estimation. *arXiv:1508.03365.*

[23] Chen, X. and Pouzo, D. (2009). Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals. *J. Econometrics* **152** 46–60.

[24] Chen, X. and Shen, X. (1998). Sieve extremum estimates for weakly dependent data. *Econometrica* **66** 289–314.

[25] Chernozhukov, V., Chetverikov, D., and Kato, K. (2013). Anti-concentration and honest adaptive confidence bands. *Ann. Statist.* **42** 1787–1818.

[26] Chernozhukov, V., Chetverikov, D., and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597.

[27] Chernozhukov, V., Chetverikov, D., and Kato, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields* **162** 47–70.

[28] Chernozhukov, V., Chetverikov, D., and Kato, K. (2016). Empirical and multiplier bootstraps for suprema of empirical processes of increasing complexity, and related Gaussian couplings. *Stochastic Processes and their Applications, forthcoming.*

[29] Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika* **96** 559–575.

[30] Chernozhukov, V., Lee, S., and Rosen, A. (2013). Intersection bounds: estimation and inference. *Econometrica* **81** 667-737.

[31] Chernozhukov, V., Hong, H., and Tamer, E. (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica* **75** 1243–1284.

[32] Claeskens, G. and Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *The Annals of Statistics* **31** 1852–1884.

[33] Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.,* **74(368)** 829–836.

[34] Dette, H., Hoderlein, S., and Neumeyer, N. (2011). Testing multivariate economic restrictions using quantiles: The example of slutsky negative semidefiniteness.

[35] DeVore, R. and Lorentz, G. (1993). *Constructive Approximation.* Springer.

[36] Dudley, R. (2000). *Uniform Cental Limit Theorems.* Cambridge Studies in advanced mathematics.

[37] Dudley, R. (2004). *Real Analysis and Probability.* Cambridge University Press

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

[39] Guédon, O. and Rudelson, M. (2007). l_p-moments of random vectors via majorizing measures. *Advances in Mathematics* **208** 798–823.

[40] Guerre, E., and Sabbah, C. (2012) Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory* **28 (1)**, 87–129.

[41] Hahn, J. (1997). Bayesian bootstrap of the quantile regression estimator: a large sample study. *Internat. Econom. Rev.* **38** 795–808.

[42] Hall, P. and Sheather, S. (1988). On the distribution of a studentized quantile. *Journal of the Royal Statistical Society. Series B (Methodological)* **50** 381–391.

[43] Härdle, W., Ritov, Y., and Song, S. (2009). Partial linear quantile regression and bootstrap confidence bands. SFB 649 Discussion Paper 2010-002.

[44] Hausman, J. and Newey, W. (1995). Nonparametric estimation of exact consumers surplus and deadweight loss. *Econometrica* **63** 1445–1476.

[45] He, X. and Shao, Q.-M. (2000). On parameters of increasing dimentions. *Journal of Multivariate Analysis* **73** 120–135.

[46] Hoderlein, S. and Mammen, E. (2009). Identification and estimation of local average derivatives in non-separable models without monotonicity. *The Econometrics Journal* **12** 1–25.

[47] Horowitz, J. and Lee, S. (2005). Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association* **100(472)** 1238–1249.

[48] Horowitz, J. and Lee, S. (2009). Uniform confidence bands for functions estimated nonparametrically with instrumental variables. *Cemmap working paper*.

[49] Huang, J. (1998). Projection estimation in multiple regression with application to functional anova models. *Ann. Statist.* **26** 242–272.

[50] Huang, J. (2003). Local asymptotics for polynomial spline regression. *Ann. Statist.* **31** 1600–1635.

[51] Koenker, R. (2005). *Quantile regression.* Cambridge University Press, New York.

[52] Koenker, R. and Basset, G. (1978). Regression quantiles. *Econometrica* **46** 33–50.

[53] Koenker, R. (2008) *quantreg: Quantile Regression.* R package version 4.24.

[54] Koenker, R. and Portnoy, S. (1987). L-estimation for linear models. *Journal of the American Statistical Association* **82(399)** 851–857.

[55] Koenker, R. and Schorfheide, F. (1994). Quantile spline models for global temperature change. *Climatic Change* **28** 395–404.

[56] Koltchinskii, V. (2011). *Oracle inequalities in empirical risk minimization and sparse recovery problems.* Springer Lecture Notes in Mathematics.

[57] Kong, E., Linton, O., and Xia, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory* **26** 1529–1564.

[58] Leadbetter, M., Lindgren, G., and Rootzen, H. (1983). *Extremes and related properties of random sequences and processes.* Springer.

[59] Lee, S. (2003). Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory* **19** 1–31.

[60] Lewbel, A. (1987). Characterizing some gorman engel curves. *Econometrica* **55** 1451–1459.

[61] Linton, O., Song, K., and Whang, Y.-J. (2010). An improved bootstrap test of stochastic dominance. *J. Econometrics* **154** 186–202.

[62] Lipsitz, M., Belloni, A., Chernozhukov, V., and I. Fernandez-Val (2016). Quantreg.nonpar: Nonparametric Series Quantile Regression in R. *Unpublished manuscript*.

[63] Lovász, L. and Vempala, S. (2006). The geometry of logconcave functions and sampling algorithms. *Random Structures and Algorithms* **30** 307–358.

[64] Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Ann. Statist.* **21** 255–285.

[65] Massart, P. (2007). *Concentration inequalities and model selection.* Springer Lecture Notes in Mathematics.

[66] Matzkin, R. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica* **71** 1339–1375.

[67] Newey, W. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics* **79** 147–168.

[68] Parzen, M., Wei, L., and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81** 341–350.

[69] Pollard, D. (2001). *A User's Guide to Measure Theoretic Probability.* Cambridge Series in Statistical and Probabilistic Mathematics.

[70] Portnoy, S. and Koenker, R. (1997). The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statist. Sci.* **12** 279–300. With comments by Ronald A. Thisted and M. R. Osborne and a rejoinder by the authors.

[71] Powell, J. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics* **25** 303–325.

[72] Præstgaard, J. and Wellner, J. (1993). Exchangeably weighted bootstraps of the general empirical process. *Ann. Probab.* **21** 2053–2086.

[73] Qu, Z. and Yoon, J. (2011). Nonparametric estimation and inference on conditional quantile processes. *Unpublished manuscript.*

[74] Rio, E. (1994). Local invariance principles and their application to density estimation. *Probability Theory and Related Fields* **98** 21–45.

[75] R Development Core Team (2008). *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

[76] Schmalensee, R. and Stoker, T. (1999). Household gasoline demand in the united states. *Econometrica* **67** 645–662.

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

[78] Stute, W. (1986). Conditional empirical processes. *The Annals of Statistics* **14** 638–647.

[79] van de Geer, S. (2002). M-estimation using penalties or sieves. *Journal of Statistical Planning and Inference* **108** 55–69.

[80] van der Vaart, A. and Wellner, J. (1996). *Weak Convergence and Empirical Processes.* Springer Series in Statistics.

[81] Vitale, R. (2000). Some comparisons for gaussian processes. *Proceedings of the American Math. Society* **128** 3043–3046.

[82] White, H. (1992). Nonparametric estimation of conditional quantiles using neural networks. *Proceedings of the Symposium on the Interface* 190–199.

[83] Yatchew, A. and No, J. (2001). Household gasoline demand in Canada. *Econometrica* **69** 1697–1709.