CONDITIONAL QUANTILE PROCESSES BASED ON SERIES OR MANY REGRESSORS

ALEXANDRE BELLONI, VICTOR CHERNOZHUKOV 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 empirical QR coefficient process, namely we obtain uniform strong approximations to the empirical QR coefficient process by conditionally pivotal and Gaussian processes, as well as by gradient and weighted bootstrap processes.

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, large sample distributions, and inference methods based on strong pivotal and Gaussian approximations and on gradient and weighted bootstraps. All of the above results are for function-valued parameters, holding uniformly in both the quantile index and in the covariate value, and covering the pointwise case as a by-product. If the function of interest is monotone, we show how to use monotonization procedures to improve estimation and inference. We demonstrate the practical utility of these results with an empirical example, where we estimate the price elasticity function of the individual demand for gasoline, as indexed by the individual unobserved propensity for gasoline consumption.

1. Introduction

Quantile regression (QR) is a principal regression method for analyzing the impact of covariates on outcomes, particularly when the impact might be heterogeneous. This impact is characterized by the conditional quantile function and its functionals [3, 7, 24]. 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 U for consuming the good, as

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

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 $Q(X, u)$ is the u-th conditional quantile of Y given X, i.e. $Q(X, u) = Q_{Y|X}(u|X)$. 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 and its linear functionals. In this framework, we approximate the entire conditional quantile function $Q_{Y|X}(u|x)$ 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, or B-splines. 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 [25]. As the number of series terms grows, the approximation error $Q_{Y|X}(u|x) - Z(x)'\beta(u)$ 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 $\widehat{\beta}(\cdot)$ denote the QR estimator of $\beta(\cdot)$. The first set of results provides large-sample theory for the empirical QR coefficient process of increasing dimension $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$. We obtain uniform strong approximations to this process by a sequence of the following stochastic processes of increasing dimension:

(i) a conditionally pivotal process,
(ii) a gradient bootstrap process,

(iii) a Gaussian process, and

(iv) a weighted bootstrap process.

To the best of our knowledge, all of the above results are new. The existence of the pivotal approximation emerges from the special nature of QR, where a (sub) gradient of the sample objective function evaluated at the truth is pivotal conditional on the regressors. This allows us to perform high-quality inference without even resorting to Gaussian approximations. We also show that the gradient bootstrap, introduced by Parzen, Wei and Ying [30] in the parametric context, is effectively a means of carrying out the conditionally pivotal approximation without explicitly estimating Jacobian matrices. The conditions for validity of these two schemes require only a mild restriction on the growth of the number of series terms in relation to the sample size. We also obtain a Gaussian approximation to the entire distribution of QR process of increasing dimension by using chaining arguments and Yurinskii's coupling. Moreover, we show that the weighted bootstrap works to approximate the distribution of QR process for the same reason as the Gaussian approximation. The conditions for validity of the Gaussian and weighted bootstrap approximations, however, appear to be substantively stronger than for the pivotal and gradient bootstrap approximations.

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

(i) the conditional quantile function itself, $(u, x) \mapsto Q_{Y|X}(u|x)$,

(ii) the partial derivative function, $(u, x) \mapsto \partial_{x_k} Q_{Y|X}(u|x)$,

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

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

where μ is a given measure and x_k is the k-th component of x. Specifically, we derive uniform rates of convergence, large sample distributions and inference methods based on the strong pivotal and Gaussian approximations and on the gradient and weighted bootstraps. It is noteworthy that all of the above results apply to function-valued parameters, holding uniformly in both the quantile index and the covariate value, and covering pointwise normality and rate results as a special case. If the function of interest is monotone, we show how to use monotonization procedures to improve estimation and inference.

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 [22] that studied the many regressors model and gave pointwise limit theorems for the QR estimator in the case of a single quantile index. 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 [28] that studied QR estimation of partially linear models in the series framework for a single quantile index, and on Horowitz and Lee [23] that studied nonparametric QR estimation of additive quantile models for a single quantile index 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. Other very important work includes Chaudhuri [10], and Chaudhuri, Doksum and Samarov [11], Härdle, Ritov, and Song [20], Cattaneo, Crump, and Jansson [8], and Kong, Linton, and Xia [27], among others, but this work 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 [35], Andrews [2], Newey [29], Chen and Shen [13], Chen [12] and others that rigorously motivated the series framework as an approximation scheme and gave pointwise normality results for least squares estimators, and on Chen [12] and van de Geer [36] that gave (non-uniform) consistency and rate results for general series estimators, including quantile regression for the case of a single quantile index. White [38] 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 entire quantile regression process.

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. One of our inference approaches involves an approximation to the entire conditional quantile process by a conditionally pivotal process, which is not Donsker in general, but can be used for high-quality inference. The utility of this new technique is particularly apparent in our high-dimensional setting. Under stronger conditions, we also establish an asymptotically valid approximation to the quantile regression process by Gaussian processes using Yurinskii's coupling. Previously, [15] used Yurinskii's coupling to obtain a strong

approximation to the least squares series estimator. The use of this technique in our context is new and much more involved, because we approximate an entire empirical QR 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, do not even have analogs in the least squares series literature (the extension of our results to least squares is a subject of ongoing research, [5]).

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. [6] 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.

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 series QR process. In Section 4, we give estimation and inference theory for linear functionals of the conditional quantile function and show how to improve estimation and inference by imposing monotonicity restrictions. 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 and the proofs of the main results are collected in the Appendices.

Notation. In what follows, S^{m-1} denotes the unit sphere in $\mathbb{R}^m$. For $x \in \mathbb{R}^m$, we define the Euclidian norm as $\|x\| := \sup_{\alpha \in S^{m-1}} |\alpha' x|$. For a set I, $\operatorname{diam}(I) = \sup_{v, \bar{v} \in I} \|v - \bar{v}\|$ denotes the diameter of I, and $\operatorname{int}(I)$ denotes the interior of I. For any two real numbers a and b, $a \vee b = \max\{a, b\}$ and $a \wedge b = \min\{a, b\}$. Calligraphic letters are used to denote the support of interest of a random variable or vector. For example, $\mathcal{U} \subset (0, 1)$ is the support of U, $\mathcal{X} \subset \mathbb{R}^d$ is the support of X, and $\mathcal{Z} = \{Z(x) \in \mathbb{R}^m : x \in \mathcal{X}\}$ is the support of $Z = Z(X)$. The relation $a_n \lesssim b_n$ means that $a_n \leq C b_n$ for a constant C and for all n large enough. We denote by P^* the probability measure induced by conditioning on any realization of the data $\mathcal{D}_n := \{(Y_i, X_i) : 1 \leq i \leq n\}$. We say that a random variable $\Delta_n = o_{P^*}(1)$ in P-probability if for any positive numbers $\epsilon > 0$ and $\eta > 0$, $P\{P^*(|\Delta_n| > \epsilon) > \eta\} = o(1)$, or

equivalently, $P^*(|\Delta_n| > \epsilon) = o_P(1)$. 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 $\sqrt{n}(\mathbb{E}_n - E)$.

2. SERIES FRAMEWORK

2.1. **The set-up.** The set-up corresponds to the nonparametric QR series framework:

$$Y = Q_{Y|X}(U|X) = Z'\beta(U) + R(X,U), \quad U|X \sim \text{Uniform}(0,1) \text{ and } \beta(u) \in \mathbb{R}^m,$$

where X is a d-vector of elementary regressors, and $Z = Z(X)$ is an m-vector of approximating functions formed by transformations of X with the first component of Z equal to one. The function $Z'\beta(u)$ is the series approximation to $Q_{Y|X}(u|X)$ and is linear in the parameter vector $\beta(u)$, which is defined below. The term $R(X, u)$ is the approximation error, with the special case of $R(X, u) = 0$ corresponding to the many regressors model. We allow the series terms $Z(X)$ to change with the sample size n, i.e. $Z(X) = Z_n(X)$, but we shall omit the explicit index indicating the dependence on n. We refer the reader to Newey [29] and Chen [12] for examples of series functions, including (i) regression splines, (ii) polynomials, (iii) trigonometric series, (iv) compactly supported wavelets, and others. Interestingly, in the latter scheme, the entire collection of series terms is dependent upon the sample size.

For each quantile index $u \in (0, 1)$, the population coefficient $\beta(u)$ is defined as a solution to

$$\min_{\beta \in \mathbb{R}^m} E[\rho_u(Y - Z'\beta)] \tag{2.1}$$

where $\rho_u(z) = (u - 1\{z < 0\})z$ is the check function ([24]). Thus, this coefficient is not a solution to an approximation problem but to a prediction problem. However, we show in Lemma 1 in Appendix B that the solution to (2.1) inherits important approximation properties from the best least squares approximation of $Q_{Y|X}(u|\cdot)$ by a linear combination of $Z(\cdot)$.

We consider estimation of the coefficient function $u \mapsto \beta(u)$ using the entire quantile regression process over $\mathcal{U}$, a compact subset of $(0, 1)$,

$$\{u \mapsto \widehat{\beta}(u), u \in \mathcal{U}\},$$

namely, for each $u \in \mathcal{U}$, the estimator $\widehat{\beta}(u)$ is a solution to the empirical analog of the population problem (2.1)

$$\min_{\beta \in \mathbb{R}^m} \mathbb{E}_n[\rho_u(Y_i - Z_i'\beta)]. \tag{2.2}$$

We are also interested in various functionals of $Q_{Y|X}(\cdot|x)$. We estimate them by the corresponding functionals of $Z(x)'\widehat{\beta}(\cdot)$, such as the estimator of the entire conditional quantile function, $(x, u) \mapsto Z(x)'\widehat{\beta}(u)$, and derivatives of this function with respect to the elementary covariates x.

2.2. **Regularity Conditions.** In our framework the entire model can change with n, although we shall omit the explicit indexing by n. We will use the following primitive assumptions on the data generating process, as $n \to \infty$ and $m = m(n) \to \infty$.

   Condition S.

   S.1 *The data $\mathcal{D}_n = \{(Y_i, X_i')', 1 \le i \le n\}$ are an i.i.d. sequence of real $(1+d)$-vectors, and $Z_i = Z(X_i)$ is a real m-vector for $i = 1, \ldots, n$.*

   S.2 *The conditional density of the response variable $f_{Y|X}(y|x)$ is bounded above by $\overline{f}$ and its derivative in y is bounded above by $\overline{f'}$, uniformly in the arguments y and $x \in \mathcal{X}$ and in n; moreover, $f_{Y|X}(Q_{Y|X}(u|x)|x)$ is bounded away from zero uniformly for all arguments $u \in \mathcal{U}$, $x \in \mathcal{X}$, and n.*

   S.3 *For every m, the eigenvalues of the Gram matrix $\Sigma_m = E[ZZ']$ are bounded from above and away from zero, uniformly in n.*

   S.4 *The norm of the series terms obeys $\max_{i \le n} \|Z_i\| \le \zeta(m, d, n) := \zeta_m$.*

   S.5 *The approximation error term $R(X, U)$ is such that $\sup_{x \in \mathcal{X}, u \in \mathcal{U}} |R(x, u)| \lesssim m^{-\kappa}$.*

Comment 2.1. Condition S is a simple set of sufficient conditions. Our proofs and lemmas gathered in the appendices hold under more general, but less primitive, conditions. Condition S.2 impose mild smoothness assumptions on the conditional density function. Condition S.3 and S.4 imposes plausible conditions on the design, see Newey [29]. Condition S.2 and S.3 together imply that the eigenvalues of the Jacobian matrix

$$J_m(u) = E[f_{Y|X}(Q_{Y|X}(u|X)|X)ZZ']$$

are bounded away from zero and from above; Lemma 12 further shows that this together with S.5 implies that the eigenvalues of $\widetilde{J}_m(u) = E[f_{Y|X}(Z'\beta(u)|X)ZZ']$ are bounded away from zero and from above, since the two matrices are uniformly close. Assumption S.4 imposes a uniform bound on the norm of the transformed vector which can grow with

the sample size. For example, we have $\zeta_m = \sqrt{m}$ for splines and $\zeta_m = m$ for polynomials. Assumption S.5 introduces a bound on the approximation error and is the only nonprimitive assumption. Deriving sharp bounds on the sup norm for the approximation error is a delicate subject of approximation theory even for least squares, with the exception of local polynomials; see, e.g. [5] for a discussion in the least squares case. However, in Lemma 1 in Appendix B, we characterize the nature of the approximation by establishing that the vector $\beta(u)$ solving (2.1) is equivalent to the least squares approximation to $Q_{Y|X}(u|x)$ in terms of the order of L_2 error. We also deduce a simple upper bound on the sup norm for the approximation error. In applications, we recommend to analyze the size of the approximation error numerically. This is possible by considering examples of plausible functions $Q_{Y|X}(u|x)$ and comparing them to the best linear approximation schemes, as this directly leads to sharp results. We give an example of how to implement this approach in the empirical application.

3. Asymptotic Theory for QR Coefficient Processes based on Series

3.1. **Uniform Convergence Rates for Series QR Coefficients.** The first main result is a uniform rate of convergence for the series estimators of the QR coefficients.

Theorem 1 (Uniform Convergence Rate for Series QR Coefficients). *Under Condition S, and provided that $\zeta_m^2 m \log n = o(n)$,*

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

Thus, up to a logarithmic factor, the uniform convergence over $\mathcal{U}$ is achieved at the same rate as for a single quantile index ([22]). The proof of this theorem relies on new concentration inequalities that control the behavior of the empirical eigenvalues of the design matrix.

3.2. **Uniform Strong Approximations to the Series QR Process.** The second main result is the approximation of the QR process by a linear, conditionally pivotal process.

Theorem 2 (Strong Approximations to the QR Process by a Pivotal Coupling). *Under Condition S, $m^3 \zeta_m^2 \log^7 n = o(n)$, and $m^{-\kappa+1} \log^3 n = o(1)$, the QR process is uniformly close to a conditionally pivotal process, namely*

$$\sqrt{n} \left(\widehat{\beta}(u) - \beta(u) \right) = J_m^{-1}(u) \mathbb{U}_n(u) + r_n(u),$$

where

$$\mathbb{U}_n(u) := \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(u - 1\{U_i \leq u\}), \tag{3.3}$$

where $U_1, \ldots, U_n$ are i.i.d. Uniform$(0,1)$, independently distributed of $Z_1, \ldots, Z_n$, and

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

The theorem establishes that the QR process is approximately equal to the process $J_m^{-1}(\cdot)\mathbb{U}_n(\cdot)$, which is pivotal conditional on $Z_1, ..., Z_n$. This is quite useful since we can simulate $\mathbb{U}_n(\cdot)$ in a straightforward fashion, conditional on $Z_1, ..., Z_n$, and we can estimate the matrices $J_m(\cdot)$ using Powell's method [32]. Thus, we can perform inference based on the entire empirical QR process in series regression or many regressors contexts. The following theorem establishes the formal validity of this approach, which we call the pivotal method.

Theorem 3 (Pivotal Method). *Let $\widehat{J}_m(u)$ denote the estimator of $J_m(u)$ defined in (3.7) with bandwidth h_n obeying $h_n = o(1)$ and $h_n\sqrt{m} \log^{3/2} n = o(1)$. Under Condition S, $m^{-\kappa+1/2} \log^{3/2} n = o(1)$, and $\zeta_m^2 m^2 \log^4 n = o(nh_n)$, the feasible pivotal process $\widehat{J}_m^{-1}(\cdot)\mathbb{U}_n^*(\cdot)$ correctly approximates a copy $J_m^{-1}(\cdot)\mathbb{U}_n^*(\cdot)$ of the pivotal process defined in Theorem 2:*

$$\widehat{J}_m^{-1}(u)\mathbb{U}_n^*(u) = J_m^{-1}(u)\mathbb{U}_n^*(u) + r_n(u),$$

where

$$\mathbb{U}_n^*(u) := \frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i(u - 1\{U_i^* \leq u\}), \tag{3.4}$$

$U_1^*, \ldots, U_n^*$ *are i.i.d. Uniform$(0,1)$, independently distributed of $Z_1, \ldots, Z_n$, and*

$$\sup_{u \in \mathcal{U}} \|r_n(u)\| \lesssim_P \sqrt{\frac{\zeta_m^2 m^2 \log^2 n}{nh_n}} + m^{-\kappa+1/2}\sqrt{\log n} + h_n\sqrt{m \log n} = o(1/\log n).$$

This method is closely related to another approach to inference, which we refer to here as the gradient bootstrap method. This approach was previously introduced by Parzen, Wei and Ying [30] for parametric models with fixed dimension. We extend it to a considerably more general nonparametric series framework. The main idea is to generate draws $\widehat{\beta}^*(\cdot)$ of the QR process as solutions to QR problems with gradients perturbed by a pivotal quantity $\mathbb{U}_n^*(\cdot)/\sqrt{n}$. In particular, let us define a gradient bootstrap draw $\widehat{\beta}^*(u)$ as the solution to the problem

$$\min_{\beta \in \mathbb{R}^m} \mathbb{E}_n[\rho_u(Y_i - Z_i'\beta)] - \mathbb{U}_n^*(u)'\beta/\sqrt{n}, \tag{3.5}$$

for each $u \in \mathcal{U}$, where $\mathbb{U}_n^*(\cdot)$ is defined in (3.4). The problem is solved many times for independent draws of $\mathbb{U}_n^*(\cdot)$, and the distribution of $\sqrt{n}(\widehat{\beta}(\cdot) - \beta(\cdot))$ is approximated by the empirical distribution of the bootstrap draws of $\sqrt{n}(\widehat{\beta}^*(\cdot) - \widehat{\beta}(\cdot))$.

Theorem 4 (Gradient Bootstrap Method). *Under Condition S, $\zeta_m^2 m^3 \log^7 n = o(n)$, and $m^{-\kappa+1/2} \log^{3/2} n = o(1)$, the gradient bootstrap process correctly approximates a copy $J_m^{-1}(\cdot)\mathbb{U}_n^*(\cdot)$ of the pivotal process defined in Theorem 2,*

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

where $\mathbb{U}_n^*(u)$ is defined in (3.4) and

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

The stated bound continues to hold in P-probability if we replace the unconditional probability P by the conditional probability P^*.

The main advantage of this method relative to the pivotal method is that it does not require to estimate $J_m(\cdot)$ explicitly. The disadvantage is that in large problems the gradient bootstrap can be computationally much more expensive.

Next we turn to a strong approximation based on a sequence of Gaussian processes.

Theorem 5 (Strong Approximation to the QR Process by a Gaussian Coupling). *Under conditions S.1-S.4 and $m^7\zeta_m^6 \log^{22} n = o(n)$, there exists a sequence of zero-mean Gaussian processes $G_n(\cdot)$ with a.s. continuous paths, that has the same covariance functions as the pivotal process $\mathbb{U}_n(\cdot)$ in (3.4) conditional on $Z_1, \ldots, Z_n$, namely,*

$$E[G_n(u)G_n(u')'] = E[\mathbb{U}_n(u)\mathbb{U}_n(u')'] = \mathbb{E}_n[Z_iZ_i'](u \wedge u' - uu'), \text{ for all } u \text{ and } u' \in \mathcal{U}.$$

Also, $G_n(\cdot)$ approximates the empirical process $\mathbb{U}_n(\cdot)$, namely,

$$\sup_{u\in\mathcal{U}} \|\mathbb{U}_n(u) - G_n(u)\| \lesssim_P o(1/\log n).$$

Consequently, if in addition S.5 holds with $m^{-\kappa+1} \log^3 n = o(1)$,

$$\sup_{u\in\mathcal{U}} \|\sqrt{n}(\widehat{\beta}(u) - \beta(u)) - J_m^{-1}(u)G_n(u)\| \lesssim_P o(1/\log n).$$

Moreover, under the conditions of Theorem 3, the feasible Gaussian process $\widehat{J}_m^{-1}(\cdot)G_n(\cdot)$ correctly approximates $J_m^{-1}(\cdot)G_n(\cdot)$:

$$\widehat{J}_m^{-1}(u)G_n(u) = J_m^{-1}(u)G_n(u) + r_n(u),$$

where

$$\sup_{u \in \mathcal{U}} \|r_n(u)\| \lesssim_P \sqrt{\frac{\zeta_m^2 m^2 \log^2 n}{n h_n}} + m^{-\kappa+1/2}\sqrt{\log n} + h_n\sqrt{m \log n} = o(1/\log n).$$

The requirement on the growth of the dimension m in Theorem 5 is a consequence of a step in the proof that relies upon Yurinskii's coupling. Therefore, improving that step through the use of another coupling could lead to significant improvements in such requirements. Note, however, that the pivotal approximation has a much weaker growth restriction, and it is also clear that this approximation should be more accurate than any further approximation of the pivotal approximation by a Gaussian process.

Another related inference method is the weighted bootstrap for the entire QR process. Consider a set of weights $h_1, ..., h_n$ that are i.i.d. draws from the standard exponential distribution. For each draw of such weights, define the weighted bootstrap draw of the QR process as a solution to the QR problem weighted by $h_1, \ldots, h_n$:

$$\widehat{\beta}^b(u) \in \arg\min_{\beta \in \mathbb{R}^m} \mathbb{E}_n[h_i \rho_u(Y_i - Z_i'\beta)], \quad \text{for } u \in \mathcal{U}.$$

The following theorem establishes that the weighted bootstrap distribution is valid for approximating the distribution of the QR process.

Theorem 6 (Weighted Bootstrap Method). *(1) Under Condition S, $\zeta_m^2 m^3 \log^9 n = o(n)$, and $m^{-\kappa+1} \log^3 n = o(1)$, the weighted bootstrap process satisfies*

$$\sqrt{n}\left(\widehat{\beta}^b(u) - \widehat{\beta}(u)\right) = \frac{J_m^{-1}(u)}{\sqrt{n}} \sum_{i=1}^{n} (h_i - 1)Z_i(u - 1\{U_i \le u\}) + r_n(u),$$

where $U_1, \ldots, U_n$ are i.i.d. Uniform$(0,1)$, independently distributed of $Z_1, \ldots, Z_n$, and

$$\sup_{u \in \mathcal{U}} \|r_n(u)\| \lesssim_P \frac{m^{3/4}\zeta_m^{1/2}\log^{5/4} n}{n^{1/4}} + \sqrt{m^{1-\kappa}\log n} = o(1/\log n).$$

The bound continues to hold in P-probability if we replace the unconditional probability P by P^*.

(2) Furthermore, under the conditions of Theorem 5, the weighted bootstrap process approximates the Gaussian process $J_m^{-1}(\cdot)G_n(\cdot)$ defined in Theorem 5, that is:

$$\sup_{u \in \mathcal{U}} \|\sqrt{n}(\widehat{\beta}^b(u) - \widehat{\beta}(u)) - J_m^{-1}(u)G_n(u)\| \lesssim_P o(1/\log n).$$

In comparison with the pivotal and gradient bootstrap methods, the Gaussian and weighted bootstrap methods require stronger assumptions.

3.3. **Estimation of Matrices.** In order to implement some of the inference methods, we need uniformly consistent estimators of the Gram and Jacobian matrices. The natural candidates are

$$\widehat{\Sigma}_m = \mathbb{E}_n[Z_i Z_i'], \tag{3.6}$$

$$\widehat{J}_m(u) = \frac{1}{2h_n}\mathbb{E}_n[1\{|Y_i - Z_i'\widehat{\beta}(u)| \le h_n\} \cdot Z_i Z_i'], \tag{3.7}$$

where h_n is a bandwidth parameter, such that $h_n \to 0$, and $u \in \mathcal{U}$. The following result establishes uniform consistency of these estimators and provides an appropriate rate for the bandwidth h_n which depends on the growth of the model.

Theorem 7 (Estimation of Gram and Jacobian Matrices). *If conditions S.1-S.4 and $\zeta_m^2 \log n = o(n)$ hold, then $\widehat{\Sigma}_m - \Sigma_m = o_P(1)$ in the eigenvalue norm . If conditions S.1-S.5, $h_n = o(1)$, and $m\zeta_m^2 \log n = o(nh_n)$ hold, then $\widehat{J}_m(u) - J_m(u) = o_P(1)$ in the eigenvalue norm uniformly in $u \in \mathcal{U}$.*

4. Estimation and Inference Methods for Linear Functionals

In this section we apply the previous results to develop estimation and inference methods for various linear functionals of the conditional quantile function.

4.1. **Functionals of Interest.** We are interested in various functions θ created as linear functionals of $Q_{Y|X}(u|\cdot)$ for all $u \in \mathcal{U}$. Particularly useful examples include, for $x = (w, v)$ and x_k denoting the k-th component of x,

1. the function itself: $\theta(u, x) = Q_{Y|X}(u|x)$;
2. the derivative: $\theta(u, x) = \partial_{x_k} Q_{Y|X}(u|x)$;
3. the average derivative: $\theta(u) = \int \partial_{x_k} Q_{Y|X}(u|x)d\mu(x)$;
4. the conditional average derivative: $\theta(u|w) = \int \partial_{x_k} Q_{Y|X}(u|w, v)d\mu(v|w)$.

The measures μ entering the definitions above are taken as known; the result can be extended to include estimated measures.

Note that in each example we could be interested in estimating $\theta(u, w)$ simultaneously for many values of (u, w). For $w \in \mathcal{W} \subset \mathbb{R}^d$, let $I \subset \mathcal{U} \times \mathcal{W}$ denote the set of indices (u, w) of interest. For example:

- the function at a particular point (u, w), in which case $I = \{(u, w)\}$,
- the function $u \mapsto \theta(u, w)$, having fixed w, in which case $I = \mathcal{U} \times \{w\}$,
- the function $w \mapsto \theta(u, w)$, having fixed u, in which case $I = \{u\} \times \mathcal{W}$,

- the entire function $(u, w) \mapsto \theta(u, w)$, in which case $I = \mathcal{U} \times \mathcal{W}$.

By the linearity of the series approximations, the above parameters can be seen as linear functions of the quantile regression coefficients $\beta(u)$ up to an approximation error, that is

$$\theta(u, w) = \ell(w)'\beta(u) + r_n(u, w), \quad (u, w) \in I, \tag{4.8}$$

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

$$\left(\mathcal{A}Q_{Y|X}(u|\cdot)\right)[w] = \left(\mathcal{A}Z(\cdot)\right)[w]'\beta(u) + \left(\mathcal{A}R(u, \cdot)\right)[w]. \tag{4.9}$$

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

1. the identity operator: $(\mathcal{A}f)[x] = f[x]$, so that

$$\ell(x) = Z(x), \quad r_n(u, x) = R(u, x) \ ;$$

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

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

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

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

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

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

For notational convenience, we use the formulation (4.8) in the analysis, instead of the motivational formulation (4.9).

We shall provide the inference tools that will be valid for inference on the series approximation

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

and, provided that the approximation error $r_n(u, w)$, $(u, w) \in I$, is small enough as compared to the estimation noise, these tools will also be valid for inference on the functional of interest:

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

Therefore, the series approximation is an important penultimate target, whereas the functional θ is the ultimate target. The inference will be based on the plug-in estimator $\widehat{\theta}(u, w) := \ell(w)'\widehat{\beta}(u)$ of the the series approximation $\ell(w)'\beta(u)$ and hence of the final target $\theta(u, w)$.

4.2. **Pointwise Rates and Inference on the Linear Functionals.** Let (u, w) be a pair of a quantile index value u and a covariate value w, at which we are interested in performing inference on $\theta(u, w)$. In principle, this deterministic value can be implicitly indexed by n, although we suppress the dependence.

Condition P.

P.1 *The approximation error is small, namely $\sqrt{n}|r_n(u, w)|/\|\ell(w)\| = o(1)$.*
P.2 *The norm of the loading $\ell(w)$ satisfies: $\|\ell(w)\| \lesssim \xi_\theta(m, w)$.*

Theorem 8 (Pointwise Convergence Rate for Linear Functionals)**.** *Assume that the conditions of Theorem 2 and Condition P hold, then*

$$|\widehat{\theta}(u, w) - \theta(u, w)| \lesssim_P \frac{\xi_\theta(m, w)}{\sqrt{n}}. \tag{4.10}$$

In order to perform inference, we construct an estimator of the variance as

$$\widehat{\sigma}_n^2(u, w) = u(1 - u)\ell(w)'\widehat{J}_m^{-1}(u)\widehat{\Sigma}_m\widehat{J}_m^{-1}(u)\ell(w)/n. \tag{4.11}$$

Under the stated conditions this quantity is consistent for

$$\sigma_n^2(u, w) = u(1 - u)\ell(w)'J_m^{-1}(u)\Sigma_m J_m^{-1}(u)\ell(w)/n. \tag{4.12}$$

Finally, consider the t-statistic:

$$t_n(u, w) = \frac{\widehat{\theta}(u, w) - \theta(u, w)}{\widehat{\sigma}_n(u, w)}.$$

Condition P also ensures that the approximation error is small, so that

$$t_n(u, w) = \frac{\ell(w)'(\widehat{\beta}(u) - \beta(u))}{\widehat{\sigma}_n(u, w)} + o_P(1).$$

We can carry out standard inference based on this statistic because $t_n(u, w) \to_d N(0, 1)$.

Moreover, we can base inference on the quantiles of the following statistics:

$$\text{pivotal coupling} \qquad t_n^*(u, w) = \frac{\ell(w)' \widehat{J}_m^{-1}(u) \mathbb{U}_n^*(u) / \sqrt{n}}{\widehat{\sigma}_n(u, w)};$$

$$\text{gradient bootstrap coupling} \quad t_n^*(u, w) = \frac{\ell(w)'(\widehat{\beta}^*(u) - \widehat{\beta}(u))}{\widehat{\sigma}_n(u, w)}; \qquad (4.13)$$

$$\text{weighted bootstrap coupling} \quad t_n^*(u, w) = \frac{\ell(w)'(\widehat{\beta}^b(u) - \widehat{\beta}(u))}{\widehat{\sigma}_n(u, w)}.$$

Accordingly let $k_n(1 - \alpha)$ be the $1 - \alpha/2$ quantile of the standard normal variable, or let $k_n(1 - \alpha)$ denote the $1 - \alpha$ quantile of the random variable $|t_n^*(u, w)|$ conditional on the data, i.e. $k_n(1 - \alpha) = \inf\{t : P(|t_n^*(u, w)| \le t | \mathcal{D}_n) \ge 1 - \alpha\}$, then a pointwise $(1 - \alpha)$-confidence interval can be formed as

$$[\grave{\iota}(u, w), \ddot{\iota}(u, w)] = [\widehat{\theta}(u, w) - k_n(1 - \alpha)\widehat{\sigma}_n(u, w), \ \widehat{\theta}(u, w) + k_n(1 - \alpha)\widehat{\sigma}_n(u, w)].$$

Theorem 9 (Pointwise Inference for Linear Functionals). *Suppose that the conditions of Theorems 2 and 7, and Condition P hold. Then,*

$$t_n(u, w) \to_d N(0, 1).$$

For the case of using the gradient bootstrap method, suppose that the conditions of Theorem 4 hold. For the case of using the weighted bootstrap method, suppose that the conditions of Theorem 6 hold. Then, $t_n^*(u, w) \to_d N(0, 1)$ conditional on the data.

Moreover, the $(1 - \alpha)$-confidence band $[\grave{\iota}(u, w), \ddot{\iota}(u, w)]$ covers $\theta(u, w)$ with probability that is asymptotically no less than $1 - \alpha$, namely

$$P\Big\{\theta(u, w) \in [\grave{\iota}(u, w), \ddot{\iota}(u, w)]\Big\} \ge 1 - \alpha + o(1).$$

4.3. **Uniform Rates and Inference on the Linear Functionals.** For $f : I \mapsto \mathbb{R}^k$, define the norm

$$\|f\|_I := \sup_{(u,w) \in I} \|f(u, w)\|.$$

We shall invoke the following assumptions to establish rates and uniform inference results over the region I.

Condition U.

U.1 *The approximation error is small, namely* $\sqrt{n} \log n \sup_{(u,w) \in I} \||r_n(u, w)/\|\ell(w)\|\|| = o(1)$.

U.2 *The loadings $\ell(w)$ are uniformly bounded and admit Lipschitz coefficients $\xi_\theta^L(m, I)$, that is,*

$$\|\ell\|_I \lesssim \xi_\theta(m, I), \ \|\ell(w) - \ell(w')\| \leq \xi_\theta^L(m, I)\|w - w'\|, \quad and$$

$$\log[\mathrm{diam}(I) \vee \xi_\theta(m, I) \vee \xi_\theta^L(m, I) \vee \zeta_m] \lesssim \log m$$

Theorem 10 (Uniform Convergence Rate for Linear Functionals). *Assume that the conditions of Theorem 2 and Condition U hold, and $d\xi_\theta^2(m, I)\zeta_m^2 \log^2 n = o(n)$, then*

$$\sup_{(w,u)\in I} |\widehat{\theta}(u, w) - \theta(u, w)| \lesssim_P \frac{\xi_\theta(m, I)\vee 1}{\sqrt{n}} \log n. \tag{4.14}$$

As in the pointwise case, consider the estimator of the variance $\widehat{\sigma}_n^2(u, w)$ in (4.11). Under our conditions this quantity is uniformly consistent for $\sigma_n^2(u, w)$ defined in (4.12), namely $\widehat{\sigma}_n^2(u, w)/\sigma_n^2(u, w) = 1 + o_P(1/\log n)$ uniformly over $(u, w) \in I$. Then we consider the t-statistic process:

$$\left\{ t_n(u, w) = \frac{\widehat{\theta}(u, w) - \theta(u, w)}{\widehat{\sigma}_n(u, w)}, \quad (u, w) \in I \right\}.$$

Under our assumptions the approximation error is small, so that

$$t_n(u, w) = \frac{\ell(w)'(\widehat{\beta}(u) - \beta(u))}{\widehat{\sigma}_n(u, w)} + o_P(1/\log n) \text{ in } \ell^\infty(I).$$

The main result on inference is that the t-statistic process can be strongly approximated by the following pivotal processes or couplings:

pivotal coupling $\quad \left\{ t_n^*(u, w) = \frac{\ell(w)'\widehat{J}_m^{-1}(u)\mathbb{U}_n^*(u)/\sqrt{n}}{\widehat{\sigma}_n(u, w)}, \quad (u, w) \in I \right\};$

gradient bootstrap coupling $\quad \left\{ t_n^*(u, w) = \frac{\ell(w)'(\widehat{\beta}^*(u) - \widehat{\beta}(u))}{\widehat{\sigma}_n(u, w)}, \quad (u, w) \in I \right\};$

Gaussian coupling $\quad \left\{ t_n^*(u, w) = \frac{\ell(w)'\widehat{J}_m^{-1}(u)G_n(u)/\sqrt{n}}{\widehat{\sigma}_n(u, w)}, \quad (u, w) \in I \right\};$

weighted bootstrap coupling $\quad \left\{ t_n^*(u, w) = \frac{\ell(w)'(\widehat{\beta}^b(u) - \widehat{\beta}(u))}{\widehat{\sigma}_n(u, w)}, \quad (u, w) \in I \right\}.$

$$\tag{4.15}$$

The following theorem shows that these couplings approximate the distribution of the t-statistic process in large samples.

Theorem 11 (Strong Approximation of Inferential Processes by Couplings). *Suppose that the conditions of Theorems 2 and 3, and Condition U hold. For the case of using the gradient bootstrap method, suppose that the conditions of Theorem 4 hold. For the case of using the Gaussian approximation, suppose that the conditions of Theorem 5 hold. For the case of using the weighted bootstrap method, suppose that the conditions of Theorem 6 hold. Then,*

$$t_n(u, w) =_d t_n^*(u, w) + o_P(1/\log n), \ \ in \ \ell^\infty(I),$$

where P can be replaced by P^*.

To construct uniform two-sided confidence bands for $\{\theta(u, w) : (u, w) \in I\}$, we consider the maximal t-statistic

$$\|t_n\|_I = \sup_{(u,w) \in I} |t_n(u, w)|,$$

as well as the couplings to this statistic in the form:

$$\|t_n^*\|_I = \sup_{(u,w) \in I} |t_n^*(u, w)|.$$

Ideally, we would like to use quantiles of the first statistic as critical values, but we do not know them. We instead use quantiles of the second statistic as large sample approximations. Let $k_n(1 - \alpha)$ denote the $1 - \alpha$ quantile of random variable $\|t_n^*\|_I$ conditional on the data, i.e.

$$k_n(1 - \alpha) = \inf\{t : P(\|t_n^*\|_I \leq t | \mathcal{D}_n) \geq 1 - \alpha\}.$$

This quantity can be computed numerically by Monte Carlo methods, as we illustrate in the empirical section.

Let $\delta_n > 0$ be a finite sample expansion factor such that $\delta_n \log^{1/2} n \to 0$ but $\delta_n \log n \to \infty$. For example, we recommend to set $\delta_n = 1/(4 \log^{3/4} n)$. Then for $c_n(1 - \alpha) = k_n(1 - \alpha) + \delta_n$ we define the confidence bands of asymptotic level $1 - \alpha$ to be

$$[i(u, w), \ddot{i}(u, w)] = [\widehat{\theta}(u, w) - c_n(1 - \alpha)\widehat{\sigma}_n(u, w), \ \widehat{\theta}(u, w) + c_n(1 - \alpha)\widehat{\sigma}_n(u, w)], \ (u, w) \in I.$$

The following theorem establishes the asymptotic validity of these confidence bands.

Theorem 12 (Uniform Inference for Linear Functionals). *Suppose that the conditions of Theorem 11 hold for a given coupling method.*
(1) Then

$$P\left\{ \|t_n\|_I \leq c_n(1 - \alpha) \right\} \geq 1 - \alpha + o(1). \tag{4.16}$$

(2) As a consequence, the confidence bands constructed above cover $\theta(u, w)$ uniformly for all $(u, w) \in I$ with probability that is asymptotically no less than $1 - \alpha$, namely

$$P\Big\{\theta(u, w) \in [i(u, w), \ddot{i}(u, w)], \text{ for all } (u, w) \in I\Big\} \geq 1 - \alpha + o(1). \qquad (4.17)$$

(3) The width of the confidence band $2c_n(1 - \alpha)\widehat{\sigma}_n(u, w)$ obeys uniformly in $(u, w) \in I$:

$$2c_n(1 - \alpha)\widehat{\sigma}_n(u, w) = 2k_n(1 - \alpha)(1 + o_P(1))\sigma_n(u, w). \qquad (4.18)$$

(4) Furthermore, if $\|t_n^*\|_I$ does not concentrate at $k_n(1 - \alpha)$ at a rate faster than $\sqrt{\log n}$, that is, it obeys the anti-concentration property $P(\|t_n^*\|_I \leq k_n(1 - \alpha) + \varepsilon_n) = 1 - \alpha + o(1)$ for any $\varepsilon_n = o(1/\sqrt{\log n})$, then the inequalities in (4.16) and (4.17) hold as equalities, and the finite sample adjustment factor δ_n could be set to zero.

The theorem shows that the confidence bands constructed above maintain the required level asymptotically and establishes that the uniform width of the bands is of the same order as the uniform rate of convergence. Moreover, under anti-concentration the confidence intervals are asymptotically similar.

Comment 4.1. This inferential strategy builds on [15], who proposed a similar strategy for inference on the minimum of a function. The idea was not to find the limit distribution, which may not exist in some cases, but to use distributions provided by couplings. Since the limit distribution needs not exist, it is not immediately clear that the confidence intervals maintain the right asymptotic level. However, the additional adjustment factor δ_n assures the right asymptotic level. A small price to pay for using the adjustment δ_n is that the confidence intervals may not be similar, i.e. remain asymptotically conservative in coverage. However, the width of the confidence intervals is not asymptotically conservative, since δ_n is negligible compared to $k_n(1 - \alpha)$. If an additional property, called anti-concentration, holds, then the confidence intervals automatically become asymptotically similar. The anti-concentration property holds if, after appropriate scaling by some deterministic sequences a_n and b_n, the inferential statistic $a_n(\|t_n\|_I - b_n)$ has a continuous limit distribution. More generally, it holds if for any subsequence of integers $\{n_k\}$ there is a further subsequence $\{n_{k_r}\}$ along which $a_{n_{k_r}}(\|t_{n_{k_r}}\|_I - b_{n_{k_r}})$ has a continuous limit distribution, possibly dependent on the subsequence. For an example of the latter see [9], where certain inferential processes converge in distribution to tight Gaussian processes subsubsequentially. We expect anti-concentration to hold in our case, but our constructions and results do not critically hinge on it.

Comment 4.2 (Primitive Bounds on $\xi_\theta(m, I)$)**.** The results of this section rely on the quantity $\xi_\theta(m, I)$. The value of $\xi_\theta(m, I)$ depends on the choice of basis for the series estimator and on the type of the linear functional. Here we discuss the case of regression splines and refer to [29] and [12] for other choices of basis. After a possible renormalization of X, we assume its support is $\mathcal{X} = [-1, 1]^d$. For splines it has been established that $\zeta_m \lesssim \sqrt{m}$ and $\max_k \sup_{x \in \mathcal{X}} \|\partial_{x_k} Z(x)\| \lesssim m^{1/2+k}$, [29]. Then we have that for

- the function itself: $\theta(u, x) = Q_{Y|X}(u|x), \ \ell(x) = Z(x), \ \xi_\theta(m, I) \lesssim \sqrt{m}$;
- the derivative: $\theta(u, x) = \partial_{x_k} Q_{Y|X}(u|x), \ \ell(x) = \partial_x Z(x), \ \xi_\theta(m, I) \lesssim m^{3/2}$;
- the average derivative: $\theta(u) = \int \partial_{x_k} Q_{Y|X}(u|x) d\mu(x), \ \mathrm{supp}(\mu) \subset \mathrm{int}\mathcal{X}, \ |\partial_{x_k}\mu(x)| \lesssim 1,$
                            $\ell = \int \partial_{x_k} Z(x)\mu(x) \ dx = -\int Z(x)\partial_{x_k}\mu(x) \ dx, \ \xi_\theta(m) \lesssim 1.$

4.4. Imposing Monotonicity on Linear Functionals.

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. $\theta(u', w') \leq \theta(u, w)$ whenever $(u', w') \leq (u, w)$ componentwise.[1] Let $\widehat{\theta}$ and $[\underline{i}, \overline{i}]$ 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 following method, suggested in [14].

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

$$\mathcal{M}q = q \quad \text{if } q \text{ monotone}; \tag{4.19}$$

(2) a distance-reducing condition

$$\|\mathcal{M}q - \mathcal{M}f\|_I \leq \|q - f\|_I; \tag{4.20}$$

and (3) an order-preserving condition

$$q \leq f \quad \text{implies} \quad \mathcal{M}q \leq \mathcal{M}f. \tag{4.21}$$

Examples of operators that satisfy these conditions include:

---

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

1. multivariate rearrangement [14],
2. isotonic projection [4],
3. convex combinations of rearrangement and isotonic regression [14], and
4. convex combinations of monotone minorants and monotone majorants.

The following result establishes that monotonizing point estimators reduces estimation error, and that monotonizing confidence bands increases coverage while reducing the length of the bands. The following result follows from Theorem 12 using the same arguments as in Propositions 2 and 3 in [14].

Corollary 1 (Inference Monotone Linear Functionals). *Let $\theta : I \mapsto K$ be weakly increasing over I and $\widehat{\theta}$ be the QR series estimator of Theorem 10. If $\mathcal{M}$ satisfies the conditions (4.19) and (4.20), then the monotonized QR estimator is necessarily closer to the true value:*

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

Let $[i, \ddot{i}]$ be a confidence band for θ of Theorem 12. If $\mathcal{M}$ satisfies the conditions (4.19) and (4.21), the monotonized confidence bands maintain the asymptotic level of the original intervals:

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

If $\mathcal{M}$ satisfies the condition (4.20), the monotonized confidence bands are shorter in length than the original monotone intervals:

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

Comment 4.3. Another strategy, as mentioned e.g. in [14], is to simply intersect the initial confidence interval with the set of monotone functions. This is done simply by taking the smallest majorant of the lower band and the largest minorant of the upper band. This, in principle, produces shortest intervals with desired asymptotic level. However, this may not have good properties under misspecification, i.e. when approximation error is not relatively small (we do not analyze such cases in this paper, but they can occur in practice), whereas the strategy explored in the corollary retains a number of good properties even in this case. See [14] for a detailed discussion of this point and for practical guidance on choosing particular monotonization schemes.

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 ([33]), using the package `quantreg` for quantile regression ([26]).

5.1. **Empirical Example.** To illustrate our methods with real data, we consider an empirical application on nonparametric estimation of the demand for gasoline. [21], [34], and [39] estimated nonparametrically the average demand function. We estimate nonparametrically the quantile demand and elasticity functions and apply our inference methods to construct confidence bands for the average quantile elasticity function. We use the same data set as in [39], which comes from the National Private Vehicle Use Survey, conducted by Statistics Canada between October 1994 and September 1996. 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 [39] for a more detailed description of the data.) Our sample selection and variable construction also follow [39]. We select into the sample households with non-zero licensed drivers, vehicles, and distance driven. We focus on regular grade gasoline consumption. This selection leaves us with a sample of 5,001 households. Fuel consumption and expenditure are recorded by the households at the purchase level.

We consider the following empirical specification:

$$Y = Q_{Y|X}(U|X), \quad Q_{Y|X}(U|X) = g(W, U) + V'\beta(U), \quad X = (W, V),$$

where Y is the log of total gasoline consumption in liters per month; W is the log of price per liter; U is the unobservable preference of the household to consume gasoline; and V is a vector of 28 covariates. Following [39], the covariate vector includes the log of age, a dummy for the top coded value of age, the log 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 $g(W, U)$, we consider three series approximations in 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 number of series terms is selected by undersmoothing over the specifications chosen by applying least squares cross validation to the corresponding conditional average demand

functions. 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 are reported in Figures 1–3. Fig. 1 plots the initial and rearranged 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.[2] The rows of the figure correspond to the three series approximations. The monotonized estimates in the right panels are obtained using the average rearrangement over both the price and quantile dimensions proposed in [14]. The power and B-spline series approximations show most noticeably non-monotone areas with respect to price at high quantiles, which are removed by the rearrangement.

Fig. 2 shows series estimates of the quantile elasticity surface as a function of price and the quantile index, that is:

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

The estimates from the linear approximation show that the elasticity decreases with the quantile index in the middle of the distribution, but this pattern is reversed at the tails. The power and B-spline estimates show substantial heterogeneity of the elasticity across prices, with individuals at the high quantiles being more sensitive to high prices.[3]

Fig. 3 shows 90% uniform confidence bands for the average quantile elasticity function

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

The rows of the figure correspond to the three series approximations and the columns correspond to the inference methods. We construct the bands using the pivotal, 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 weighted bootstrap uses standard exponential weights and 199 repetitions. The confidence bands show that the

---

[2]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, 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.

[3]These estimates are smoothed by local weighted polynomial regression across the price dimension ([16]), because the unsmoothed elasticity estimates display very erratic behavior.

Linear   **Rearranged**

Power   **Rearranged**

B−splines   **Rearranged**
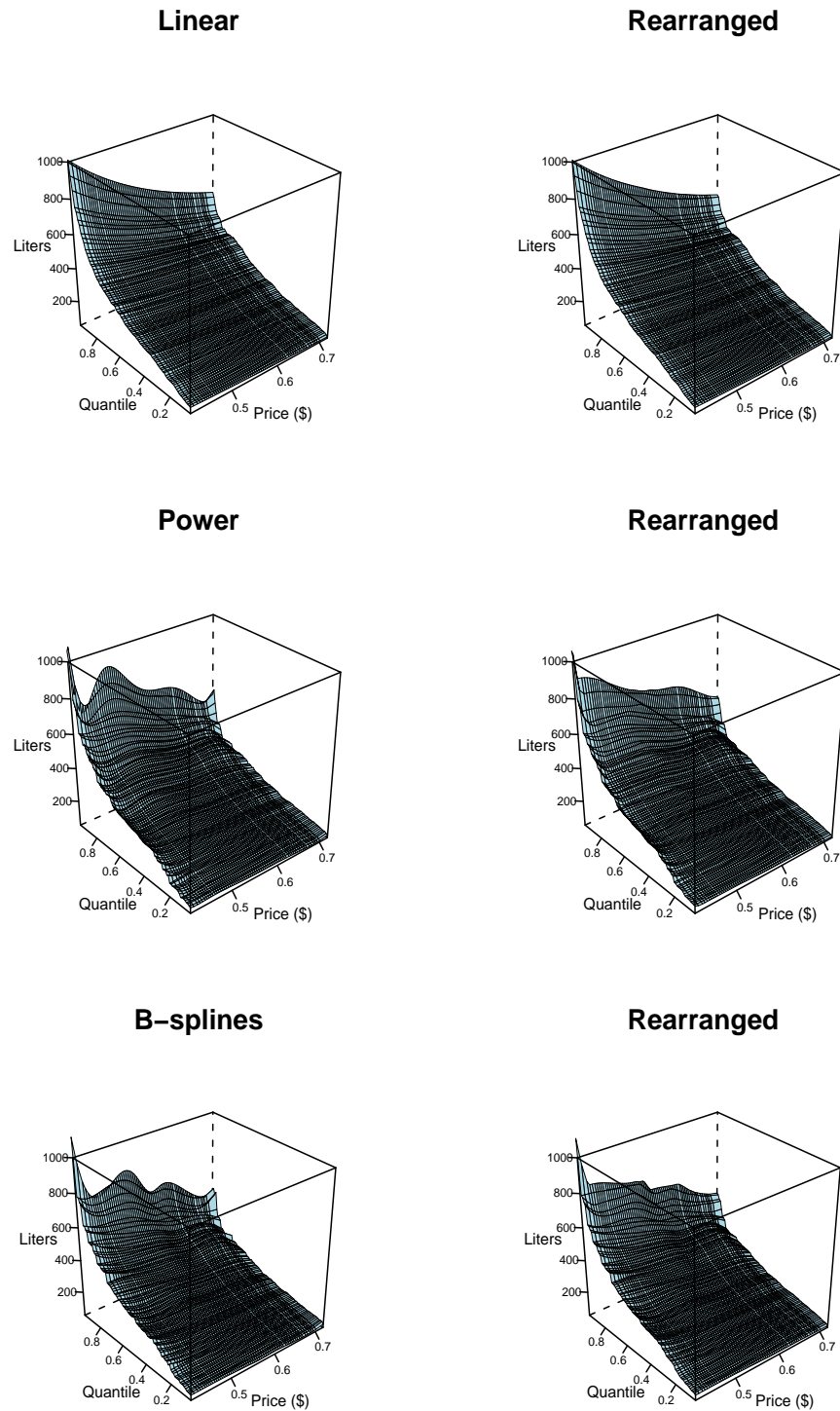
FIGURE 1. Quantile demand surfaces for gasoline as a function of price and
the quantile index. The left panels display linear, power and B-spline series
estimates and the right panels shows the corresponding estimates mono-
tonized by rearrangement over both dimensions.
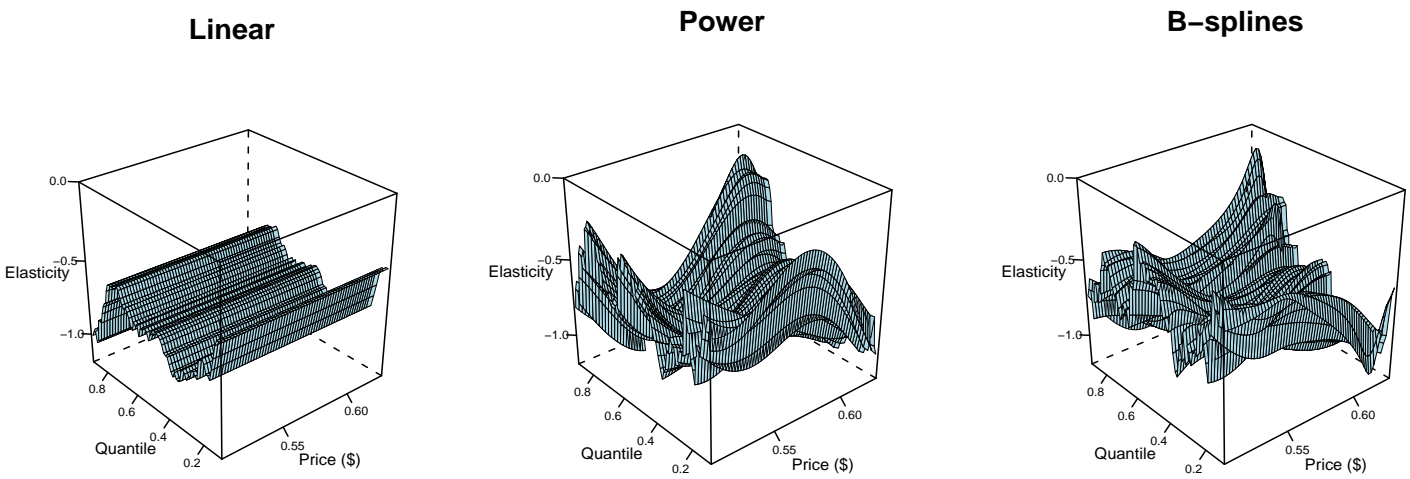
FIGURE 2. Quantile elasticity surfaces as a function of price and the quantile index. The power and B-spline series estimates are smoothed by local weighted polynomial regression with bandwidth 0.5.

evidence of heterogeneity in the elasticities across quantiles is not statistically significant, because we can trace a horizontal line within the bands. They show, however, that there is significant evidence of negative price sensitivity at most quantiles as the bands are bounded away from zero for most quantiles.

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:

$$Y = g(W) + V'\beta + \sigma\Phi^{-1}(U), \tag{5.22}$$

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 4 and 5 examine the quality of the series approximations in 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 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. 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 (5.22). 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.
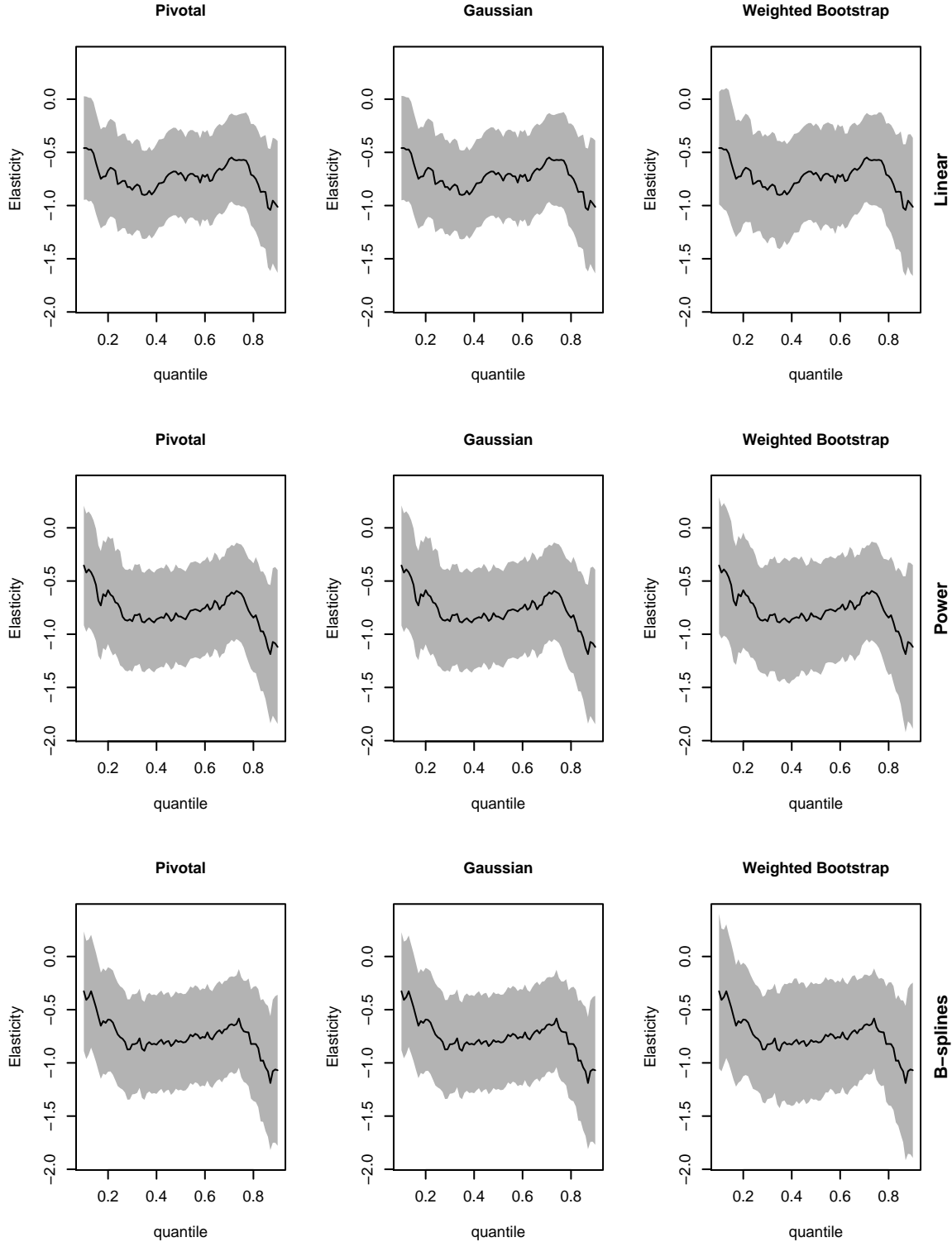
FIGURE 3. 90% Confidence bands for the average quantile elasticity function. Pivotal and Gaussian bands are obtained by 1,000 simulations. Weighted bootstrap bands are based on 199 bootstrap repetitions with standard exponential weights.

True

Linear

Power

B−spline

Figure 4. Estimands of the quantile demand surface. Estimands for the linear, power and B-spline series estimators are obtained numerically using 500,100 simulations.

FIGURE 5. Estimands of the quantile elasticity surface. Estimands for the linear, power and B-spline series estimators are obtained numerically using 500,100 simulations.

To analyze the properties of the inferenge methods in finite samples, we draw 500 samples from the DGP in equation (5.22) with 3 sample sizes, n: $5,001$, $1,000$, and 500 observations. Fo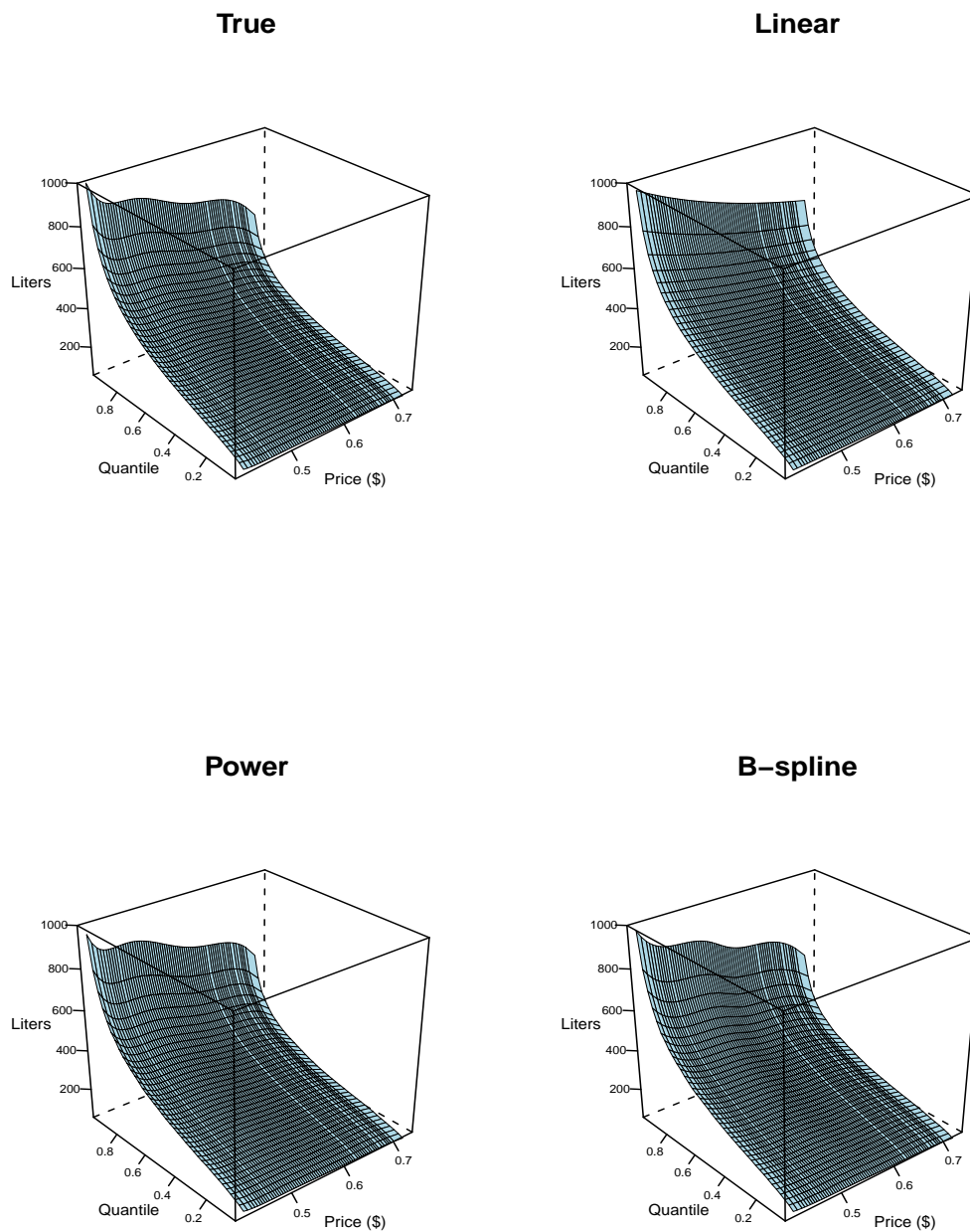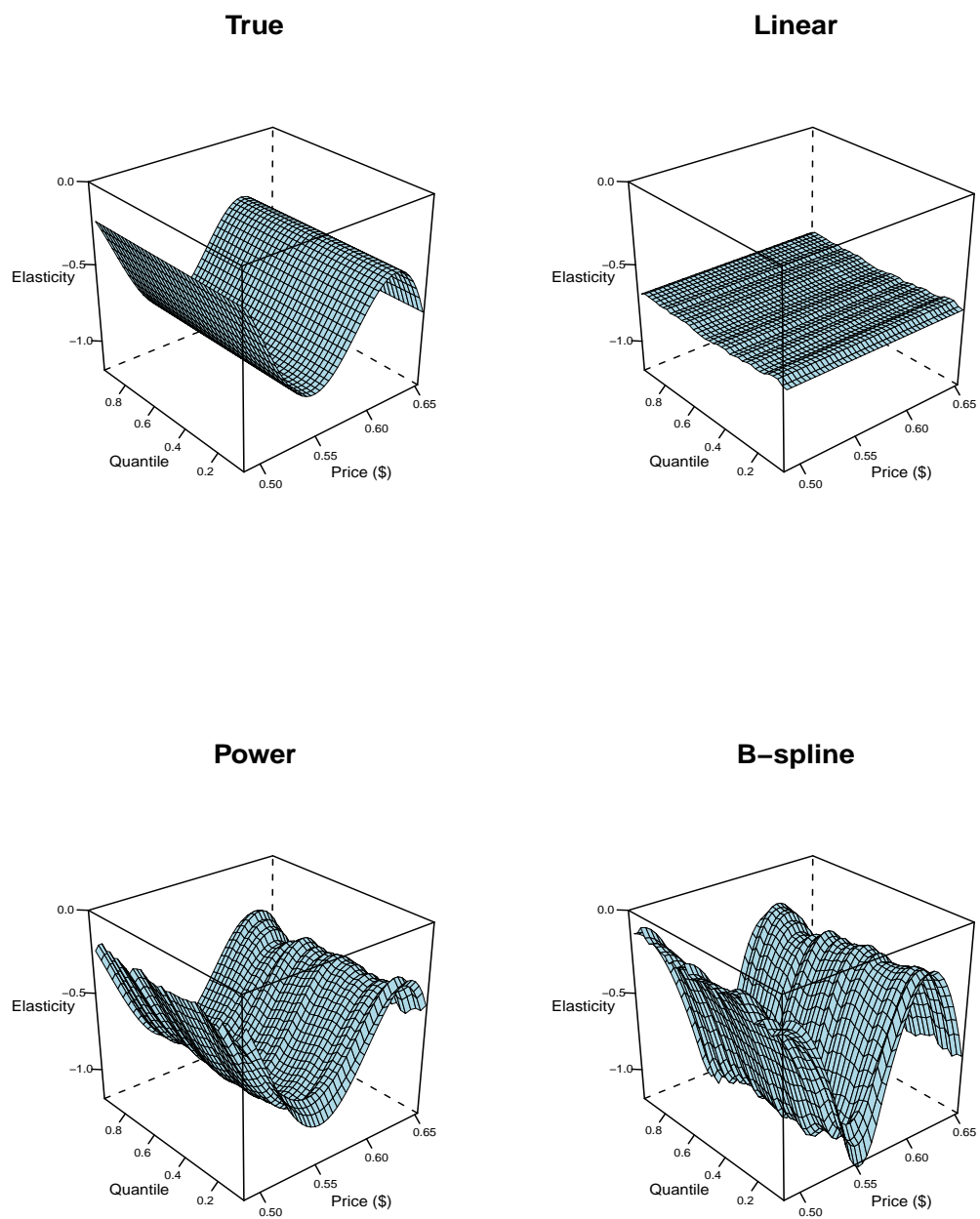r $n = 5,001$ we fix W to the values in the data set, whereas for the smaller sample sizes we draw W with replacement from the values in the data set and keep it fixed across samples. To speed up computation, we drop the vector V by fixing it at the sample median values of the ordinal components and at one for the dummies corresponding to the sample modal values for all the individuals. We focus on the average quantile 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, 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 1 reports estimation and inference results averaged across 200 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 (4.11). The bandwidth for $\widehat{J}_m(u)$ is chosen using the Hall-Sheather option of the `quantreg R` package ([19]). 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 1 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 weighted bootstrap bands tend to have larger average length than the pivotal and Gaussian bands.

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.

TABLE 1. Finite Sample Properties of Estimation and Inference Methods for Average Quantile Elasticity Function

| | Bias | RMSE | SE/SD | Pivotal | | | Gaussian | | | Weighted Bootstrap | | |
				Cover	Length	Stat	Cover	Length	Stat	Cover	Length	Stat
						$n = 5,001$						
Linear	0.05	0.14	1.04	90	0.77	2.64	90	0.76	2.64	87	0.82	2.87
Power	0.00	0.15	1.03	91	0.85	2.65	91	0.85	2.65	88	0.91	2.83
B-spline	0.01	0.15	1.02	90	0.86	2.64	88	0.86	2.64	90	0.93	2.84
						$n = 1,000$						
Linear	0.03	0.29	1.09	92	1.78	2.64	93	1.78	2.64	90	1.96	2.99
Power	0.03	0.33	1.07	92	2.01	2.66	91	2.00	2.65	95	2.17	2.95
B-spline	0.02	0.35	1.05	90	2.08	2.65	90	2.07	2.65	96	2.21	2.95
						$n = 500$						
Linear	0.04	0.45	1.01	88	2.60	2.64	88	2.60	2.64	90	2.84	3.05
Power	0.02	0.52	1.04	90	3.12	2.65	90	3.13	2.66	95	3.29	3.01
B-spline	0.02	0.52	1.04	90	3.25	2.65	90	3.25	2.65	96	3.35	3.00

Notes: 200 repetitions. Simulation standard error for coverage probability is 2%.

Appendix A. Implementation Algorithms

Throughout this section we assume that we have a random sample $\{(Y_i, Z_i) : 1 \leq i \leq n\}$. 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_m by $\widehat{\Sigma}_m = \mathbb{E}_n[Z_i Z_i']$, and the Jacobian matrix $J_m(u)$ by Powell [32] estimator $\widehat{J}_m(u) = \mathbb{E}_n[1\{|Y_i - Z_i'\widehat{\beta}(u)| \leq h_n\} \cdot Z_i Z_i']/2h_n$, where we recommend choosing the bandwidth h_n as in the `quantreg` R package with the Hall-Sheather option ([19]).

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 $\mathcal{U}$.

Algorithm 1 (Pivotal 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 $\mathbb{U}_n^b(u) = n^{-1/2} \sum_{i=1}^n Z_i(u - 1\{U_i^b \leq u\})$, $u \in \mathcal{U}$. (2) Approximate the distribution of $\{\sqrt{n}(\widehat{\beta}(u) - \beta(u)) : u \in \mathcal{U}\}$ by the empirical distribution of $\{\widehat{J}_m^{-1}(u)\mathbb{U}_n^b(u) : u \in \mathcal{U}, 1 \leq b \leq B\}$.*

Algorithm 2 (Gaussian method). *(1) For $b = 1, \ldots, B$, generate a m-dimensional standard Brownian bridge on $\mathcal{U}$, $B_m^b(\cdot)$. Define $G_n^b(u) = \widehat{\Sigma}_m^{-1/2} B_m^b(u)$ 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 $\{\widehat{J}_m^{-1}(u)G_n^b(u) : u \in \mathcal{U}, 1 \leq b \leq B\}$.*

Algorithm 3 (Weighted bootstrap method). *(1) For $b = 1, \ldots, B$, draw $h_1^b, \ldots, h_n^b$ i.i.d. from the standard exponential distribution and compute the weighted quantile regression process $\widehat{\beta}^b(u) = \arg\min_{\beta \in \mathbb{R}^m} \sum_{i=1}^n h_i^b \cdot \rho_u(Y_i - Z_i'\beta)$, $u \in \mathcal{U}$. (2) Approximate the distribution of $\{\sqrt{n}(\widehat{\beta}(u) - \beta(u)) : u \in \mathcal{U}\}$ by the empirical distribution of $\{\sqrt{n}(\widehat{\beta}^b(u) - \widehat{\beta}(u)) : u \in \mathcal{U}, 1 \leq b \leq B\}$.*

Algorithm 4 (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 $\mathbb{U}_n^b(u) = n^{-1/2} \sum_{i=1}^n Z_i(u - 1\{U_i^b \leq u\})$, $u \in \mathcal{U}$. (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 $\{\sqrt{n}(\widehat{\beta}^b(u) - \widehat{\beta}(u)) : u \in \mathcal{U}, 1 \leq b \leq B\}$.*

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 $Q_{Y|X}(\cdot|X)$ including the conditional quantile functions, provided the approximation error is small as stated in Theorems 9 and 11. 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.

Let B be a pre-specified number of bootstrap or simulation repetitions.

Algorithm 5 (Pointwise Inference for Linear Functionals). *(1) Compute the variance estimate $\widehat{\sigma}_n^2(u,w) = u(1-u)\ell(w)'\widehat{J}_m^{-1}(u)\widehat{\Sigma}_m\widehat{J}_m^{-1}(u)\ell(w)/n$. (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 $t_n^{*b}(u,w) = \left|\frac{\ell(w)'V_b(u)}{\sqrt{n}\widehat{\sigma}_n(u,w)}\right|$. (4) Form a $(1-\alpha)$-confidence interval for $\theta(u,w)$ as $\ell(w)'\widehat{\beta}(u) \pm k_n(1-\alpha)\widehat{\sigma}_n(u,w)$, where $k_n(1-\alpha)$ is the $1 - \alpha$ sample quantile of $\{t_n^{*b}(u,w) : 1 \le b \le B\}$.*

Algorithm 6 (Uniform Inference for Linear Functionals). *(1) Compute the variance estimates $\widehat{\sigma}_n^2(u,w) = u(1-u)\ell(w)'\widehat{J}_m^{-1}(u)\widehat{\Sigma}_m\widehat{J}_m^{-1}(u)\ell(w)/n$ for $(u,w) \in I$. (2) Using any of the Algorithms 1-4, compute the processes $V_1(\cdot), \ldots, V_B(\cdot)$ whose empirical distribution approximates the distribution of $\{\sqrt{n}(\widehat{\beta}(u) - \beta(u)) : u \in \mathcal{U}\}$. (3) For $b = 1, \ldots, B$, compute the maximal t-statistic $\|t_n^{*b}\|_I = \sup_{(u,w)\in I} \left|\frac{\ell(w)'V_b(u)}{\sqrt{n}\widehat{\sigma}_n(u,w)}\right|$. (4) Form a $(1-\alpha)$-confidence band for $\{\theta(u,w) : (u,w) \in I\}$ as $\{\ell(w)'\widehat{\beta}(u) \pm k_n(1-\alpha)\widehat{\sigma}_n(u,w) : (u,w) \in I\}$, where $k_n(1-\alpha)$ is the $1 - \alpha$ sample quantile of $\{\|t_n^{*b}\|_I : 1 \le b \le B\}$.*

Appendix B. A result on identification of QR series approximation in population and its relation to the best L_2-approximation

In this section we provide results on identification and approximation properties for the QR series approximation. In what follows, we denote $z = Z(x) \in \mathcal{Z}$ for some $x \in \mathcal{X}$, and for a function $h : \mathcal{X} \to \mathbb{R}$ we define

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

so that $\beta(u) \in \arg\min_{\beta \in \mathbb{R}^m} Q_u(Z'\beta)$. Also let $\underline{f} := \inf_{u \in \mathcal{U}, X \in \mathcal{X}} f_{Y|X}(Q_{Y|X}(u|X)|X)$, where $\underline{f} > 0$ by condition S.2.

Consider the best L_2-approximation to the conditional quantile function $g_u(\cdot) = Q_{Y|X}(u|\cdot)$ by a linear combination of the chosen basis, namely

$$\tilde{\beta}^*(u) \in \arg\min_{\beta \in \mathbb{R}^m} E\left[|Z'\beta - g_u(X)|^2\right]. \tag{B.23}$$

We consider the following approximation rates associated with $\tilde{\beta}^*(u)$:

$$c_{u,2}^2 = E\left[|Z'\tilde{\beta}^*(u) - g_u(X)|^2\right] \quad \text{and} \quad c_{u,\infty} = \sup_{x \in \mathcal{X}, z = Z(x)} |z'\tilde{\beta}^*(u) - g_u(x)|.$$

Lemma 1. *Assume that conditions S.2-S.4, $\zeta_m c_{u,2} = o(1)$ and $c_{u,\infty} = o(1)$ hold. Then, as n grows, we have the following approximation properties for $\beta(u)$:*

$$E\left[|Z'\beta(u) - g_u(X)|^2\right] \leq (16 \vee 3\bar{f}/\underline{f})c_{u,2}^2, \quad E\left[|Z'\beta(u) - Z'\tilde{\beta}^*(u)|^2\right] \leq (9 \vee 8\bar{f}/\underline{f})c_{u,2}^2 \quad and$$

$$\sup_{x \in \mathcal{X}, z = Z(x)} |z'\beta(u) - g_u(x)| \lesssim c_{u,\infty} + \zeta_m \sqrt{(9 \vee 8\bar{f}/\underline{f})} c_{u,2}/\sqrt{\text{mineig}(\Sigma_m)}.$$

Proof of Lemma 1. We assume that $E\left[|Z'\beta(u) - Z'\tilde{\beta}^*(u)|^2\right]^{1/2} \geq 3E\left[|Z'\tilde{\beta}^*(u) - g_u(X)|^2\right]^{1/2}$, otherwise the statements follow directly. The proof proceeds in 4 steps.

Step 1 (Main argument). For notational convenience let

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

By Steps 2 and 3 below we have respectively

$$Q_u(Z'\beta(u)) - Q_u(g_u) \leq \bar{f}c_{u,2}^2 \quad \text{and} \tag{B.24}$$

$$Q_u(Z'\beta(u)) - Q_u(g_u) \geq \frac{\underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right]}{3} \wedge \left(\frac{\bar{q}}{3}\sqrt{\underline{f}E\left[|Z'\beta(u) - g_u(X)|^2\right]}\right). \tag{B.25}$$

Thus

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

As n grows, since $\sqrt{\bar{f}c_{u,2}^2} < \bar{q}/\sqrt{3}$ by Step 4 below, it follows that $E\left[|Z'\beta(u) - g_u(X)|^2\right] \leq 3c_{u,2}^2\bar{f}/\underline{f}$ which proves the first statement regarding $\beta(u)$.

The second statement regarding $\beta(u)$ follows since

$$\sqrt{E\left[|Z'\beta(u) - Z'\tilde{\beta}^*(u)|^2\right]} \leq \sqrt{E\left[|Z'\beta(u) - g_u(X)|^2\right]} + \sqrt{E\left[|Z'\tilde{\beta}^*(u) - g_u(X)|^2\right]} \leq \sqrt{3c_{u,2}^2\bar{f}/\underline{f}} + c_{u,2}.$$

Finally, the third statement follows by the triangle inequality, S.4, and the second statement

$$
\begin{aligned}
\sup_{x\in\mathcal{X}, z=Z(x)}|z'\beta(u)-g_u(x)| &\leq \sup_{x\in\mathcal{X}, z=Z(x)}|z'\tilde{\beta}^*(u)-g_u(x)|+\zeta_m\|\beta(u)-\tilde{\beta}^*(u)\| \\
&\leq c_{u,\infty}+\zeta_m\sqrt{E\left[|Z'\tilde{\beta}^*(u)-Z'\beta(u)|^2\right]/\mathrm{mineig}(\Sigma_m)} \\
&\leq c_{u,\infty}+\zeta_m\sqrt{(9\vee 8\bar{f}/\underline{f})c_{u,2}}/\sqrt{\mathrm{mineig}(\Sigma_m)}.
\end{aligned}
$$

Step 2 (Upper Bound). For any two scalars w and v we have that

$$
\rho_u(w-v)-\rho_u(w) = -v(u-1\{w\leq 0\})+\int_0^v (1\{w\leq t\}-1\{w\leq 0\})dt. \tag{B.26}
$$

By (B.26) and the law of iterated expectations, for any measurable function h

$$
\begin{aligned}
Q_u(h)-Q_u(g_u) &= E\left[\int_0^{h-g_u} F_{Y|X}(g_u+t|X)-F_{Y|X}(g_u|X)dt\right] \\
&= E\left[\int_0^{h-g_u} tf_{Y|X}(g_u+\tilde{t}_{X,t}|X)dt\right]\leq (\bar{f}/2)E[|h-g_u|^2]
\end{aligned} \tag{B.27}
$$

where $\tilde{t}_{X,t}$ lies between 0 and t for each $t\in[0,h(x)-g_u(x)]$.

Thus, (B.27) with $h(X)=Z'\tilde{\beta}^*(u)$ and $Q_u(Z'\beta(u))\leq Q_u(Z'\tilde{\beta}^*(u))$ imply that

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

Step 3 (Lower Bound). To establish a lower bound note that for a measurable function h

$$
\begin{aligned}
Q_u(h)-Q_u(g_u) &= E\left[\int_0^{h-g_u} F_{Y|X}(g_u+t|X)-F_{Y|X}(g_u|X)dt\right] \\
&= E\left[\int_0^{h-g_u} tf_{Y|X}(g_u|X)+\tfrac{t^2}{2}f'_{Y|X}(g_u+\tilde{t}_{X,t}|X)dt\right] \\
&\geq (\underline{f}/2)E[|h-g_u|^2]-\tfrac{1}{6}\bar{f}'E[|h-g_u|^3].
\end{aligned} \tag{B.28}
$$

If $\sqrt{\underline{f}E\left[|Z'\beta(u)-g_u(X)|^2\right]}\leq \bar{q}$, then $\bar{f}'E\left[|Z'\beta(u)-g_u(X)|^3\right]\leq \underline{f}E\left[|Z'\beta(u)-g_u(X)|^2\right]$ and (B.28) with the function $h(Z)=Z'\beta(u)$ yields

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

On the other hand, if $\sqrt{\underline{f}E\left[|Z'\beta(u)-g_u(X)|^2\right]}>\bar{q}$, let $h_u(X)=(1-\alpha)Z'\beta(u)+\alpha g_u(X)$ where $\alpha\in(0,1)$ is picked so that $\sqrt{\underline{f}E\left[|h_u-g_u|^2\right]}=\bar{q}$. Then by (B.28) and convexity of Q_u we have

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

Next note that $h_u(X) - g_u(X) = (1-\alpha)(Z'\beta(u) - g_u(X))$, thus

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

Using that and applying (B.28) with h_u we obtain

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

Therefore

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

Step 4. ($\bar{f}c_{u,2}^2 < \bar{q}^2/3$ as n grows) Recall that by S.4 $\|Z\| \leq \zeta_m$ and that we can assume

$$E\left[|Z'\beta(u) - Z'\tilde{\beta}^*(u)|^2\right]^{1/2} \geq 3E\left[|Z'\tilde{\beta}^*(u) - g_u(X)|^2\right]^{1/2}.$$

Then, using the relation above (in the second inequality)

$$
\begin{aligned}
\frac{E[|Z'\beta(u) - g_u(X)|^2]^{3/2}}{E[|Z'\beta(u) - g_u(X)|^3]} \;&\geq\; \frac{E[|Z'\beta(u) - g_u(X)|^2]^{1/2}}{\sup_{x \in \mathcal{X}, z = Z(x)}|z'\beta(u) - g_u(x)|} \\[4pt]
&\geq \frac{1}{2}\frac{E\left[|Z'\beta(u) - Z'\tilde{\beta}^*(u)|^2\right]^{1/2} + E\left[|Z'\tilde{\beta}^*(u) - g_u(X)|^2\right]^{1/2}}{\sup_{x \in \mathcal{X}, z = Z(z)}|z'\beta(u) - z'\tilde{\beta}^*(u)| + \sup_{x \in \mathcal{X}, z = Z(x)}|z'\tilde{\beta}^*(u) - g_u(x)|} \\[4pt]
&\geq \frac{1}{2}\left(\frac{E\left[|Z'\beta(u) - Z'\tilde{\beta}^*(u)|^2\right]^{1/2}}{\sup_{x \in \mathcal{X}, z = Z(x)}|z'\beta(u) - z'\tilde{\beta}^*(u)|} \wedge \frac{E\left[|Z'\tilde{\beta}^*(u) - g_u(X)|^2\right]^{1/2}}{\sup_{x \in \mathcal{X}, z = Z(x)}|z'\tilde{\beta}^*(u) - g_u(x)|}\right) \\[4pt]
&\geq \frac{1}{2}\left(\frac{\bar{\kappa}\|\beta(u) - \tilde{\beta}^*(u)\|}{\zeta_m\|\beta(u) - \tilde{\beta}^*(u)\|} \wedge \frac{c_{u,2}}{c_{u,\infty}}\right)
\end{aligned}
$$

where $\bar{\kappa} = \sqrt{\mathrm{mineig}(\Sigma_m)}$.

Finally, $\sqrt{\bar{f}}c_{u,2} < (\underline{f}^{3/2}/\overline{f'})(1/2)(\frac{\bar{\kappa}}{\zeta_m} \wedge \frac{c_{u,2}}{c_{u,\infty}})/\sqrt{3} \leq \bar{q}/\sqrt{3}$ as n grows under the condition $\zeta_m c_{u,2} = o(1)$, $c_{u,\infty} = o(1)$, and conditions S.2 and S.3. $\qquad \square$

APPENDIX C. PROOF OF THEOREMS 1-7

In this section we gather the proofs of the theorems 1-7 stated in the main text. We adopt the standard notation of the empirical process literature [37]. We begin by assuming that the sequences m and n satisfy $m/n \to 0$ as $m, n \to \infty$. For notational convenience we write $\psi_i(\beta, u) = Z_i(1\{Y_i \leq Z_i'\beta\} - u)$, where $Z_i = Z(X_i)$. Also for any sequence $r = r_n = o(1)$ and any fixed $0 < B < \infty$, we define the set

$$R_{n,m} := \{(u, \beta) \in \mathcal{U} \times \mathbb{R}^m : \|\beta - \beta(u)\| \leq Br\}$$

and the following error terms:

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

$$\epsilon_1(m,n) \quad := \quad \sup_{(u,\beta) \in R_{n,m}} \|\mathbb{G}_n(\psi_i(\beta,u)) - \mathbb{G}_n(\psi_i(\beta(u),u))\|,$$

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

In what follows, we say that the data are in general position if for any $\gamma \in \mathbb{R}^m$, $P(Y_i = Z_i'\gamma, \text{for at least one } i) = 0$. Under the bounded density condition S.2, the data are in the general position. We also assume i.i.d. sampling throughout the appendix, although this condition is not needed for most of the results.

C.1. **Proof of Theorem 1.** We start by establishing uniform rates of convergence. The following technical lemma will be used in the proof of Theorem 1.

Lemma 2 (Rates in Euclidian Norm for Perturbed QR Process). *Suppose that $\widehat{\beta}(u)$ is a minimizer of*

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

for each $u \in \mathcal{U}$, and that the perturbation term obeys $\sup_{u \in \mathcal{U}}\|\mathcal{A}_n(u)\| \lesssim_P r = o(1)$. The unperturbed case corresponds to $\mathcal{A}_n(\cdot) = 0$. If $\inf_{u \in \mathcal{U}} \text{mineig}[J_m(u)] > J > 0$ and the conditions

R1. $\epsilon_0(m,n) \lesssim_P \sqrt{n}r$,
R2. $\epsilon_1(m,n) \lesssim_P \sqrt{n}r$,
R3. $\epsilon_2(m,n) \lesssim_P \sqrt{n}r$,

hold, where the constants in the bounds above can be taken to be independent of the constant B in the definition of $R_{m,n}$, then for any $\varepsilon > 0$, there is a sufficiently large B such that with probability at least $1 - \varepsilon$, $(u, \widehat{\beta}(u)) \in R_{m,n}$ uniformly over $u \in \mathcal{U}$, that is,

$$\sup_{u \in \mathcal{U}}\left\|\widehat{\beta}(u) - \beta(u)\right\| \lesssim_P r. \tag{C.29}$$

Proof of Lemma 2. Due to the convexity of the objective function, it suffices to show that for any $\varepsilon > 0$, there exists $B < \infty$ such that

$$P\left(\inf_{u \in \mathcal{U}} \inf_{\|\eta\|=1} \eta'\left[\mathbb{E}_n\left[\psi_i(\beta,u)\right] + \mathcal{A}_n(u)\right]|_{\beta = \beta(u) + Br\eta} > 0\right) \geq 1 - \varepsilon. \tag{C.30}$$

Indeed, the quantity $\mathbb{E}_n\left[\psi_i\left(\beta,u\right)\right] + \mathcal{A}_n(u)$ is a subgradient of the objective function at β. Observe that uniformly in $u \in \mathcal{U}$,

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

since $E\left[\psi_i(\beta(u), u)\right] = 0$ by definition of $\beta(u)$ (see argument in the proof of Lemma 3).

Invoking R2 and R3,

$$\epsilon_1(m,n) + \epsilon_2(m,n) \lesssim_P \sqrt{n}r,$$

and by R1, uniformly in $\eta \in S^{m-1}$ we have

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

Then the event of interest in (C.30) is implied by the event

$$\left\{\eta'J_m(u)\eta B\sqrt{n}r - \epsilon_0(m,n) - \epsilon_1(m,n) - \epsilon_2(m,n) - \sqrt{n}\sup_{u \in \mathcal{U}}\|\mathcal{A}_n(u)\| > 0\right\},$$

whose probability can be made arbitrarily close to 1, for large n, by setting B sufficiently large since $\sup_{u \in \mathcal{U}}\|\mathcal{A}_n(u)\| \lesssim_P r$, and $\eta'J_m(u)\eta \geq J > 0$ by the condition on the eigenvalues of $J_m(u)$. $\qquad\square$

Proof of Theorem 1. Let $\phi_n = \sup_{\alpha \in S^{m-1}} E[(Z_i'\alpha)^2] \vee \mathbb{E}_n[(Z_i'\alpha)^2]$. Under $\zeta_m^2 \log n = o(n)$ and S.3, $\phi_n \lesssim_P 1$ by Corollary 2 in Appendix G.

Next recall that $R_{m,n} = \{(u,\beta) \in \mathcal{U} \times \mathbb{R}^m : \|\beta - \beta(u)\| \leq Br\}$ for some fixed B large enough.

Under S.1-S.5, $\epsilon_0(m,n) \lesssim_P \sqrt{m\phi_n \log n}$ by Lemma 23, and $\epsilon_1(m,n) \lesssim_P \sqrt{m\phi_n \log n}$ by Lemma 22, where none of these bounds depend on B. Under S.1-S.5, $\epsilon_2(m,n) \lesssim \sqrt{n}\zeta_m B^2 r^2 + \sqrt{n}m^{-\kappa}Br \lesssim \sqrt{n}r$ by Lemma 24 provided $\zeta_m B^2 r = o(1)$ and $m^{-\kappa}B = o(1)$.

Finally, since $\zeta_m^2 m \log n = o(n)$ by the growth condition of the theorem, we can take $r = \sqrt{(m\log n)/n}$ in Lemma 2 with $\mathcal{A}_n(u) = 0$ and the result follows. $\qquad\square$

C.2. **Proof of Theorems 2, 3, 4, 5 and 6.** In order to establish our uniform linear approximations for the perturbed QR process, it is convenient to define the following approximation error:

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

Lemma 3 (Uniform Linear Approximation)**.** *Suppose that* $\widehat{\beta}(u)$ *is a minimizer of*

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

for each $u \in \mathcal{U}$, the data are in general position, and the conditions of Lemma 2 hold. The unperturbed case corresponds to $\mathcal{A}_n(\cdot) = 0$. Then

$$\sqrt{n}J_m(u)\left(\widehat{\beta}(u) - \beta(u)\right) = \frac{-1}{\sqrt{n}}\sum_{i=1}^{n}\psi_i(\beta(u), u) - \mathcal{A}_n(u) + r_n(u), \tag{C.31}$$

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

Proof of Lemma 3. First note that $E[\psi_i(\beta(u), u)] = 0$ by definition of $\beta(u)$. Indeed, despite the possible approximation error, $\beta(u)$ minimizes $E[\rho_u(Y - Z'\beta)]$ which yields $E[\psi_i(\beta(u), u)] = 0$ by the first order conditions.

Therefore equation (C.31) can be recast as

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

Next note that if $(u, \widehat{\beta}(u)) \in R_{n,m}$ uniformly in $u \in \mathcal{U}$ by the triangle inequality, uniformly in $u \in \mathcal{U}$

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

The result follows by Lemma 2 which bounds from below the probability of $(u, \widehat{\beta}(u)) \in R_{n,m}$ uniformly in $u \in \mathcal{U}$.

□

Proof of Theorem 2. We first control the approximation errors ϵ_0, ϵ_1, ϵ_2, and ϵ_3. Lemma 23 implies that $\epsilon_0(m, n) \lesssim_P \sqrt{m\phi_n \log n}$. Lemma 24 implies that

$$\epsilon_1(m, n) \lesssim_P \sqrt{m\zeta_m r \log n} + \frac{m\zeta_m \log n}{\sqrt{n}} \quad \text{and} \quad \epsilon_2(m, n) \lesssim_P \sqrt{n}\zeta_m r^2 + \sqrt{n}m^{-\kappa}r.$$

Next note that under the bounded density condition in S.2, the data are in general position. Therefore, by Lemma 26 with probability 1 we have

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

Thus, the assumptions required by Lemma 3 follow under $m^3\zeta_m^2 \log^3 n = o(n)$, the uniform rate $r = \sqrt{(m \log n)/n}$, and the condition on κ.

Thus by Lemma 3 with $\mathcal{A}_n(\cdot) = 0$

$$\sqrt{n} J_m(u) \left(\widehat{\beta}(u) - \beta(u) \right) = \frac{-1}{\sqrt{n}} \sum_{i=1}^{n} \psi_i(\beta(u), u) + r_n(u), \quad \text{where}$$

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

Next, let

$$\widetilde{r}_u := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i (1\{Y_i \leq Z_i'\beta(u) + R(X_i, u)\} - 1\{Y_i \leq Z_i'\beta(u)\}).$$

By Lemma 25 $\sup_{u \in \mathcal{U}} \|\widetilde{r}_u\| \lesssim_P \sqrt{m^{1-\kappa} \log n} + \frac{m \zeta_m \log n}{\sqrt{n}}$. The result follows since $\mathbb{U}_n(u) =_d \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i(u - 1\{Y_i \leq Z_i'\beta(u) + R(X_i, u)\})$. □

Proof of Theorem 3. Note that

$$\sup_{u \in \mathcal{U}} \|r_n(u)\| \leq \sup_{\alpha \in S^{m-1}, u \in \mathcal{U}} |\alpha'(J_m^{-1}(u) - \widehat{J}_m^{-1}(u))\alpha| \sup_{u \in \mathcal{U}} \|\mathbb{U}_n^*(u)\|.$$

To bound the second term, by Lemma 23

$$\sup_{u \in \mathcal{U}} \|\mathbb{U}_n^*(u)\| \lesssim_P \sqrt{m \log n} \tag{C.32}$$

since $\phi_n \lesssim_P 1$ by Corollary 2 in Appendix G under $\zeta_m^2 \log n = o(n)$.

Next since $J_m(u)$ has eigenvalues bounded away from zero by S.2 and S.3, $J_m^{-1}(u)$ also has eigenvalues bounded above by a constant. Moreover, by Lemma 4

$$\sup_{\alpha \in S^{m-1}, u \in \mathcal{U}} |\alpha'(\widehat{J}_m(u) - J_m(u))\alpha| \lesssim_P \epsilon_5(m, n) + \epsilon_6(m, n)$$

$$\lesssim_P \sqrt{\frac{\zeta_m^2 m \log n}{n h_n}} + \frac{m \zeta_m^2 \log n}{n h_n} + m^{-\kappa} + \sqrt{\frac{m \log n}{n}} \zeta_m + h_n,$$

where the second inequality follows by Lemma 28 with $r = \sqrt{(m \log n)/n}$. Under the growth conditions stated in the Theorem,

$$\sup_{\alpha \in S^{m-1}, u \in \mathcal{U}} |\alpha'(\widehat{J}_m(u) - J_m(u))\alpha| \lesssim o_P \left(1/\sqrt{m \log^3 n} \right).$$

Thus, with probability approaching one, $\widehat{J}_m^{-1}(u)$ has eigenvalues bounded above by a constant uniformly in $u \in \mathcal{U}$.

Moreover, by the matrix identity $A^{-1} - B^{-1} = B^{-1}(B - A)A^{-1}$

$$
\begin{aligned}
|\alpha'(J_m^{-1}(u) - \widehat{J}_m^{-1}(u))\alpha| &= |\alpha'\widehat{J}_m^{-1}(u)(\widehat{J}_m(u) - J_m(u))J_m^{-1}(u)\alpha| \\
&\leq \|\alpha'\widehat{J}_m^{-1}(u)\| \sup_{\tilde{\alpha} \in S^{m-1}} |\tilde{\alpha}'(\widehat{J}_m(u) - J_m(u))\tilde{\alpha}| \; \|J_m^{-1}(u)\alpha\| \quad \text{(C.33)} \\
&\lesssim o_P\left(1/\sqrt{m\log^3 n}\right),
\end{aligned}
$$

which holds uniformly over $\alpha \in S^{m-1}$ and $u \in \mathcal{U}$.

Finally, the stated bound on $\sup_{u \in \mathcal{U}} \|r_n(u)\|$ follows by combining the relations (C.32) and (C.33). $\qquad \square$

Proof of Theorem 4. The first part of the proof is similar to the proof of Theorem 2 but applying Lemma 3 twice, one to the unperturbed problem and one to the perturbed problem with $\mathcal{A}_n(u) = -\mathbb{U}_n^*(u)/\sqrt{n}$, for every $u \in \mathcal{U}$. Consider the set E of realizations of the data $\mathcal{D}_n$ such that $\sup_{\|\alpha\|=1} \mathbb{E}_n[(\alpha'Z_i)^2] \lesssim 1$. By Corollary 2 in Appendix G and assumptions S.3 and S.4, $P(E) = 1 - o(1)$ under our growth conditions. Thus, by Lemma 23

$$
\sup_{u \in \mathcal{U}} \|\mathcal{A}_n(u)\| \lesssim_P r = \sqrt{\frac{m\log n}{n}}
$$

where we used that $\phi_n \lesssim_P 1$ by Corollary 2 in Appendix G under $\zeta_m^2 \log n = o(n)$ and S.3.

Then,

$$
\begin{aligned}
\sqrt{n}J_m(u)\left(\widehat{\beta}^*(u) - \widehat{\beta}(u)\right) &= \sqrt{n}J_m(u)\left(\widehat{\beta}^*(u) - \beta(u)\right) - \sqrt{n}J_m(u)\left(\widehat{\beta}(u) - \beta(u)\right) \\
&= \mathbb{U}_n^*(u) + r_n^{pert}(u) - r_n^{unpert}(u), \quad \text{where}
\end{aligned}
$$

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

Note also that the results continue to hold in P-probability if we replace P by P^*, 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 clearly follows from the first from the Markov inequality, observing that $E[P^*(|B_n| > \ell_n)] = P(|B_n| > \ell_n) = o(1)$. $\qquad \square$

Proof of Theorem 5. The existence of the Gaussian process with the specified covariance structure is trivial. To establish the additional coupling, note that under S.1-S.4 and the growth conditions, $\sup_{\|\alpha\|=1} \mathbb{E}_n[(\alpha'Z_i)^2] \lesssim_P 1$ by Corollary 2 in Appendix G. Conditioning on any realization of $Z_1, ..., Z_n$ such that $\sup_{\|\alpha\|=1} \mathbb{E}_n[(\alpha'Z_i)^2] \lesssim 1$, the existence of the Gaussian process with the specified covariance structure that also satisfies the coupling

condition follows from Lemma 8. Under the conditions of the theorem $\sup_{u\in\mathcal{U}} \|\mathbb{U}_n(u) - G_n(u)\| = o_P(1/\log n)$. Therefore, the second statement follows since

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

$$
\leq \sup_{u\in\mathcal{U}} \|J_m^{-1}(u)\mathbb{U}_n(u) - J_m^{-1}(u)G_n(u)\| + o_P(1/\log n)
$$

$$
\leq \sup_{u\in\mathcal{U}} \|J_m^{-1}(u)\| \sup_{u\in\mathcal{U}} \|\mathbb{U}_n(u) - G_n(u)\| + o_P(1/\log n) = o_P(1/\log n),
$$

where we invoke Theorem 2 under the growth conditions, and that the eigenvalues of $J_m(u)$ are bounded away from zero uniformly in n by S.2 and S.3.

The last statement proceeds similarly to the proof of Theorem 3.

$\square$

Proof of Theorem 6. Note that $\widehat{\beta}^b(u)$ solves the quantile regression problem for the rescaled data $\{(h_iY_i, h_iZ_i) : 1 \leq i \leq n\}$. The weight h_i is independent of (Y_i, Z_i), $E[h_i] = 1$, $E[h_i^2] = 1$, and $\max_{1\leq i\leq n} h_i \lesssim_P \log n$. That allows us to extend all results from $\widehat{\beta}(u)$ to $\widehat{\beta}^b(u)$ replacing ζ_m by $\zeta_m^b = \zeta_m \log n$ to account for the larger envelope, and $\psi_i^b(\beta, u) = h_i\psi_i(\beta, u)$.

The first part of the proof is similar to the proof of Theorem 2 but applying Lemma 3 twice, one to the original problem and one to the problem weighted by $\{h_i\}$. Then

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\beta}^b(u) - \widehat{\beta}(u)\right) &= \sqrt{n}\left(\widehat{\beta}^b(u) - \beta(u)\right) + \sqrt{n}\left(\beta(u) - \widehat{\beta}(u)\right)\\
&= \frac{J_m^{-1}(u)}{\sqrt{n}} \sum_{i=1}^n h_iZ_i(u - 1\{Y_i \leq Z_i'\beta(u)\}) + r_n^b(u) -\\
&\quad -\frac{J_m^{-1}(u)}{\sqrt{n}} \sum_{i=1}^n Z_i(u - 1\{Y_i \leq Z_i'\beta(u)\}) - r_n(u)
\end{aligned}
$$

where $\sup_{u\in\mathcal{U}} \|r_n(u) - r_n^b(u)\| \lesssim_P \sqrt{m\zeta_m r \log^2 n} + \frac{m\zeta_m \log^2 n}{\sqrt{n}} + m^{-\kappa}\sqrt{m\log n} + \frac{m\zeta_m \log n}{\sqrt{n}}$.

Next, let

$$
\widetilde{r}_u := \frac{1}{\sqrt{n}} \sum_{i=1}^n (h_i - 1)Z_i(1\{Y_i \leq Z_i'\beta(u) + R(X_i, u)\} - 1\{Y_i \leq Z_i'\beta(u)\}).
$$

By Lemma 25, $\sup_{u\in\mathcal{U}} \|\widetilde{r}_u\| \lesssim_P \sqrt{m^{1-\kappa}\log n} + \frac{\zeta_m m \log^2 n}{\sqrt{n}}$. The result follows since $1\{U_i \leq u\} = 1\{Y_i \leq Z_i'\beta(u) + R(X_i, u)\}$.

Note also that the results continue to hold in P-probability if we replace P by P^*, 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 clearly follows from the first from the Markov inequality, observing that $E[P^*(|B_n| > \ell_n)] = P(|B_n| > \ell_n) = o(1)$.

The second part of the theorem follows by applying Lemma 8 with $v_i = h_i - 1$, where $h_i \sim \exp(1)$ so that $E[v_i^2] = 1$, $E[|v_i|^3] \lesssim 1$ and $\max_{1 \leq i \leq n} |v_i| \lesssim_P \log n$. Lemma 8 implies that there is a Gaussian process $G_n(\cdot)$ with covariance structure $\mathbb{E}_n[Z_i Z_i'](u \wedge u' - uu')$ such that

$$\sup_{u \in \mathcal{U}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n (h_i - 1) Z_i(u - 1\{U_i \leq u\}) - G_n(u) \right\| \lesssim_P o(1/\log n).$$

Combining the result above with the first part of the theorem, the second part follows by the triangle inequality. $\qquad\square$

C.3. **Proof of Theorem 7 on Matrices Estimation.** Consider the following quantities

$$\epsilon_4(m, n) = \sup_{\alpha \in S^{m-1}} \left| n^{-1/2} \mathbb{G}_n \left((\alpha' Z_i)^2 \right) \right|,$$

$$\epsilon_5(m, n) = \frac{n^{-1/2}}{2h_n} \sup_{\substack{\alpha \in S^{m-1}, \\ (u, \beta) \in R_{m,n}}} \left| \mathbb{G}_n \left(1\{|Y_i - Z_i'\beta| \leq h_n\}(\alpha' Z_i)^2 \right) \right|,$$

$$\epsilon_6(m, n) = \sup_{\substack{\alpha \in S^{m-1}, \\ (u, \beta) \in R_{m,n}}} \left| \frac{1}{2h_n} E \left[1\{|Y_i - Z_i'\beta| \leq h_n\}(\alpha' Z_i)^2 \right] - \alpha' J_m(u) \alpha \right|,$$

where h_n is a bandwidth parameter such that $h_n \to 0$.

Lemma 4 (Estimation of Variance Matrices and Jacobians). *Under the definitions of $\epsilon_4, \epsilon_5,$ and ϵ_6 we have*

$$\sup_{\alpha \in S^{m-1}} \left| \alpha'(\widehat{\Sigma}_m - \Sigma_m)\alpha \right| \leq \epsilon_4(m, n) \qquad \sup_{\substack{\alpha \in S^{m-1}, \\ u \in \mathcal{U}}} \left| \alpha'(\widehat{J}_m(u) - J_m(u))\alpha \right| \leq \epsilon_5(m, n) + \epsilon_6(m, n).$$

Proof. The first relation follows by the definition of $\epsilon_4(m, n)$ and the second by the triangle inequality and definition of $\epsilon_5(m, n)$ and $\epsilon_6(m, n)$. $\qquad\square$

Proof of Theorem 7. This follows immediately from Lemma 4, Lemma 27, and Lemma 28.

$\qquad\square$

APPENDIX D. PROOFS OF THEOREMS 8 AND 9

Proof of Theorem 8. By Theorem 2 and P1

$$
\begin{aligned}
|\widehat{\theta}(u,w) - \theta(u,w)| \quad &\leq |\ell(u)'(\widehat{\beta}(u) - \beta(u))| + |r_n(u,w)| \\
&\leq \frac{|\ell(w)'J_m^{-1}(u)\mathbb{U}_n(u)|}{\sqrt{n}} + o_P\left(\frac{\|\ell(w)\|}{\sqrt{n}\log n}\right) + o(\|\ell(w)\|/\sqrt{n}) \\
&\leq \frac{|\ell(w)'J_m^{-1}(u)\mathbb{U}_n(u)|}{\sqrt{n}} + o_P\left(\frac{\xi_\theta(m,w)}{\sqrt{n}\log n}\right) + o(\xi_\theta(m,w)/\sqrt{n})
\end{aligned}
$$

where the last inequality follows by $\|\ell(w)\| \lesssim \xi_\theta(m,w)$ assumed in P2. Finally, since $E[|\ell(w)'J_m^{-1}(u)\mathbb{U}_n(u)|^2] \lesssim \|\ell(w)\|^2 \|J_m^{-1}(u)\|^2 \sup_{\alpha \in S^{m-1}} E[(\alpha'Z)^2] \lesssim \xi_\theta^2(m,w)$, the result follows by applying the Chebyshev inequality to establish that $|\ell(w)J_m^{-1}(u)\mathbb{U}_n(u)| \lesssim_P \xi_\theta(m,w)$. $\square$

Proof of Theorem 9. By Assumption P1 and Theorem 2

$$
t_n(u,w) \quad = \frac{\ell(w)'(\widehat{\beta}(u)-\beta(u))}{\widehat{\sigma}_n(u,w)} + \frac{r_n(u,w)}{\widehat{\sigma}_n(u,w)} = \frac{\ell(w)'J_m^{-1}(u)\mathbb{U}_n(u)}{\sqrt{n}\widehat{\sigma}_n(u,w)} + o_P\left(\frac{\|\ell(w)\|}{\sqrt{n}\widehat{\sigma}_n(u,w)\log n}\right) + o_P\left(\frac{\|\ell(w)\|}{\sqrt{n}\widehat{\sigma}_n(u,w)}\right).
$$

To show that the last two terms are $o_P(1)$, note that under conditions S.2 and S.3, $\widehat{\sigma}_n(u,w) \gtrsim_P \|\ell(w)\|/\sqrt{n}$ since $\widehat{\sigma}_n(u,w) = (1 + o_P(1))\sigma_n(u,w)$ by Theorem 7.

Finally, for any $\epsilon > 0$, $E[(\ell(w)'J_m^{-1}(u)Z_i)^2/\sigma_n^2(u,w)1\{|\ell(w)'J_m^{-1}(u)Z_i| \geq \epsilon\sqrt{n}\sigma_n(u,w)\}] \rightarrow 0$ under our growth conditions. Thus, by Lindeberg-Feller central limit theorem and $\widehat{\sigma}_n(u,w) = (1 + o_P(1))\sigma_n(u,w)$, it follows that

$$
\frac{\ell(w)'J_m^{-1}(u)\mathbb{U}_n(u)}{\sqrt{n}\widehat{\sigma}_n(u,w)} \rightarrow_d N(0,1)
$$

and the first result follows.

The remaining results for $t_n^*(u,w)$ follow similarly under the corresponding conditions. $\square$

APPENDIX E. PROOFS OF THEOREMS 10-12

Lemma 5 (Entropy Bound). *Let $\mathcal{W} \subset \mathbb{R}^d$ be a bounded set, and $\ell : \mathcal{W} \to \mathbb{R}^m$ be a mapping in $\mathcal{W}$, such that for $\xi_\theta(m)$ and $\xi_\theta^L(m) \geq 1$,*

$$
\|\ell(w)\| \leq \xi_\theta(m) \quad and \quad \|\ell(w) - \ell(\tilde{w})\| \leq \xi_\theta^L(m)\|w - \tilde{w}\|, \quad for \ all \ w, \tilde{w} \in \mathcal{W}.
$$

Moreover, assume that $\|J_m(u) - J_m(u')\| \leq \xi(m)|u - u'|$ for $u, u' \in \mathcal{U}$ in the operator norm, $\xi(m) \geq 1$, and let $\mu = \sup_{u \in \mathcal{U}} \|J_m(u)^{-1}\| > 0$. Let $\mathcal{L}_m$ be the class of functions

$$
\mathcal{L}_m = \{f_{u,w}(g) = \ell(w)'J_m^{-1}(u)g : \ u \in \mathcal{U}, \ w \in \mathcal{W}\} \cup \{f(g) = 1\}
$$

where $g \in B(0, \zeta_m)$, and let F denote the envelope of $\mathcal{L}_m$, i.e. $F = \sup_{f \in \mathcal{L}_m} |f|$. Then, for any $\varepsilon < \varepsilon_0$, the uniform entropy of $\mathcal{L}_m$ is bounded by

$$\sup_Q N(\varepsilon \|F\|_{Q,2}, \mathcal{L}_m, L_2(Q)) \leq \left(\frac{\varepsilon_0 + \mathrm{diam}(\mathcal{U} \times \mathcal{W})K}{\varepsilon} \right)^{d+1}$$

where $K := \mu \xi_\theta^L \zeta_m + \mu^2 \xi_\theta(m) \xi(m) \zeta_m$.

Proof. The uniform entropy bound is based on the proof of Theorem 2 of Andrews [1]. We first exclude the function constant equal to 1 and compute the Lipschitz constant associated with the other functions in $\mathcal{L}_m$ as follows

$$
\begin{aligned}
|\ell(w)' J_m^{-1}(u)g - \ell(\tilde{w})' J_m^{-1}(\tilde{u})g| &= |(\ell(w) - \ell(\tilde{w}))' J_m^{-1}(u)g + \ell(\tilde{w})'(J_m^{-1}(u) - J_m^{-1}(\tilde{u}))g| \\
&= |(\ell(w) - \ell(\tilde{w}))' J_m^{-1}(u)g| + \\
&\quad + |\ell(\tilde{w})' J_m^{-1}(\tilde{u})(J_m(\tilde{u}) - J_m(u)) J_m^{-1}(u)g| \\
&\leq \xi_\theta^L(m) \|w - \tilde{w}\| \|J_m^{-1}(u)g\| + \\
&\quad + \|\ell(\tilde{w})' J_m^{-1}(\tilde{u})\| \, \xi(m) |\tilde{u} - u| \, \|J_m^{-1}(u)g\| \\
&\leq K(\|w - \tilde{w}\| + |\tilde{u} - u|)
\end{aligned}
$$

where $K := \mu \xi_\theta^L \zeta_m + \mu^2 \xi_\theta(m) \xi(m) \zeta_m$.

Consider functions $f_{\tilde{u}_j, \tilde{w}_j}$, where $(\tilde{u}_j, \tilde{w}_j)$'s are points at the centers of disjoint cubes of diameter $\varepsilon \|F\|_{Q,2}/K$ whose union covers $\mathcal{U} \times \mathcal{W}$. Thus, for any $(u, w) \in \mathcal{U} \times \mathcal{W}$

$$\min_{\tilde{u}_j, \tilde{w}_j} \|f_{u,w} - f_{\tilde{u}_j, \tilde{w}_j}\|_{Q,2} \leq K \min_{\tilde{u}_j, \tilde{w}_j} (\|w - \tilde{w}_j\| + |\tilde{u}_j - u|) \leq \varepsilon \|F\|_{Q,2}.$$

Adding (if necessary) the constant function in to the cover, for any measure Q we obtain

$$N(\varepsilon \|F\|_{Q,2}, \mathcal{L}_m, L_2(Q)) \leq 1 + (\mathrm{diam}(\mathcal{U} \times \mathcal{W})K/\varepsilon \|F\|_{Q,2})^{d+1}.$$

The result follows by noting that $\varepsilon < \varepsilon_0$ and that $\|F\|_{Q,2} \geq 1$ since $\mathcal{L}_m$ contains the function constant equal to 1. $\square$

Lemma 6. *Let $\mathcal{F}$ and $\mathcal{G}$ be two classes of functions with envelopes F and G respectively. Then the entropy of $\mathcal{FG} = \{fg : f \in \mathcal{F}, \ g \in \mathcal{G}\}$ satisfies*

$$\sup_Q N(\varepsilon \|FG\|_{Q,2}, \mathcal{FG}, L_2(Q)) \leq \sup_Q N((\varepsilon/2) \|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \sup_Q N((\varepsilon/2) \|G\|_{Q,2}, \mathcal{G}, L_2(Q)).$$

Proof. The proof is similar to Theorem 3 of Andrews [1]. For any measure Q we denote $c_F = \|F\|_{Q,2}^2$ and $c_G = \|G\|_{Q,2}^2$. Note for any measurable set A, $Q_F(A) = \int_A F^2(x) dQ(x)/c_F$ and $Q_G(A) = \int_A G^2(x) dQ(x)/c_G$ are also measures. Let

$$K = \sup_{Q_F} N(\varepsilon \|G\|_{Q_F,2}, \mathcal{G}, L_2(Q_F)) = \sup_Q N(\varepsilon \|G\|_{Q,2}, \mathcal{G}, L_2(Q)) \text{ and}$$

$$L = \sup_{Q_G} N(\varepsilon\|F\|_{Q_G,2}, \mathcal{F}, L_2(Q_G)) = \sup_{Q} N(\varepsilon\|G\|_{Q,2}, \mathcal{G}, L_2(Q)).$$

Let $g_1, \ldots, g_K$ and $f_1, \ldots, f_L$ denote functions in $\mathcal{G}$ and $\mathcal{F}$ used to build a cover of $\mathcal{G}$ and $\mathcal{F}$ of cubes with diameter $\varepsilon\|FG\|_{Q,2}$. Since $F \geq |f|$ and $G \geq |g|$ we have

$$
\begin{aligned}
\min_{\ell \leq L, k \leq K} \|fg - f_\ell g_k\|_{Q,2} &\leq \min_{\ell \leq L, k \leq K} \|f(g - g_k)\|_{Q,2} + \|g_k(f - f_\ell)\|_{Q,2} \\
&\leq \min_{k \leq K} \|F(g - g_k)\|_{Q,2} + \min_{\ell \leq L} \|G(f - f_\ell)\|_{Q,2} \\
&= \min_{k \leq K} \|(g - g_k)\|_{Q_F,2} + \min_{\ell \leq L} \|(f - f_\ell)\|_{Q_G,2} \\
&\leq \varepsilon\|G\|_{Q_F,2} + \varepsilon\|F\|_{Q_G,2} = 2\varepsilon\|FG\|_{Q,2}.
\end{aligned}
$$

Therefore, by taking pairwise products of $g_1, \ldots, g_K$ and $f_1, \ldots, f_L$ to create a net we have

$$N(\varepsilon\|FG\|_{Q,2}, \mathcal{F}\mathcal{G}, L_2(Q)) \leq N((\varepsilon/2)\|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \, N((\varepsilon/2)\|G\|_{Q,2}, \mathcal{G}, L_2(Q)).$$

The result follows by taking the supremum over Q on both sides. $\qquad\square$

Proof of Theorem 10. By the triangle inequality

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

where the second term satisfies $\sup_{(u,w)\in I} |\, r_n(u,w)/\|\ell(w)\| \,| = o(n^{-1/2}\log^{-1} n)$ by condition U.1.

By Theorem 2, the first term is bounded uniformly over I by

$$|\ell(w)'(\widehat{\beta}(u) - \beta(u))| \lesssim_P \frac{|\ell(w)'J_m^{-1}(u)\mathbb{U}_n(u)|}{\sqrt{n}} + o_P(\xi_\theta(m,I)/[\sqrt{n}\log n]) \qquad \text{(E.34)}$$

since $\|\ell(w)\| \leq \xi_\theta(m,I)$ by U.2 and the remainder term of the linear representation in Theorem 2 satisfies $\sup_{u\in\mathcal{U}} \|r_n(u)\| = o_P(1/\log n)$.

Given $Z_i = Z(X_i)$ consider the classes of functions

$$\mathcal{F}_m = \{\ell(w)'J_m^{-1}(u)Z_i \, : \, (u,w)\in I\}, \;\; \mathcal{G} = \{(1\{U_i \leq u\} - u) : u \in \mathcal{U}\} \text{ and } \mathcal{L}_m = \mathcal{F}_m \cup \{f(Z_i) = 1\}.$$

By Lemma 5 and Lemma 6 the entropy of the class of function $\mathcal{L}_m\mathcal{G} = \{fg : f \in \mathcal{L}_m,\, g \in \mathcal{G}\}$ satisfies, for $K \lesssim (\xi_\theta(m,I) \vee 1)(\xi_\theta^L(m,I) \vee 1)(\zeta_m \vee 1)$,

$$\sup_{Q} \log N(\varepsilon\|F\|_{2,Q}, \mathcal{L}_m\mathcal{G}, L_2(Q)) \lesssim (d+1)\log[(1+\text{diam}(I)K)/\varepsilon] \lesssim \log(n/\varepsilon)$$

by our assumptions. Noting that under S.3

$$\sup_{(u,w)\in I} \sqrt{u(1-u)\ell(w)'J_m^{-1}(u)\Sigma_m J_m^{-1}(u)\ell(w)} \lesssim \xi_\theta(m,I),$$

Lemma 16 applied to $\mathcal{L}_m\mathcal{G}$ with $J(m) = \sqrt{d\log n}$, $M(m,n) \lesssim 1 \vee \xi_\theta(m,I)\zeta_m$ yields

$$\sup_{(u,w)\in I} |\ell(w)'J_m(u)^{-1}\mathbb{U}_n(u)| \lesssim_P \sqrt{d\log n}\left(1 \vee \xi_\theta^2(m,I) + \frac{d\log(n)\,(1 \vee \xi_\theta^2(m,I)\,\zeta_m^2)\log n}{n}\right)^{1/2}\log^{1/2} n$$

where the second term in the parenthesis is negligible compared to the first under our growth conditions. The result follows using this bound into (E.34).

$\square$

Proof of Theorem 11. Under the conditions of Theorems 2 and 3, $\sup_{u\in\mathcal{U}}|\text{maxeig}(\widehat{J}_m^{-1}(u) - J_m^{-1}(u))| \lesssim_P o(1/\sqrt{m}\log^{3/2} n)$, and $\widehat{\sigma}_n(u,w) = (1 + o_P(1/\sqrt{m}\log^{3/2} n))\sigma_n(u,w)$ uniformly in $(u,w) \in I$. Note that $\widehat{\sigma}_n(u,w) \gtrsim_P \|\ell(w)\|/\sqrt{n}$. Next define $\bar{t}_n^*(u,w)$ as follows:

For pivotal and gradient bootstrap coupling: $\qquad \bar{t}_n^*(u,w) = \dfrac{\ell(w)'J_m^{-1}(u)\mathbb{U}_n^*(u)/\sqrt{n}}{\sigma_n(u,w)}$;

For Gaussian and weighted bootstrap coupling: $\quad \bar{t}_n^*(u,w) = \dfrac{\ell(w)'J_m^{-1}(u)G_n(u)/\sqrt{n}}{\sigma_n(u,w)}$.

Note that $\bar{t}_n^*$ is independent of the data conditional on the regressor sequence $\mathcal{Z}_n = (Z_1,...,Z_n)$, unlike t_n^* which has some dependence on the data through various estimated quantities.

For the case of pivotal and gradient bootstrap couplings by Theorem 2 and condition U

$$t_n(u,w) =_d \bar{t}_n^*(u,w) + o_P(1/\log n) \quad \text{in } \ell^\infty(I).$$

Moreover, for the case of Gaussian and weighted bootstrap couplings under the conditions of Theorem 5

$$t_n(u,n) =_d \bar{t}_n^*(u,w) + o_P(1/\log n) \quad \text{in } \ell^\infty(I).$$

Finally, under the growth conditions

$$\bar{t}_n^*(u,w) = t_n^*(u,w) + o_P(1/\log n) \quad \text{in } \ell^\infty(I).$$

Thus, it follows that uniformly in $(u,w) \in I$

$$t_n(u,w) =_d t_n^*(u,w) + o_P(1/\log n) = t_n(u,w) + o_P(1/\log n)$$

and the result follows.

$\square$

Proof of Theorem 12. Let $\varepsilon_n = 1/\log n$, and δ_n such that $\delta_n \log^{1/2} n \to 0$, and $\delta_n/\varepsilon_n \to \infty$.

Step 1. By the proof of Theorem 11 there is an approximation $\|\bar{t}_n^*\|_I = \sup_{(u,w)\in I} |\bar{t}_n^*(u,w)|$ to $\|t_n^*\|_I = \sup_{(u,w)\in I} |t_n^*(u,w)|$, which does not depend on the data conditional on the regressor sequence $\mathscr{Z}_n = (Z_1, ..., Z_n)$, such that

$$P(|\ \|\bar{t}_n^*\|_I - \|t_n^*\|_I\ | \leq \varepsilon_n) = 1 - o(1).$$

Now let

$$k_n(1-\alpha) := (1-\alpha)-\ \text{quantile of}\ \|t_n^*\|_I,\ \text{conditional on}\ \mathcal{D}_n,$$

and let

$$\kappa_n(1-\alpha) := (1-\alpha)-\ \text{quantile of}\ \|\bar{t}_n^*\|_I,\ \text{conditional on}\ \mathcal{D}_n.$$

Note that since $\|\bar{t}_n^*\|_I$ is conditionally independent of $Y_1, \dots, Y_n$,

$$\kappa_n(1-\alpha) = (1-\alpha)-\ \text{quantile of}\ \|\bar{t}_n^*\|_I,\ \text{conditional on}\ \mathscr{Z}_n.$$

Then applying Lemma 7 to $\|\bar{t}_n^*\|_I$ and $\|t_n^*\|_I$, we get that for some $\nu_n \searrow 0$

$$P[\kappa_n(p) \geq k_n(p-\nu_n) - \varepsilon_n\ \text{and}\ k_n(p) \geq \kappa_n(p-\nu_n) - \varepsilon_n] = 1 - o(1).$$

Step 2. Claim (1) now follows by noting that

$$
\begin{aligned}
P\{\|t_n\|_I > k_n(1-\alpha) + \delta_n\} &\leq& P\{\|t_n\|_I > \kappa_n(1-\alpha-\nu_n) - \varepsilon_n + \delta_n\} + o(1)\\
&\leq& P\{\|t_n^*\|_I > \kappa_n(1-\alpha-\nu_n) - 2\varepsilon_n + \delta_n\} + o(1)\\
&\leq& P\{\|t_n^*\|_I > k_n(1-\alpha-2\nu_n) - 3\varepsilon_n + \delta_n\} + o(1)\\
&\leq& P\{\|t_n^*\|_I > k_n(1-\alpha-2\nu_n)\} + o(1)\\
&=& E_P[P\{\|t_n^*\|_I > k_n(1-\alpha-2\nu_n)|\mathcal{D}_n\}] + o(1)\\
&\leq& E_P[\alpha+2\nu_n] + o(1) = \alpha + o(1).
\end{aligned}
$$

Claim (2) follows from the equivalence of the event $\{\theta(u,w) \in [i(u,w), \ddot{i}(u,w)],\ \text{for all}\ (u,w) \in I\}$ and the event $\{\|t_n\|_I \leq c_n(1-\alpha)\}$.

To prove Claim (3) note that $\widehat{\sigma}_n(u,w) = (1+o_P(1))\sigma_n(u,w)$ uniformly in $(u,w) \in I$ under the conditions of Theorems 2 and 7. Moreover, $c_n(1-\alpha) = k_n(1-\alpha)(1+o_P(1))$ because $1/k_n(1-\alpha) \lesssim_P 1$ and $\delta_n \to 0$. Combining these relations the result follows.

Claim (4) follows from Claim (1) and from the following lower bound.

By Lemma 7, we get that for some $\nu_n \searrow 0$

$$P[\kappa_n(p+\nu_n) + \varepsilon_n \geq k_n(p)\ \text{and}\ k_n(p+\nu_n) + \varepsilon_n \geq \kappa_n(p)] = 1 - o(1).$$

Then

$$
\begin{aligned}
P\{\|t_n\|_I \geq k_n(1-\alpha) + \delta_n\} \;\geq\;& P\{\|t_n\|_I \geq \kappa_n(1-\alpha+\nu_n) + \varepsilon_n + \delta_n\} - o(1) \\
\geq\;& P\{\|t_n^*\|_I \geq \kappa_n(1-\alpha+\nu_n) + 2\varepsilon_n + \delta_n\} - o(1) \\
\geq\;& P\{\|t_n^*\|_I \geq k_n(1-\alpha+2\nu_n) + 3\varepsilon_n + \delta_n\} - o(1) \\
\geq\;& P\{\|t_n^*\|_I \geq k_n(1-\alpha+2\nu_n) + 2\delta_n\} - o(1) \\
\geq\;& E[P\{\|t_n^*\|_I \geq k_n(1-\alpha+2\nu_n) + 2\delta_n|\mathcal{D}_n\}] - o(1) \\
=\;& \alpha - 2\nu_n - o(1) = \alpha - o(1),
\end{aligned}
$$

where we used the anti-concentration property in the last step.

$\square$

We use this lemma in the proof of Theorem 12.

Lemma 7 (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 $|X_n - Y_n| = o_P(\varepsilon)$, then for some $\nu_n \searrow 0$ with probability converging to one*

$$
F_{X_n}^{-1}(p|\mathcal{D}_n) \leq F_{Y_n}^{-1}(p+\nu_n|\mathcal{D}_n) + \varepsilon \text{ and } F_{Y_n}^{-1}(p|\mathcal{D}_n) \leq F_{X_n}^{-1}(p+\nu_n|\mathcal{D}_n) + \varepsilon, \forall p \in (\nu_n, 1-\nu_n).
$$

Proof. We have that for some $\nu_n \searrow 0$, $P\{|X_n - Y_n| > \varepsilon\} = o(\nu_n)$. This implies that $P[P\{|X_n - Y_n| > \varepsilon|\mathcal{D}_n\} \leq \nu_n] \to 1$, i.e. there is a set Ω_n such that $P(\Omega_n) \to 1$ and $P\{|X_n - Y_n| > \varepsilon|\mathcal{D}_n\} \leq \nu_n$ for all $\mathcal{D}_n \in \Omega_n$. So, for all $\mathcal{D}_n \in \Omega_n$

$$
F_{X_n}(x|\mathcal{D}_n) \geq F_{Y_n+\varepsilon}(x|\mathcal{D}_n) - \nu_n \text{ and } F_{Y_n}(x|\mathcal{D}_n) \geq F_{X_n+\varepsilon}(x|\mathcal{D}_n) - \nu_n, \forall x \in \mathbb{R},
$$

which implies the inequality stated in the lemma, by definition of the conditional quantile function and equivariance of quantiles to location shifts. $\square$

APPENDIX F. A LEMMA ON STRONG APPROXIMATION OF AN EMPIRICAL PROCESS OF AN INCREASING DIMENSION BY A GAUSSIAN PROCESS

Lemma 8. *(Approximation of a Sequence of Empirical Processes of Increasing Dimension by a Sequence of Gaussian Processes) Consider the empirical process $\mathbb{U}_n$ in $[\ell^\infty(\mathcal{U})]^m$, $\mathcal{U} \subseteq (0,1)$, conditional on $Z_i \in \mathbb{R}^m$, $i = 1, \ldots, n$, defined by*

$$
\mathbb{U}_n(u) = \mathbb{G}_n\left(v_i Z_i \psi_i(u)\right), \quad \psi_i(u) = u - 1\{U_i \leq u\},
$$

where $U_i, i = 1, \ldots, n$, is an i.i.d. sequence of standard uniform random variables, $v_i, i = 1, \ldots, n$, is an i.i.d. sequence of real valued random variables such that $E[v_i^2] = 1$, $E[|v_i|^3] \lesssim 1$, and $\max_{1 \le i \le n} |v_i| \lesssim_P \log n$. Suppose that $Z_i, i = 1, \ldots, n$, are such that

$$\sup_{\|\alpha\| \le 1} \mathbb{E}_n \left[(\alpha' Z_i)^2 \right] \lesssim 1, \quad \max_{1 \le i \le n} \|Z_i\| \lesssim \zeta_m, \quad and \quad m^7 \zeta_m^6 \log^{22} n = o(n).$$

There exists a sequence of zero-mean Gaussian processes G_n with a.s. continuous paths, that has the same covariance functions as $\mathbb{U}_n$, conditional on $Z_1, \ldots, Z_n$, namely,

$$E[G_n(u) G_n(u')'] = E[\mathbb{U}_n(u) \mathbb{U}_n(u')'] = \mathbb{E}_n[Z_i Z_i'](u \wedge u' - uu'), \quad for \ all \ u \ and \ u' \in \mathcal{U},$$

and that approximates the empirical process $\mathbb{U}_n$, namely,

$$\sup_{u \in \mathcal{U}} \|\mathbb{U}_n(u) - G_n(u)\| \lesssim_P o\left(\frac{1}{\log n}\right).$$

Proof. The proof is based on the use of maximal inequalities and Yurinskii's coupling. Throughout the proof all the probability statements are conditional on $Z_1, \ldots, Z_n$.

We define the sequence of projections $\pi_j : \mathcal{U} \to \mathcal{U}$, $j = 0, 1, 2, \ldots, \infty$ by $\pi_j(u) = 2^{k-1}/2^j$ if $u \in ((2^k - 2)/2^j, 2^k/2^j), k = 1, \ldots, j$, and $\pi_j(u) = u$ if $u = 0$ or 1. 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 2^j steps, we shall identify the process $G \circ \pi_j$ with a random vector $G \circ \pi_j$ in $\mathbb{R}^{2^j m}$, when convenient. Analogously, given a random vector W in $\mathbb{R}^{2^j m}$ we identify it with a process W in $[\ell^\infty(\mathcal{U})]^m$, whose paths are step functions with 2^j steps.

The following relations proven below:

(1) (Finite-Dimensional Approximation)

$$r_1 = \sup_{u \in \mathcal{U}} \|\mathbb{U}_n(u) - \mathbb{U}_n \circ \pi_j(u)\| \lesssim_P o\left(\frac{1}{\log n}\right);$$

(2) (Coupling with a Normal Vector) there exists $\mathcal{N}_{nj} =_d N(0, \operatorname{var}[\mathbb{U}_n \circ \pi_j])$ such that,

$$r_2 = \|\mathcal{N}_{nj} - \mathbb{U}_n \circ \pi_j\|_2 \lesssim_P o\left(\frac{1}{\log n}\right);$$

(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)

$$r_3 = \sup_{u \in \mathcal{U}} \|G_n(u) - G_n \circ \pi_j(u)\| \lesssim_P o\left(\frac{1}{\log n}\right).$$

The result then follows from the triangle inequality

$$\sup_{u \in \mathcal{U}} \|\mathbb{U}_n(u) - G_n(u)\| \le r_1 + r_2 + r_3.$$

Relation (1) follows from

$$r_1 = \sup_{u \in \mathcal{U}} \|\mathbb{U}_n(u) - \mathbb{U}_n \circ \pi_j(u)\| \le \sup_{|u-u'| \le 2^{-j}} \|\mathbb{U}_n(u) - \mathbb{U}_n(u')\|$$

$$\lesssim_P \sqrt{2^{-j} m \log n} + \sqrt{\frac{m^2 \zeta_m^2 \log^4 n}{n}} \lesssim_P o(1/\log n),$$

where the last inequality holds by Lemma 9, and the final rate follows by choosing here

$$2^j = (m \log^3 n)\ell_n \text{ for some } \ell_n \to \infty \text{ slowly enough.}$$

Relation (2) follows from the use of Yurinskii's coupling (Pollard [31], Chapter 10, Theorem 10): Let $\xi_1, \ldots, \xi_n$ be independent p-vectors with $E[\xi_i] = 0$ for each i, and with $\kappa := \sum_i E\left[\|\xi_i\|^3\right]$ finite. Let $S = \xi_1 + \cdots + \xi_n$. For each $\delta > 0$ there exists a random vector T with a $N(0, \text{var}(S))$ distribution such that

$$P\{\|S - T\| > 3\delta\} \le C_0 B \left(1 + \frac{|\log(1/B)|}{p}\right) \text{ where } B := \kappa p \delta^{-3},$$

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 $2^j = m \log^3 n$,

$$P\left\{\left\|\frac{\sum_{i=1}^n \xi_i}{\sqrt{n}} - \mathcal{N}_{n,j}\right\| \ge 3\delta\right\} \lesssim \frac{n 2^{3j/2} \zeta_m^3 2^j m}{(\delta\sqrt{n})^3} = \frac{2^{5j/2} m \zeta_m^3}{\delta^3 n^{1/2}} \to 0$$

by setting $\delta = \left(\frac{2^{5j} m^2 \zeta_m^6}{n} \log n\right)^{1/6}$. This verifies relation (2) with

$$r_2 \lesssim_P \left(\frac{2^{5j} m^2 \zeta_m^6}{n} \log n\right)^{1/6} = o(1/\log n),$$

provided j is chosen as above.

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 11.

Relation (4) follows from

$$r_3 = \sup_{u \in \mathcal{U}} \|G_n(u) - G_n \circ \pi_j(u)\| \leq \sup_{|u-u'| \leq 2^{-j}} \|G_n(u) - G_n(u')\|$$
$$\lesssim_P \sqrt{2^{-j} m \log n} \lesssim_P o(1/\log n),$$

where the last inequality holds by Lemma 10 since by assumption of this lemma,

$$\sup_{\|\alpha\| \leq 1} E\mathbb{E}_n\left[v_i^2(\alpha' Z_i)^2\right] = \sup_{\|\alpha\| \leq 1} \mathbb{E}_n\left[(\alpha' Z_i)^2\right] \lesssim 1$$

and the rate follows from setting j as above.

Note that putting bounds together we also get an explicit bound on the approximation error:

$$\sup_{u \in \mathcal{U}} \|\mathbb{U}_n(u) - G_n(u)\| \lesssim_P \sqrt{2^{-j} m \log n} + \sqrt{\frac{m^2 \zeta_m^2 \log^4 n}{n}} + \left(\frac{2^{5j} m^2 \zeta_m^6 \log n}{n}\right)^{1/6},$$

$\square$

Next we establish the auxiliary relations (1) and (4) appearing in the preceding proof.

Lemma 9 (Finite-Dimensional Approximation). *Let $Z_1, \ldots, Z_n \in \mathbb{R}^m$ be such that $\max_{i \leq n} \|Z_i\| \lesssim \zeta_m$, and $\varphi = \sup_{\|\alpha\| \leq 1} \mathbb{E}_n[(\alpha' Z_i)^2]$, let v_i be i.i.d. random variables such that $E[v_i^2] = 1$ and $\max_{1 \leq i \leq n} |v_i| \lesssim_P \log n$, and let $\psi_i(u) = u - 1\{U_i \leq u\}$, $i = 1, \ldots, n$, where $U_1, \ldots, U_n$ are i.i.d. Uniform$(0,1)$ random variables. Then, for $\gamma > 0$, and $\mathbb{U}_n(u) = \mathbb{G}_n(v_i Z_i \psi_i(u))$,*

$$\sup_{\|\alpha\| \leq 1, |u-u'| \leq \gamma} \left|\alpha'\left(\mathbb{U}_n(u) - \mathbb{U}_n(u')\right)\right| \lesssim_P \sqrt{\gamma \varphi m \log n} + \sqrt{\frac{m^2 \zeta_m^2 \log^4 n}{n}}.$$

Proof. For notational convenience let $A_n := \sqrt{\frac{m^2 \zeta_m^2 \log^4 n}{n}}$. Using the second maximal inequality of Lemma 16 with $M(m,n) = \zeta_m \log n$

$$\begin{aligned}\varepsilon(m,n,\gamma) &= \sup_{\|\alpha\| \leq 1, |u-u'| \leq \gamma} \left|\alpha'\left(\mathbb{U}_n(u) - \mathbb{U}_n(u')\right)\right| \\ &\lesssim_P \sqrt{m \log n} \sup_{\|\alpha\| \leq 1, |u-u'| \leq \gamma} \sqrt{E\mathbb{E}_n[v_i^2(\alpha' Z_i)^2(\psi_i(u) - \psi_i(u'))^2]} + A_n.\end{aligned}$$

By the independence between Z_i, v_i and U_i, and $E[v_i^2] = 1$,

$$\varepsilon(m,n,\gamma) \lesssim_P \sqrt{\varphi m \log n} \sup_{|u-u'| \leq \gamma} \sqrt{E[(\psi_i(u) - \psi_i(u'))^2]} + A_n.$$

Since $U_i \sim \text{Uniform}(0,1)$ we have $(\psi_i(u) - \psi_i(u'))^2 =_d (|u - u'| - 1\{U_i \le |u - u'|\})^2$. Thus, since $|u - u'| \le \gamma$

$$\varepsilon(m, n, \gamma) \lesssim_P \sqrt{\varphi m \log n} \sqrt{\gamma(1 - \gamma)} + A_n.$$

$\square$

Lemma 10 (Infinite-Dimensional Approximation). *Let* $G_n : \mathcal{U} \to \mathbb{R}^m$ *be a zero-mean Gaussian process whose covariance structure conditional on* $Z_1, \ldots, Z_n$ *is given by*

$$E\left[G_n(u)G_n(u')'\right] = \mathbb{E}_n[Z_i Z_i'](u \vee u' - uu')$$

for any $u, u' \in \mathcal{U} \subset (0,1)$*, where* $Z_i \in \mathbb{R}^m$*,* $i = 1, \ldots, n$*. Then, for any* $\gamma > 0$ *we have*

$$\sup_{|u-u'| \le \gamma} \|G_n(u) - G_n(u')\| \lesssim_P \sqrt{\varphi \gamma m \log m}$$

where $\varphi = \sup_{\|\alpha\| \le 1} \mathbb{E}_n[(\alpha' Z_i)^2]$.

Proof. We will use the following maximal inequality for Gaussian processes (Proposition A.2.7 [37]) Let X be a separable zero-mean Gaussian process indexed by a set T. Suppose that for some $K > \sigma(X) = \sup_{t \in T} \sigma(X_t)$, $0 < \epsilon_0 \le \sigma(X)$, we have

$$N(\varepsilon, T, \rho) \le \left(\frac{K}{\varepsilon}\right)^V, \text{ for } 0 < \varepsilon < \epsilon_0,$$

where $N(\varepsilon, T, \rho)$ 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 $\lambda \ge \sigma^2(X)(1 + \sqrt{V})/\epsilon_0$

$$P\left(\sup_{t \in T} X_t > \lambda\right) \le \left(\frac{DK\lambda}{\sqrt{V}\sigma^2(X)}\right)^V \bar{\Phi}(\lambda/\sigma(X)),$$

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 $X_n : S^{m-1} \times \mathcal{U} \times \mathcal{U} \to \mathbb{R}$ defined as

$$X_{n,t} = \alpha'(G_n(u) - G_n(u')), \quad t = (\alpha, u, u'), \ \alpha \in S^{m-1}, \ |u - u'| \le \gamma.$$

It follows that $\sup_{t \in T} X_{n,t} = \sup_{|u-u'| \le \gamma} \|G_n(u) - G_n(u')\|$.

For the process X_n we have:

$$\sigma(X_n) \le \sqrt{\gamma \sup_{\|\alpha\| \le 1} \mathbb{E}_n[(\alpha' Z_i)^2]}, \quad K \lesssim \sqrt{\sup_{\|\alpha\| \le 1} \mathbb{E}_n[(\alpha' Z_i)^2]}, \quad \text{and } V \lesssim m.$$

Therefore the result follows by setting $\lambda \simeq \sqrt{\gamma m \log m \sup_{\|\alpha\| \le 1} \mathbb{E}_n[(\alpha' Z_i)^2]}$. $\square$

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 2^j steps, we shall identify the process $G \circ \pi_j$ with a random vector $G \circ \pi_j$ in $\mathbb{R}^{2^j m}$, when convenient. Analogously, given a random vector W in $\mathbb{R}^{2^j m}$ we identify it with a process W in $[\ell^\infty(\mathcal{U})]^m$, whose paths are step functions with 2^j steps.

Lemma 11. *(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 $\ell > j$. Then $\tilde{\mathcal{N}}_j = \tilde{G} \circ \pi_j$ is a subvector of $\tilde{G} \circ \pi_\ell = \tilde{\mathcal{N}}_\ell$. Thus, denote the remaining components of $\tilde{\mathcal{N}}_\ell$ as $\tilde{\mathcal{N}}_{\ell \setminus j}$. 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

$$\mathcal{N}_j \text{ and } \mathcal{N}_{\ell \setminus j},$$

arranged in appropriate order, namely that $\mathcal{N}_\ell \circ \pi_j = \mathcal{N}_j$, where

$$\mathcal{N}_{\ell \setminus j} = \Sigma_{\ell \setminus j, j} \Sigma_{j,j}^{-1} \mathcal{N}_j + \eta_j,$$

where $\eta_j \perp \mathcal{N}_j$ and $\eta_j =_d N(0, \Sigma_{\ell \setminus j, \ell \setminus j} - \Sigma_{\ell \setminus j, j} \Sigma_{j,j}^{-1} \Sigma_{j, \ell \setminus j})$, where

$$\begin{pmatrix} \Sigma_{j,j} & \Sigma_{\ell \setminus j, j} \\ \Sigma_{j, \ell \setminus j} & \Sigma_{\ell \setminus j, \ell \setminus j} \end{pmatrix} := \mathrm{var} \begin{pmatrix} \tilde{N}_j \\ \tilde{N}_{\ell \setminus j} \end{pmatrix}.$$

We then identify the vector $\mathcal{N}_\ell$ with a process $\mathcal{N}_\ell$ in $\ell^\infty(\mathcal{U})$, and define the pointwise limit G of this process as

$$G(u) := \lim_{\ell \to \infty} \mathcal{N}_\ell(u) \text{ for each } u \in \mathcal{U}_0,$$

where $\mathcal{U}_0 = \cup_{j=1}^\infty \cup_{k=1}^{2^j} u_{kj}$ is a countable dense subset of $\mathcal{U}$. 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.

By construction $G_\ell = G_\ell \circ \pi_\ell =_d \tilde{G} \circ \pi_\ell$. Therefore, for each $\epsilon > 0$, there exists $\eta(\epsilon) > 0$ small enough such that

$$
P \left(\sup_{u,u' \in \mathcal{U}_0 : |u-u'| \leq \eta(\epsilon)} \|G(u) - G(u')\| \geq \epsilon \right) \leq P \left(\sup_{|u-u| \leq \eta(\epsilon)} \sup_k \|G \circ \pi_k(u) - G \circ \pi_k(u')\| \geq \epsilon \right)
$$

$$
\leq P \left(\sup_{|u-u| \leq \eta(\epsilon)} \sup_k \|\tilde{G} \circ \pi_k(u) - \tilde{G} \circ \pi_k(u')\| \geq \epsilon \right)
$$

$$
\leq P \left(\sup_{|u-u| \leq \eta(\epsilon)} \|\tilde{G}(u) - \tilde{G}(u')\| \geq \epsilon \right) \leq \epsilon,
$$

where the last display is true because $\sup_{|u-u| \leq \eta} \|\tilde{G}(u) - \tilde{G}(u')\| \to 0$ 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 side. Conclude that by the Borel-Cantelli lemma, for almost all $\omega \in \Omega$, $|G(u) - G(u')| \leq 2^{-m}$ for all $|u - u'| \leq \eta(2^{-m})$ for all sufficiently large m. This implies that almost all sample paths are uniformly continuous on $\mathcal{U}_0$, and we can extend the process by continuity to a process $\{G(u), u \in \mathcal{U}\}$ with almost all paths that are uniformly continuous.

In order to show that the law of G is equal to the law of $\tilde{G}$ in $\ell^\infty(\mathcal{U})$, it suffices to demonstrate that

$$
E[g(G)] = E[g(\tilde{G})] \text{ for all } g : [\ell^\infty(\mathcal{U})]^m \to \mathbb{R} : |g(z) - g(\tilde{z})| \leq \sup_{u \in \mathcal{U}} \|z(u) - \tilde{z}(u)\| \wedge 1.
$$

We have that

$$
|E[g(G)] - E[g(\tilde{G})]| \leq |E[g(G \circ \pi_\ell)] - E[g(\tilde{G} \circ \pi_\ell)]| +
$$

$$
+ E \left[\sup_{u \in \mathcal{U}} \|G \circ \pi_\ell(u) - G(u)\| \wedge 1 \right] +
$$

$$
+ E \left[\sup_{u \in \mathcal{U}} \|\tilde{G} \circ \pi_\ell(u) - \tilde{G}(u)\| \wedge 1 \right]
$$

$$
\to 0 \text{ as } \ell \to \infty.
$$

The first term converges to zero by construction, and the second and third terms converge to zero by the dominated convergence theorem and by

$$
G \circ \pi_\ell \to G \text{ and } \tilde{G} \circ \pi_\ell \to \tilde{G} \text{ in } [\ell^\infty(\mathcal{U})]^m \text{ as } \ell \to \infty \text{ a.s.,}
$$

holding due to a.s. uniform continuity of sample paths of G and $\tilde{G}$. $\qquad\square$

APPENDIX G. TECHNICAL LEMMAS ON BOUNDING EMPIRICAL ERRORS

In Appendix G.2 we establish technical results needed for our main results - uniform rates of convergence, uniform linear approximations, and uniform central limit theorem - under high-level conditions. In Appendix G.3 we verify that these conditions are implied by the primitive Condition S stated in Section 2.

G.1. **Some Preliminary Lemmas.**

Lemma 12. *Under the conditions S.2 and S.5, for any $u \in \mathcal{U}$ and $\alpha \in S^{m-1}$*

$$|\alpha'(J_m(u) - \widetilde{J}_m(u))\alpha| \lesssim m^{-\kappa} = o(1),$$

where $\widetilde{J}_m(u) = E[f_{Y|X}(Z'\beta(u)|X)ZZ']$.

Proof of Lemma 12. For any $\alpha \in S^{m-1}$

$$
\begin{aligned}
|\alpha'(J_m(u) - \widetilde{J}_m(u))\alpha| &= E[|f_{Y|X}(Z'\beta(u) + R(X,u)|X) - f_{Y|X}(Z'\beta(u)|X)|(Z'\alpha)^2] \\
&\lesssim \alpha'\Sigma_m\alpha \cdot \overline{f'}m^{-\kappa}.
\end{aligned}
$$

The result follows since Σ_m has bounded eigenvalues, $\kappa > 0$ and $m \to \infty$ as $n \to \infty$. $\qquad\square$

Lemma 13 (Auxiliary Matrix)**.** *Under conditions S.1-S.5, for $u', u \in \mathcal{U}$ we have that uniformly over $z \in \mathcal{Z}$*

$$\left| \frac{z'(J_m^{-1}(u') - J_m^{-1}(u))\mathbb{U}_n(u')}{\sqrt{u(1-u)z'J_m^{-1}(u)\Sigma_m J_m^{-1}(u)z}} \right| \lesssim_P |u - u'|\sqrt{m \log n}.$$

Proof. Recall that $J_m(u) = E\left[f_{Y|X}(Q_{Y|X}(u|X)|X)ZZ' \right]$ for any $u \in \mathcal{U}$. Moreover, under S.1-S.5, we have $\|J_m^{-1}(u')\mathbb{U}_n(u')\| \lesssim_P \sqrt{m \log n}$ uniformly in $u' \in \mathcal{U}$ by Lemma 23 and Corollary 2 of Appendix G.

Using the matrix identity $A^{-1} - B^{-1} = B^{-1}(B - A)A^{-1}$

$$J_m^{-1}(u') - J_m^{-1}(u) = J_m^{-1}(u)(J_m(u) - J_m(u'))J_m^{-1}(u').$$

Moreover, since $|f_{Y|X}(Q_{Y|X}(u|x)|x) - f_{Y|X}(Q_{Y|X}(u'|x)|x)| \le (\bar{f}'/\underline{f})|u - u'|$ by Lemma 14,

$$J_m(u) - J_m(u') \preccurlyeq (\bar{f}'/\underline{f})|u - u'|\Sigma_m, \quad \text{and} \quad J_m(u') - J_m(u) \preccurlyeq (\bar{f}'/\underline{f})|u - u'|\Sigma_m$$

where the inequalities are in the semi-definite positive sense. Using these relations and the definition of $s_n(u, x)$ we obtain

$$\left| \frac{z'(J_m^{-1}(u') - J_m^{-1}(u))\mathbb{U}_n(u')}{\sqrt{u(1-u)z'J_m^{-1}(u)\Sigma_m J_m^{-1}(u)z}} \right| = \left| \frac{z'J_m^{-1}(u)}{s_n(u,x)}(J_m(u) - J_m(u'))J_m^{-1}(u')\mathbb{U}_n(u') \right|$$

$$\lesssim_P \quad (\bar{f}'/\underline{f})|u - u'|\mathrm{maxeig}(\Sigma_m)\sqrt{m \log n}.$$

The result follows since $\bar{f}'$ is bounded above, $\underline{f}$ is bounded from below, and the eigenvalues of Σ_m are bounded above and below by constants uniformly in n by S.2 and S.3. $\qquad\square$

Lemma 14 (Primitive Condition S.2). *Under the condition S.2, there are positive constants c, C, C_1', C_2', C'' such that the conditional quantile functions satisfy the following properties uniformly over $u, u' \in \mathcal{U}$, $x \in \mathcal{X}$,*

(i) *$c|u - u'| \leq |Q_{Y|X}(u|x) - Q_{Y|X}(u'|x)| \leq C|u - u'|$;*
(ii) *$|f_{Y|X}(Q_{Y|X}(u|x)|x) - f_{Y|X}(Q_{Y|X}(u'|x)|x)| \leq C_1'|u - u'|$;*
(iii) *$f_{Y|X}(y|x) \leq 1/c$ and $|f_{Y|X}'(y|x)| \leq C_2'$;*
(iv) *$\left| \frac{d^2}{du^2}Q_{Y|X}(u|x) \right| \leq C''$.*

Proof. Under S.2, $f_{Y|X}(\cdot|x)$ is a differentiable function so that $Q_{Y|X}(\cdot|x)$ is twice differentiable.

To show the first statement note that $\frac{d}{du}Q_{Y|X}(u|x) = \frac{1}{f_{Y|X}(Q_{Y|X}(u|x)|x)}$, by an application of the inverse function theorem. Recall that $\underline{f} = \inf_{x \in \mathcal{X}, u \in \mathcal{U}} f_{Y|X}(Q_{Y|X}(u|x)|x) > 0$, and $\sup_{x \in \mathcal{X}, y \in \mathbb{R}} f_{Y|X}(y|x) \leq \bar{f}$. This proves the first statement for $c = 1/\bar{f}$ and $C = 1/\underline{f}$.

To show the second statement let $\bar{f}' = \sup_{x \in \mathcal{X}} \left| \frac{d}{dy}f_{Y|X}(y|x) \right|$, and

$$C_1' = \sup_{x \in \mathcal{X}, u \in \mathcal{U}} \left| \frac{d}{du}f_{Y|X}(Q_{Y|X}(u|x)|x) \right|$$

$$= \sup_{x \in \mathcal{X}, u \in \mathcal{U}} \left| \frac{d}{dy}f_{Y|X}(y|x) \mid_{y=Q_{Y|X}(u|x)} \frac{d}{du}Q_{Y|X}(u|x) \right| \leq \frac{\bar{f}'}{\underline{f}}.$$

By a Taylor expansion we have

$$|f_{Y|X}(Q_{Y|X}(u|x)|x) - f_{Y|X}(Q_{Y|X}(u'|x)|x)| \leq C_1'|u - u'|.$$

The second part of the third statement follows with $C_2' = \bar{f}'$. The first statement was already shown in the proof of part (i).

For the fourth statement, using the implicit function theorem for second order derivatives

$$\frac{d^2}{du^2}Q_{Y|X}(u|x) = -\frac{d^2}{dy^2}F_{Y|X}(y|x)\Big|_{y=Q_{Y|X}(u|X)} \left(\frac{d}{du}Q_{Y|X}(u|x) \right)^3 = -\frac{f_{Y|X}'(Q_{Y|X}(u|x)|x)}{f_{Y|X}^3(Q_{Y|X}(u|x)|x)}.$$

Thus, the statement holds with $C'' = \bar{f}'/\underline{f}^3$.

Under S.2, we can take c, C, C_1', C_2', and C'' to be fixed positive constants uniformly over n. □

G.2. **Maximal Inequalities.** In this section we derive maximal inequalities that are needed for verifying the preliminary high-level conditions. These inequalities rely mainly on uniform entropy bounds and VC classes of functions. In what follows $\mathcal{F}$ denotes a class of functions whose envelope is F. Recall that for a probability measure Q with $\|F\|_{Q,p} > 0$, $N(\varepsilon\|F\|_{Q,p}, \mathcal{F}, L_p(Q))$ denotes the covering number under the specified metric (i.e., the minimum number of $L_p(Q)$-balls of radius $\varepsilon\|F\|_{Q,p}$ needed to cover $\mathcal{F}$). We refer to Dudley [17] for the details of the definitions.

Suppose that we have the following upper bound on the $L_2(P)$ covering numbers for $\mathcal{F}$:

$$N(\epsilon\|F\|_{P,2}, \mathcal{F}, L_2(P)) \leq n(\epsilon, \mathcal{F}, P) \text{ for each } \epsilon > 0,$$

where $n(\epsilon, \mathcal{F}, P)$ is increasing in $1/\epsilon$, and $\epsilon\sqrt{\log n(\epsilon, \mathcal{F}, P)} \to 0$ as $1/\epsilon \to \infty$ and is decreasing in $1/\epsilon$. Let $\rho(\mathcal{F}, P) := \sup_{f\in\mathcal{F}} \|f\|_{P,2}/\|F\|_{P,2}$. Let us call a threshold function $x : \mathbb{R}^n \mapsto \mathbb{R}$ k-sub-exchangeable if, for any $v, w \in \mathbb{R}^n$ and any vectors $\tilde{v}, \tilde{w}$ created by the pairwise exchange of the components in v with components in w, we have that $x(\tilde{v}) \vee x(\tilde{w}) \geq [x(v) \vee x(w)]/k$. Several functions satisfy this property, in particular $x(v) = \|v\|$ with $k = \sqrt{2}$ and constant functions with $k = 1$.

Lemma 15 (Exponential inequality for separable empirical process). *Consider a separable empirical process* $\mathbb{G}_n(f) = n^{-1/2} \sum_{i=1}^{n} \{f(Z_i) - E[f(Z_i)]\}$ *and the empirical measure* $\mathbb{P}_n$ *for* $Z_1, \ldots, Z_n$, *an underlying independent data sequence. 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

$$\|F\|_{\mathbb{P}_n,2} \int_0^{\rho(\mathcal{F},\mathbb{P}_n)/4} \sqrt{\log n(\epsilon, \mathcal{F}, \mathbb{P}_n)} d\epsilon \leq e_n(\mathcal{F}, \mathbb{P}_n) \text{ and } \sup_{f\in\mathcal{F}} var_{\mathbb{P}} f \leq \frac{\tau}{2}(4kcKe_n(\mathcal{F}, \mathbb{P}_n))^2$$

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 [6], Lemma 18 and note that the proof does not use that Z_i's are i.i.d., only independent which was the requirement of Lemma 17 of [6]. □

The next lemma establishes a new maximal inequality which will be used in the following sections.

Lemma 16. *Suppose for all large m and all $0 < \varepsilon \le \varepsilon_0$*

$$n(\epsilon, \mathcal{F}_m, P) \le (\omega/\varepsilon)^{J(m)^2} \quad and \quad n(\epsilon, \mathcal{F}_m^2, P) \le (\omega/\varepsilon)^{J(m)^2}, \tag{G.35}$$

for some ω such that $\log \omega \lesssim \log n$, and let $F_m = \sup_{f \in \mathcal{F}_m} |f|$ denote the envelope function associated with $\mathcal{F}_m$.

1. *(A Maximal Inequality Based on Entropy and Moments) Then, as n grows we have*

$$\sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f)| \lesssim_P J(m) \left(\sup_{f \in \mathcal{F}_m} E[f^2] + n^{-1/2} J(m) \log^{1/2} n \left(\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f^4] \vee E[f^4] \right)^{1/2} \right)^{1/2} \log^{1/2} n.$$

2. *(A Maximal Inequality Based on Entropy, Moments, and Extremum) Suppose that $F_m \le M(m, n)$ with probability going to 1 as n grows. Then as n grows we have*

$$\sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f)| \lesssim_P J(m) \left(\sup_{f \in \mathcal{F}_m} E[f^2] + n^{-1} J(m)^2 M^2(m, n) \log n \right)^{1/2} \log^{1/2} n.$$

Proof. We divide the proof into steps. Step 1 is the main argument, Step 2 is an application of Lemma 15, and Step 3 contains some auxiliary calculations.

Step 1. (Main Argument) **Proof of Part 1.** By Step 2 below, which invokes Lemma 15,

$$\sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f)| \lesssim_P J(m) \sqrt{\log n} \sup_{f \in \mathcal{F}_m} (\mathbb{E}_n[f^2]^{1/2} \vee E[f^2]^{1/2}). \tag{G.36}$$

We can assume that $\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f^2]^{1/2} \ge \sup_{f \in \mathcal{F}_m} E[f^2]^{1/2}$ throughout the proof otherwise we are done with both bounds.

Again by Step 2 and (G.35) applied to $\mathcal{F}_m^2$, we also have

$$\begin{aligned} \sup_{f \in \mathcal{F}_m} |\mathbb{E}_n[f^2] - E[f^2]| \quad &= \quad n^{-1/2} \sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f^2)| \\ &\lesssim_P \quad n^{-1/2} J(m) \sqrt{\log n} \sup_{f \in \mathcal{F}_m} (\mathbb{E}_n[f^4] \vee E[f^4])^{1/2}. \end{aligned} \tag{G.37}$$

Thus we have

$$\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f^2] \lesssim_P \sup_{f \in \mathcal{F}_m} E[f^2] + n^{-1/2} J(m) \log^{1/2} n \sup_{f \in \mathcal{F}_m} (\mathbb{E}_n[f^4] \vee E[f^4])^{1/2}. \tag{G.38}$$

Therefore, inserting the bounds (G.38) in equation (G.36) yields the result.

Proof of Part 2. One more time we can assume that $\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f^2] \ge \sup_{f \in \mathcal{F}_m} E[f^2]$ otherwise we are done. By (G.37) we have

$$\begin{aligned} \sup_{f \in \mathcal{F}_m} |\mathbb{E}_n[f^2] - E[f^2]| \quad &\lesssim_P \quad n^{-1/2} J(m) \log^{1/2} n \sup_{f \in \mathcal{F}_m} (\mathbb{E}_n[f^4] \vee E[f^4])^{1/2} \\ &\lesssim_P \quad n^{-1/2} J(m) \log^{1/2} n M(m, n) \sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f^2]^{1/2} \end{aligned}$$

where we used that $f^4 \leq f^2 M^2(m, n)$ with probability going to 1. Since for positive numbers a, c, and x, $x \leq a + c|x|^{1/2}$ implies that $x \leq 4a + 4c^2$ we conclude

$$\sup_{f \in \mathcal{F}_m} \mathbb{E}_n\left[f^2\right] \lesssim_P \sup_{f \in \mathcal{F}_m} E\left[f^2\right] + n^{-1} J(m)^2 M^2(m, n) \log n.$$

Inserting the bound in equation (G.36) gives the result.

Step 2. (Applying Lemma 15) We apply Lemma 15 to $\mathcal{F}_m$ with $\tau_m = 1/(4J(m)^2[K^2-1])$ for some large constant K to be set later, and

$$e_n(\mathcal{F}_m, \mathbb{P}_n) = J(m)\sqrt{\log n}\left(\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f^2]^{1/2} \vee E[f^2]^{1/2}\right)$$

assuming that n is sufficiently large (i.e., $n \geq \omega$). We observe that by (G.35), the bound $\epsilon \mapsto n(\epsilon, \mathcal{F}_m, \mathbb{P}_n)$ satisfies the monotonicity hypotheses of Lemma 15. Next note that $e_n(\mathcal{F}_m, \mathbb{P}_n)$ is $\sqrt{2}$-sub-exchangeable, because $\sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P}_n,2}$ is $\sqrt{2}$-sub-exchangeable, and $\rho(\mathcal{F}_m, \mathbb{P}_n) := \sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P}_n,2}/\|F_m\|_{\mathbb{P}_n,2} \geq 1/\sqrt{n}$ by Step 3 below. Thus,

$$
\begin{aligned}
\|F_m\|_{\mathbb{P}_n,2} \int_0^{\rho(\mathcal{F}_m, \mathbb{P}_n)/4} \sqrt{\log n(\epsilon, \mathcal{F}_m, P)} d\epsilon &\leq \|F_m\|_{\mathbb{P}_n,2} \int_0^{\rho(\mathcal{F}_m, \mathbb{P}_n)/4} J(m)\sqrt{\log(\omega/\epsilon)} d\epsilon \\
&\leq J(m)\sqrt{\log(n \vee \omega)} \sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P}_n,2}/2 \\
&\leq e_n(\mathcal{F}_m, \mathbb{P}_n),
\end{aligned}
$$

which follows by $\int_0^\rho \sqrt{\log(\omega/\epsilon)} d\epsilon \leq \left(\int_0^\rho 1 d\epsilon\right)^{1/2}\left(\int_0^\rho \log(\omega/\epsilon) d\epsilon\right)^{1/2} \leq \rho\sqrt{2\log(n \vee \omega)}$, for $1/\sqrt{n} \leq \rho \leq 1$.

Let $K > 1$ be sufficiently large (to be set below). Recall that $4\sqrt{2}c > 4$ where $c > 1$ is universal. Note that for any $f \in \mathcal{F}_m$, by Chebyshev inequality

$$P(|\mathbb{G}_n(f)| > 4\sqrt{2}cK e_n(\mathcal{F}_m, \mathbb{P}_n)\,) \leq \frac{\sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P},2}^2}{(4\sqrt{2}cK e_n(\mathcal{F}_m, \mathbb{P}_n))^2} \leq \frac{1}{(4\sqrt{2}cK)^2 J(m)^2 \log n} \leq \tau_m/2.$$

By Lemma 15 with our choice of τ_m, $\omega > 1$, and $\rho(\mathcal{F}_m, \mathbb{P}_n) \leq 1$,

$$
\begin{aligned}
\mathbb{P}\left\{\sup_{f \in \mathcal{F}_m} |\mathbb{G}_n(f)| > 4\sqrt{2}cK e_n(\mathcal{F}_m, \mathbb{P}_n)\right\} &\leq \frac{4}{\tau_m} \int_0^{1/2} (\omega/\epsilon)^{1-J(m)^2[K^2-1]} d\epsilon + \tau_m \\
&\leq \frac{4}{\tau_m} \frac{(1/[2\omega])^{J(m)^2[K^2-1]}}{J(m)^2[K^2-1]} + \tau_m,
\end{aligned}
$$

which can be made arbitrary small by choosing K sufficiently large (and recalling that $\tau_m \to 0$ as K grows).

Step 3. (Auxiliary calculations.) To establish that $\sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P}_n,2}$ is $\sqrt{2}$-sub-exchangeable, define $\tilde{Z}$ and $\tilde{Y}$ by exchanging any components in Z with corresponding components in Y. Then

$$\sqrt{2}(\sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P}_n(\tilde{Z}),2} \vee \sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P}_n(\tilde{Y}),2}) \geq (\sup_{f \in \mathcal{F}_m} \|f\|^2_{\mathbb{P}_n(\tilde{Z}),2} + \sup_{f \in \mathcal{F}_m} \|f\|^2_{\mathbb{P}_n(\tilde{Y}),2})^{1/2}$$

$$\geq (\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f(\tilde{Z}_i)^2] + \mathbb{E}_n[f(\tilde{Y}_i)^2])^{1/2} = (\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[f(Z_i)^2] + \mathbb{E}_n[f(Y_i)^2])^{1/2}$$

$$\geq (\sup_{f \in \mathcal{F}_m} \|f\|^2_{\mathbb{P}_n(Z),2} \vee \sup_{f \in \mathcal{F}_m} \|f\|^2_{\mathbb{P}_n(Y),2})^{1/2} = \sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P}_n(Z),2} \vee \sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P}_n(Y),2}.$$

Next we show that $\rho(\mathcal{F}_m, \mathbb{P}_n) := \sup_{f \in \mathcal{F}_m} \|f\|_{\mathbb{P}_n,2}/\|F_m\|_{\mathbb{P}_n,2} \geq 1/\sqrt{n}$. The latter bound follows from $\mathbb{E}_n[F_m^2] = \mathbb{E}_n[\sup_{f \in \mathcal{F}_m} |f(Z_i)|^2] \leq \sup_{i \leq n} \sup_{f \in \mathcal{F}_m} |f(Z_i)|^2$, and from the inequality $\sup_{f \in \mathcal{F}_m} \mathbb{E}_n[|f(Z_i)|^2] \geq \sup_{f \in \mathcal{F}_m} \sup_{i \leq n} |f(Z_i)|^2/n$. $\qquad\square$

The last technical lemma in this section bounds the uniform entropy for VC classes of functions (we refer Dudley [17] for formal definitions).

Lemma 17 (Uniform Entropy of VC classes). *Suppose $\mathcal{F}$ has VC index V, as $\varepsilon > 0$ goes to zero we have for $J = O(V)$*

$$\sup_Q N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \lesssim (1/\varepsilon)^J$$

where Q ranges over all discrete probabilities measures.

Proof. Being a VC class of index V, by Theorem 2.6.7 in [37] we have that the bound $\sup_Q \log N(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \lesssim V \log(1/\varepsilon)$ holds for ε sufficiently small (also making the expression bigger than 1). $\qquad\square$

Comment G.1. Although the product of two VC classes of functions may not be a VC class, if $\mathcal{F}$ has VC index V, the square of $\mathcal{F}$ is still a VC class whose VC index is at most $2V$.

G.3. **Bounds on Various Empirical Errors.** In this section we provide probabilistic bounds for the error terms under the primitive Condition S. Our results rely on empirical processes techniques. In particular, they rely on the maximal inequalities derived in Section G.2.

We start with a sequence of technical lemmas which are used in the proofs of the lemmas that bound the error terms $\epsilon_0 - \epsilon_6$.

Lemma 18. *Let $r = o(1)$. The class of functions*

$$\mathcal{F}_{m,n} = \{\alpha'(\psi_i(\beta, u) - \psi_i(\beta(u), u)) \ : \ u \in \mathcal{U}, \|\alpha\| \leq 1, \|\beta - \beta(u)\| \leq r\}$$

has VC index of $O(m)$.

Proof. Consider the classes $\mathcal{W} := \{Z_i'\alpha \ : \ \alpha \in \mathbb{R}^m\}$ and $\mathcal{V} := \{1\{Y_i \leq Z_i'\beta\} \ : \ \beta \in \mathbb{R}^m\}$ (for convenience let $A_i = (Z_i, Y_i)$). Their VC index is bounded by $m + 2$. Next consider $f \in \mathcal{F}_{m,n}$ which can be written in the form $f(A_i) := g(A_i)(1\{h(A_i) \leq 0\} - 1\{p(A_i) \leq 0\})$ where $g \in \mathcal{W}$, $1\{h \leq 0\}$ and $1\{p \leq 0\} \in \mathcal{V}$.

$$
\begin{aligned}
\{(A_i, t) : f(A_i) \leq t\} \ &= \ \{(A_i, t) : g(A_i)(1\{h(A_i) \leq 0\} - 1\{p(A_i) \leq 0\}) \leq t\} \\
&= \ \{(A_i, t) : h(A_i) > 0, p(A_i) > 0, t \geq 0\} \cup \\
&\cup \ \{(A_i, t) : h(A_i) \leq 0, p(A_i) \leq 0, t \geq 0\} \cup \\
&\cup \ \{(A_i, t) : h(A_i) \leq 0, p(A_i) > 0, g(A_i) \leq t\} \cup \\
&\cup \ \{(A_i, t) : h(A_i) > 0, p(A_i) \leq 0, -g(A_i) \leq t\}.
\end{aligned}
$$

Since each one of the sets can be written as three intersections of basic sets, it follows that $\mathcal{F}_{m,n}$ has VC index at most $O(m)$. $\qquad \square$

Lemma 19. *The class of functions*

$$\mathcal{H}_{m,n} = \{1\{|Y_i - Z_i'\beta| \leq h\}(\alpha'Z_i)^2 \ : \ \|\beta - \beta(u)\| \leq r, \ h \in (0, H], \alpha \in S^{m-1}\}$$

has VC index of $O(m)$.

Proof. The proof is similar to that of Lemma 18. $\qquad \square$

Lemma 20. *The family of functions $\mathcal{G}_{m,n} = \{\alpha'\psi_i(\beta(u), u) \ : \ u \in \mathcal{U}, \alpha \in S^{m-1}\}$ has VC index of $O(m)$.*

Proof. The proof is similar to the proof of Lemma 18. $\qquad \square$

Lemma 21. *The family of functions*

$$\mathcal{A}_{n,m} = \{\alpha'Z\left(1\{Y \leq Z'\beta(u) + R(X, u)\} - 1\{Y \leq Z'\beta(u)\}\right) \ : \ \alpha \in S^{m-1}, \ u \in \mathcal{U}\}$$

has VC index of $O(m)$.

Proof. The key observation is that the function $Z'\beta(u) + R(X, u)$ is monotone in u so that

$$\{1\{Y \leq Z'\beta(u) + R(X, u)\} : u \in \mathcal{U}\}$$

has VC index of 1 and that $\{1\{Y \leq Z'\beta(u)\} : u \in \mathcal{U}\} \subset \{1\{Y \leq Z'\beta : \beta \in \mathbb{R}^m\}$. The proof then follows similarly to Lemma 18. $\qquad \square$

Consider the maximum between the maximum eigenvalue associated with the empirical Gram matrix and the population Gram matrix

$$\phi_n = \max_{\alpha \in S^{m-1}} \mathbb{E}_n \left[(\alpha' Z_i)^2 \right] \vee E \left[(\alpha' Z_i)^2 \right]. \tag{G.39}$$

The factor ϕ_n will be used to bound the quantities ϵ_0 and ϵ_1 in the analysis for the rate of convergence. Next we state a result due to Guédon and Rudelson [18] specialized to our framework.

Theorem 13 (Guédon and Rudelson [18]). *Let $Z_i \in \mathbb{R}^m$, $i = 1, \ldots, n$, be random vectors such that*

$$\delta^2 := \frac{\log n}{n} \cdot \frac{E \left[\max_{i \leq n} \| Z_i \|^2 \right]}{\max_{\alpha \in S^{m-1}} E \left[(Z_i' \alpha)^2 \right]} < 1$$

we have

$$E \left[\max_{\alpha \in S^{m-1}} \left| \frac{1}{n} \sum_{i=1}^n (Z_i' \alpha)^2 - E \left[(Z_i' \alpha)^2 \right] \right| \right] \leq 2\delta \cdot \max_{\alpha \in S^{m-1}} E \left[(Z_i' \alpha)^2 \right].$$

Corollary 2. *Under Condition S and $\zeta_m^2 \log n = o(n)$, for $\lambda_{max} = \max_{\alpha \in S^{m-1}} E \left[(Z_i' \alpha)^2 \right]$, we have that for n large enough ϕ_n as defined in (G.39) satisfies*

$$E \left[\phi_n \right] \leq \left(1 + 2\sqrt{\frac{\zeta_m^2 \log n}{n \lambda_{max}}} \right) \lambda_{max} \quad and \quad P(\phi_n > 2\lambda_{max}) \leq 2\sqrt{\frac{\zeta_m^2 \log n}{n \lambda_{max}}}.$$

Proof. Let δ be defined as in Theorem 13. Next note that $E \left[\max_{i=1,\ldots,n} \| Z_i \|^2 \right] \lesssim \zeta_m^2$ under S.4, and $\lambda_{max} \lesssim 1$ and $\lambda_{max} \gtrsim 1$ under S.3. Therefore, $\delta^2 \lesssim (\zeta_m^2 \log n)/n$ in Theorem 13. The growth condition $\zeta_m^2 \log n = o(n)$ yields $\delta < 1$ as n grows.

The first result follows by applying Theorem 13 and the triangle inequality.

To show the second relation note that the event $\{\phi_n > 2\lambda_{max}\}$ cannot occur if $\phi_n = \max_{\alpha \in S^{m-1}} E[(Z_i' \alpha)^2] = \lambda_{max}$. Thus

$$
\begin{aligned}
P(\phi_n > 2\lambda_{max}) \quad &= P(\max_{\alpha \in S^{m-1}} \mathbb{E}_n[(Z_i' \alpha)^2] > 2\lambda_{max}) \\
&\leq P(\max_{\alpha \in S^{m-1}} \left| \mathbb{E}_n[(Z_i' \alpha)^2] - E \left[(Z_i' \alpha)^2 \right] \right| > \lambda_{max}) \\
&\leq E \left[\max_{\alpha \in S^{m-1}} \left| \mathbb{E}_n[(Z_i' \alpha)^2] - E \left[(Z_i' \alpha)^2 \right] \right| \right] / \lambda_{max} \\
&\leq 2\delta,
\end{aligned}
$$

by the triangle inequality, the Markov inequality and Theorem 13. $\qquad \square$

Next we proceed to bound the various approximation errors terms.

Lemma 22 (Controlling error ϵ_1). *Under conditions S.1-S.4 we have*

$$\epsilon_1(m,n) \quad \lesssim_P \quad \sqrt{m \log n \, \phi_n}.$$

Proof. Consider the class of functions $\mathcal{F}_{m,n}$ defined in Lemma 18 so that

$$\epsilon_1(m,n) = \sup_{f \in \mathcal{F}_{m,n}} |\mathbb{G}_n(f)|.$$

From Lemma 17 we have that $J(m) \lesssim \sqrt{m}$. By Step 2 of Lemma 16, see equation (G.36),

$$\sup_{f \in \mathcal{F}_{m,n}} |\mathbb{G}_n(f)| \lesssim_P \sqrt{m \log n} \sup_{f \in \mathcal{F}_{m,n}} \left(\mathbb{E}_n[f^2] \vee E[f^2]\right)^{1/2} \qquad (G.40)$$

The score function $\psi_i(\cdot, \cdot)$ satisfies the following inequality for any $\alpha \in S^{m-1}$

$$\left|(\psi_i(\beta,u) - \psi_i(\beta(u),u))' \alpha\right| = |\alpha' Z_i| \left|1\{y_i \le Z_i'\beta\} - 1\{y_i \le Z_i'\beta(u)\}\right| \le |\alpha' Z_i|.$$

Therefore

$$\mathbb{E}_n[f^2] \le \mathbb{E}_n[|\alpha' Z_i|^2] \le \phi_n \quad \text{and} \quad E[f^2] \le E\left[\frac{1}{n}\sum_{i=1}^n |\alpha' Z_i|^2\right] \le \phi_n \qquad (G.41)$$

by definition (G.39). Combining (G.41) with (G.40) we obtain the result. $\square$

Lemma 23 (Controlling error ϵ_0 and Pivotal process norm). *Under the conditions S.1-S.4*

$$\epsilon_0(m,n) \quad \lesssim_P \quad \sqrt{m \log n \, \phi_n} \quad \text{and} \quad \sup_{u \in \mathcal{U}} \|\mathbb{U}_n(u)\| \lesssim_P \sqrt{m \log n \, \phi_n}.$$

Proof. For ϵ_0, the proof is similar to the proof of Lemma 22 and relying on Lemma 20 and noting that for $g \in \mathcal{G}_{m,n}$

$$\mathbb{E}_n[g^2] = \mathbb{E}_n[(\alpha' \psi_i(\beta(u),u))^2] = \mathbb{E}_n[(\alpha' Z_i)^2 (1\{y_i \le Z_i'\beta(u)\} - u)^2] \le \mathbb{E}_n[(\alpha' Z_i)^2] \le \phi_n.$$

Similarly, $E[g^2] \le \phi_n$.

The second relation follows similarly. $\square$

Lemma 24 (Bounds on $\epsilon_1(m,n)$ and $\epsilon_2(m,n)$). *Under conditions S.1-S.4 we have*

$$\epsilon_1(m,n) \lesssim_P \sqrt{m\zeta_m r \log n} + \frac{m\zeta_m}{\sqrt{n}} \log n \quad \text{and} \quad \epsilon_2(m,n) \lesssim_P \sqrt{n}\zeta_m r^2 + \sqrt{n}m^{-\kappa} r.$$

Proof. Part 1 ($\epsilon_1(m,n)$) The first bound will follow from the application of the maximal inequality derived in Lemma 16.

For any $B \ge 0$, define the class of functions $\mathcal{F}_{n,m}$ as in Lemma 18 which characterize ϵ_1. By the (second) maximal inequality in Lemma 16

$$\epsilon_1(m,n) = \sup_{f \in \mathcal{F}_{n,m}} |\mathbb{G}_n(f)| \lesssim_P J(m) \left(\sup_{f \in \mathcal{F}_m} E\left[f^2\right] + n^{-1}J(m)^2 M^2(m,n) \log n\right)^{1/2} \log^{1/2} n,$$

where $J(m) \lesssim \sqrt{m}$ by the VC dimensions of $\mathcal{F}_{n,m}$ being of $O(m)$, Lemma 18, and $M(m,n) = \max_{1 \leq i \leq n} \|Z_i\| \leq \zeta_m$. The bound stated in the lemma holds provided we can show that

$$\sup_{f \in \mathcal{F}_m} E\left[f^2\right] \lesssim r\zeta_m. \tag{G.42}$$

The score function $\psi_i(\cdot, \cdot)$ satisfies the following inequality for any $\alpha \in S^{m-1}$

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

$$\leq |Z_i'\alpha| \cdot 1\{|Y_i - Z_i'\beta(u)| \leq |Z_i'(\beta - \beta(u))|\}.$$

Thus,

$$E\left[(\alpha'(\psi_i(\beta, u) - \psi_i(\beta(u), u)))^2\right] \leq E\left[\int |Z'\alpha|^2 1\{|Y - Z'\beta(u)| \leq |Z'(\beta - \beta(u))|\} f_{Y|Z}(y|Z) dy\right]$$

$$\leq E\left[(Z'\alpha)^2 \cdot \min(2\bar{f}|Z'(\beta - \beta(u))|, 1)\right]$$

$$\leq 2\bar{f}\|\beta - \beta(u)\| \sup_{\alpha \in S^{m-1}, \gamma \in S^{m-1}} E\left[|Z_i'\alpha|^2 |Z_i'\gamma|\right]$$

$$= 2\bar{f}\|\beta - \beta(u)\| \sup_{\alpha \in S^{m-1}} E\left[|Z_i'\alpha|^3\right],$$

where $\sup_{\alpha \in S^{m-1}} E[|Z_i'\alpha|^3] \lesssim \zeta_m$ by S.3 and S.4.

Therefore, we have the upper bound (G.42).

Part 2 ($\epsilon_2(m,n)$) To show the bound for ϵ_2, note that by Lemma 12, for any $\alpha \in S^{m-1}$

$$\sqrt{n}|\alpha'(J_m(u) - \widetilde{J}_m(u))(\beta - \beta(u))| \lesssim \sqrt{n} m^{-\kappa} r.$$

For $\alpha \in S^{m-1}$ and $\widetilde{J}_m(u) = E[f_{Y|X}(Z'\beta(u)|X) ZZ']$, define

$$\epsilon_2(m,n,\alpha) = n^{1/2} |\alpha'(E[\psi(Y,Z,\beta,u)] - E[\psi(Y,Z,\beta(u),u)]) - \alpha'\widetilde{J}_m(u)(\beta - \beta(u))|.$$

Thus $\epsilon_2(m,n) \lesssim \sup_{\alpha \in S^{m-1}, (u,\beta) \in R_{n,m}} \epsilon_2(m,n,\alpha) + \sqrt{n} m^{-\kappa} r$.

Note that since $E[\alpha'\psi_i(\beta(u), u)] = 0$, for some $\tilde{\beta}$ in the line segment between $\beta(u)$ and β, $E[\alpha'\psi_i(\beta, u)] = E[f_{Y|X}(Z_i'\tilde{\beta}|X)(Z_i'\alpha)Z_i'](\beta - \beta(u))$. Thus, using that $|f_{Y|X}(Z'\tilde{\beta}|X) - f_{Y|X}(Z'\beta(u)|X)| \leq \overline{f'}|Z'(\tilde{\beta} - \beta(u))| \leq \overline{f'}|Z'(\beta - \beta(u))|$,

$$\epsilon_2(m, n, \alpha) \quad = n^{1/2} \left| E \left[(\alpha' Z)(f_{Y|X}(Z'\tilde{\beta}|X) - f_{Y|X}(Z'\beta(u)|X))Z' \right] (\beta - \beta(u)) \right|$$

$$\leq n^{1/2} E \left[|\alpha' Z| \; |Z'(\beta - \beta(u))|^2 \overline{f'} \right]$$

$$\leq n^{1/2} \overline{f'} \|\beta - \beta(u)\|^2 \sup_{\alpha \in S^{m-1}} E[|\alpha' Z|^3],$$

where $\sup_{\alpha \in S^{m-1}} E[|Z_i'\alpha|^3] \lesssim \zeta_m$ by S.3 and S.4.

$\square$

Lemma 25 (Bounds on Approximation Error for Uniform Linear Approximation). *Under Condition S,*

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

satisfies

$$\sup_{\alpha \in S^{m-1}, u \in \mathcal{U}} |\alpha' \tilde{r}_n(u)| \lesssim_P \min \left\{ \sqrt{m^{1-\kappa} \log n} + \frac{\zeta_m m \log n}{\sqrt{n}}, \; \sqrt{n \phi_n m^{-\kappa} \bar{f}} \right\}.$$

Proof. The second bound follows by Cauchy-Schwarz, $|R(X_i, u)| \lesssim m^{-\kappa}$, and bounded conditional probability density function of Y given X.

The proof of the first bound is similar to the bound on ϵ_1 in Lemma 24, it will also follow from the application of the maximal inequality derived in Lemma 16.

Define the class of functions

$$\mathcal{A}_{n,m} = \{\alpha' Z \left(1\{Y \leq Z'\beta(u) + R(X, u)\} - 1\{Y \leq Z'\beta(u)\} \right) \; : \; \alpha \in S^{m-1}, \; u \in \mathcal{U}\}.$$

By the (second) maximal inequality in Lemma 16,

$$\sup_{\alpha \in S^{m-1}, u \in \mathcal{U}} |\alpha' \tilde{r}_n(u)| \quad = \quad \sup_{f \in \mathcal{A}_{n,m}} |\mathbb{G}_n(f)|$$

$$\lesssim_P \quad J(m) \left(\sup_{f \in \mathcal{A}_{n,m}} E\left[f^2\right] + n^{-1} J(m)^2 M^2(m, n) \log n \right)^{1/2} \log^{1/2} n,$$

where $J(m) \lesssim \sqrt{m}$ by the VC dimensions of $\mathcal{A}_{n,m}$ being of $O(m)$ by Lemma 21, and $M(m, n) = \zeta_m$ by S.4. The bound stated in the lemma holds provided we can show that

$$\sup_{f \in \mathcal{A}_{n,m}} E\left[f^2\right] \lesssim m^{-\kappa}. \tag{G.43}$$

For any $f \in \mathcal{A}_{n,m}$

$$|f(Y_i, Z_i, X_i)| = |Z_i'\alpha| \, |1\{Y_i \leq Z_i'\beta(u) + R(X_i, u)\} - 1\{Y_i \leq Z_i'\beta(u)\}|$$

$$\leq |Z_i'\alpha| \cdot 1\{|Y_i - Z_i'\beta(u)| \leq |R(X_i, u)|\}.$$

Thus, since $|f_{Y|X}(y|X)| \leq \bar{f}$,

$$E\left[f^2(Y, Z, X)\right] \leq E\left[\int |Z'\alpha|^2 1\{|y - Z'\beta(u)| \leq |R(X, u)|\} f_{Y|X}(y|X) dy\right]$$

$$\leq E\left[(Z'\alpha)^2 \cdot \min(2\bar{f}R(X, u), 1)\right]$$

$$\leq 2\bar{f}m^{-\kappa} \sup_{\alpha \in S^{m-1}} E\left[|Z_i'\alpha|^2\right],$$

where $\sup_{\alpha \in S^{m-1}} E[|Z_i'\alpha|^2] \lesssim 1$ by invoking S.3. Therefore, we have the upper bound (G.43).

$\square$

Lemma 26 (Bound on $\epsilon_3(m, n)$). *Let $\widehat{\beta}(u)$ be a solution to the perturbed QR problem*

$$\widehat{\beta}(u) \in \arg\min_{\beta \in \mathbb{R}^m} \mathbb{E}_n[\rho_u(Y_i - Z_i'\beta)] + \mathcal{A}_n(u)'\beta.$$

If the data are in general position,

$$\epsilon_3(m, n) = \sup_{u \in \mathcal{U}} n^{1/2} \|\mathbb{E}_n[\psi_i(\widehat{\beta}(u), u)] + \mathcal{A}_n(u)\| \leq \min\left\{\frac{m}{\sqrt{n}}\zeta_m, \ \phi_n\sqrt{m}\right\}$$

holds with probability 1.

Proof. Note that the dual problem associated with the perturbed QR problem 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 $a_i(\widehat{\beta}(u)) := (u - 1\{Y_i \leq Z_i'\widehat{\beta}(u)\})$, by the triangle inequality

$$\epsilon_3(m, n) \leq \sup_{\|\alpha\| \leq 1, u \in \mathcal{U}} \sqrt{n} \left|\mathbb{E}_n\left[(Z_i'\alpha)(a_i(\widehat{\beta}(u)) - \widehat{a}_i(u))\right]\right| +$$
$$+ \sup_{u \in \mathcal{U}} \sqrt{n} \|\mathbb{E}_n[Z_i\widehat{a}_i(u)] + \mathcal{A}_n(u)\|.$$

By dual feasibility $\mathbb{E}_n[Z_i\widehat{a}_i(u)] = -\mathcal{A}_n(u)$, and the second term is identically equal to zero.

We 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 $|a_i(\widehat{\beta}(u)) - \widehat{a}_i(u)| \leq 1$

$$\epsilon_3(m, n) \leq \sup_{\|\alpha\| \leq 1, u \in \mathcal{U}} \sqrt{n} \sqrt{\mathbb{E}_n\left[(Z_i'\alpha)^2\right]} \sqrt{\mathbb{E}_n\left[\{a_i(\widehat{\beta}(u)) - \widehat{a}_i(u)\}^2\right]} \leq \phi_n \sqrt{m}$$

and, with probability 1,

$$\epsilon_3(m, n) \leq \sup_{\|\alpha\| \leq 1, u \in \mathcal{U}} \sqrt{n} \mathbb{E}_n\left[1\{a_i(\widehat{\beta}(u)) \neq \widehat{a}_i(u)\} \max_{1 \leq i \leq n} \|Z_i\|\right] \leq \frac{m}{\sqrt{n}} \max_{1 \leq i \leq n} \|Z_i\|.$$

$\square$

Lemma 27 (Bound on $\epsilon_4(m, n)$). *Under conditions $S.1 - S.4$, and $\zeta_m^2 \log n = o(n)$,*

$$\epsilon_4(m, n) \lesssim_P \sqrt{\frac{\zeta_m^2 \log n}{n}} = o(1).$$

Proof. The result follows from Theorem 13 of Guédon and Rudelson [18] under our assumptions. $\square$

Lemma 28 (Bounds on $\epsilon_5(m, n)$ and $\epsilon_6(m, n)$). *Under $S.2$, $h_n = o(1)$, and $r = o(1)$,*

$$\epsilon_5(m, n) \lesssim_P \sqrt{\frac{\zeta_m^2 m \log n}{nh_n}} + \frac{m\zeta_m^2}{nh_n} \log n \quad and \quad \epsilon_6(m, n) \lesssim m^{-\kappa} + r\zeta_m + h_n.$$

Proof. To bound ϵ_5 first let

$$\mathcal{H}_{m,n} = \left\{1\{|Y_i - Z_i'\beta| \leq h\}(\alpha'Z_i)^2 : \|\beta - \beta(u)\| \leq r, \ h \in (0, H], \alpha \in S^{m-1}\right\}.$$

Then, by Lemma 16

$$\begin{aligned}
\epsilon_5(m, n) &= \frac{n^{-1/2}}{h_n} \sup_{f \in \mathcal{H}_{n,m}} |\mathbb{G}_n(f)| \\
&\lesssim_P \frac{n^{-1/2}}{h_n} J(m) \left(\sup_{f \in \mathcal{H}_m} E\left[f^2\right] + n^{-1} J(m)^2 M^2(m, n) \log n\right)^{1/2} \log^{1/2} n,
\end{aligned}$$

where $J(m) \lesssim \sqrt{m}$ by the VC dimensions of $\mathcal{H}_{n,m}$ being of $O(m)$ by Lemma 19, and $M(m, n) = \max_{i \leq n} H_{m,n,i}$, where $H_{m,n,i} = \|Z_i\|^2$, is the envelope of $\mathcal{H}_{m,n}$ in the sample of size n. Therefore the envelope is bounded by $\max_{1 \leq i \leq n} \|Z_i\|^2$. We also get that

$$\begin{aligned}
\sup_{f \in \mathcal{H}_{m,n}} E\left[f^2\right] &\leq \sup_{\beta \in \mathbb{R}^m, \alpha \in S^{m-1}} E\left[1\{|Y_i - Z_i'\beta| \leq h_n\}(\alpha'Z_i)^4\right] \\
&\leq 2\bar{f}h_n \sup_{\alpha \in S^{m-1}} E\left[(\alpha'Z_i)^4\right].
\end{aligned}$$

By S.2 $\bar{f}$ is bounded, and by S.4

$$\sup_{f \in \mathcal{H}_{m,n}} E\left[f^2\right] \le 2\bar{f}\zeta_m^2 h_n \sup_{\alpha \in S^{m-1}} E\left[(\alpha' Z_i)^2\right] \lesssim \zeta_m^2 h_n.$$

Collecting terms,

$$
\begin{aligned}
\epsilon_5(m,n) \ \lesssim_P \ & \frac{n^{-1/2}}{h_n}\sqrt{m}\left(\zeta_m^2 h_n + (m\max_{1\le i\le n}\|Z_i\|^4/n)\log n\right)^{1/2}\log^{1/2}n \\
\lesssim_P \ & \sqrt{\frac{\zeta_m^2 m \log n}{n h_n}} + \sqrt{\frac{m^2}{n h_n^2}\frac{\max_{1\le i\le n}\|Z_i\|^4}{n}}\log n \\
\lesssim_P \ & \sqrt{\frac{\zeta_m^2 m}{n h_n}}\log^{1/2}n + \frac{m\max_{1\le i\le n}\|Z_i\|^2}{n h_n}\log n.
\end{aligned}
$$

To show the bound on ϵ_6 note that $|f'_{Y|X}(y|x)| \le \bar{f}'$ by S.2. Therefore

$$
\begin{aligned}
E\left[1\{|Y - Z'\beta| \le h_n\}(\alpha' Z)^2\right] &= E\left[(\alpha' Z)^2 \int_{-h_n}^{h_n} f_{Y|X}(Z'\beta + t|X)dt\right] \\
&= E\left[(\alpha' Z)^2 \int_{-h_n}^{h_n} f_{Y|X}(Z'\beta|X) + t f'_{Y|X}(Z'\beta + \tilde{t}|X)dy\right] \\
&= 2h_n E\left[f_{Y|X}(Z'\beta|X)(\alpha' Z)^2\right] + O(2h_n^2 \bar{f}' E\left[(Z'\alpha)^2\right])
\end{aligned}
$$

by the mean-value theorem. Moreover, we have for any $(u,\beta) \in R_{m,n}$

$$
\begin{aligned}
E\left[f_{Y|X}(Z'\beta|X)(\alpha' Z)^2\right] &= E\left[f_{Y|X}(Z'\beta(u)|X)(\alpha' Z)^2\right] + \\
&\quad + E\left[(f_{Y|X}(Z'\beta|X) - f_{Y|X}(Z'\beta(u)|X))(\alpha' Z)^2\right] \\
&= E\left[f_{Y|X}(Z'\beta(u)|X)(\alpha' Z)^2\right] + O(E\left[\bar{f}'Z'(\beta - \beta(u))(\alpha' Z)^2\right]) \\
&= \alpha' \widetilde{J}_m(u)\alpha + O(\bar{f}'r\sup_{\alpha \in S^{m-1}} E\left[|\alpha' Z|^3\right]) \\
&= \alpha' J_m(u)\alpha + O(m^{-\kappa}) + O(\bar{f}'r\sup_{\alpha \in S^{m-1}} E\left[|\alpha' Z|^3\right])
\end{aligned}
$$

where the last line follows from Lemma 13, S.2 and S.5. Since $\bar{f}'$ is bounded by assumption S.2, we obtain $\epsilon_6(m,n) \lesssim h_n + m^{-\kappa} + r\sup_{\alpha \in S^{m-1}} E\left[|\alpha' Z|^3\right]$. Finally, conditions S.3 and S.4 yields $\sup_{\alpha \in S^{m-1}} E\left[|\alpha' Z|^3\right] \le \zeta_m \sup_{\alpha \in S^{m-1}} E\left[|\alpha' Z|^2\right] \lesssim \zeta_m$ and the results follows.

$\square$

References

[1] D. W. K. Andrews. Empirical process methods in econometrics. *Handbook of Econometrics, Volume IV, Chapter 37, Edited by R. F. Engle and D. L. McFadden, Elsvier Science B.V.*, pages 2247–2294, 1994.

[2] Donald W. K. Andrews. Asymptotic normality of series estimators for nonparametric and semiparametric regression models. *Econometrica*, 59(2):307–345, 1991.

[3] O. Arias, K. F. Hallock, and Walter Sosa-Escudero. Individual heteonerity in the returns to schooling: instrumental variables quantile regression using twins data. *Empirical Economics*, 26(1):7–40, 2001.

[4] R. Barlow, D. Bartholemew, J. Bremner, and H. Brunk. *Statistical Inference Under Order Restrictions*. John Wiley, New York, 1972.

[5] A. Belloni, X. Chen, and V. Chernozhukov. New asymptotic theory for series estimators. 2009.

[6] A. Belloni and V. Chernozhukov. ℓ_1-penalized quantile regression for high dimensional sparse models. *Ann. Statist.*, 39(1):82–130, 2011.

[7] M. Buchinsky. Changes in the u.s. wage structure 1963-1987: Application of quantile regression. *Econometrica*, 62(2):405–458, Mar. 1994.

[8] Matias D. Cattaneo, Richard K. Crump, and Michael Jansson. Robust data-driven inference for density-weighted average derivatives. *J. Amer. Statist. Assoc.*, 105(491):1070–1083, 2010. With supplementary material available online.

[9] Arun G. Chandrasekhar, Victor Chernozhukov, Francesca Molinari, and Paul Schrimpf. Inference on sets of best linear approximations to functions. *MIT Working Paper*, 2010.

[10] Probal Chaudhuri. Nonparametric estimates of regression quantiles and their local Bahadur representation. *Ann. Statist.*, 19(2):760–777, 1991.

[11] Probal Chaudhuri, Kjell Doksum, and Alexander Samarov. On average derivative quantile regression. *Ann. Statist.*, 25(2):715–744, 1997.

[12] Xiaohong Chen. Large sample sieve estimation of semi-nonparametric models. *In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics*, 6B(Chapter 76), 2006.

[13] Xiaohong Chen and Xiaotong Shen. Sieve extremum estimates for weakly dependent data. *Econometrica*, 66(2):289–314, 1998.

[14] V. Chernozhukov, I. Fernández-Val, and A. Galichon. Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3):559–575, 2009.

[15] V. Chernozhukov, Sokbae Lee, and Adam M. Rosen. Intersection bounds: estimation and inference. *arXiv*, (0907.3503v1), 2009.

[16] William S. Cleveland. Robust locally weighted regression and smoothing scatterplots. *J. Amer. Statist. Assoc.*, 74(368):829–836, 1979.

[17] R. Dudley. *Uniform Cental Limit Theorems*. Cambridge Studies in advanced mathematics, 2000.

[18] O. Guédon and M. Rudelson. l_p-moments of random vectors via majorizing measures. *Advances in Mathematics*, 208:798–823, 2007.

[19] Peter Hall and Simon J. Sheather. On the distribution of a studentized quantile. *Journal of the Royal Statistical Society. Series B (Methodological)*, 50(3):381–391, 1988.

[20] Wolfgang K. Härdle, Yaacov Ritov, and Song Song. Partial linear quantile regression and bootstrap confidence bands. SFB 649 Discussion Paper 2010-002, 2009.

[21] Jerry A. Hausman and Whitney K. Newey. Nonparametric estimation of exact consumers surplus and deadweight loss. *Econometrica*, 63(6):pp. 1445–1476, 1995.

[22] X. He and Q.-M. Shao. On parameters of increasing dimenions. *Journal of Multivariate Analysis*, 73:120–135, 2000.

[23] J. L. Horowitz and S. Lee. Nonparametric estimation of an additive quantile regression model. *Journal of the American Statistical Association*, 100(472):1238–1249, 2005.

[24] R. Koenker. *Quantile regression*. Cambridge University Press, New York, 2005.

[25] R. Koenker and G. Basset. Regression quantiles. *Econometrica*, 46(1):33–50, 1978.

[26] Roger Koenker. *quantreg: Quantile Regression*, 2008. R package version 4.24.

[27] Efang Kong, Oliver Linton, and Yingcun Xia. Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory*, 26(5):1529–1564, 2010.

[28] S. Lee. Efficient semiparametric estimation of a partially linear quantile regression model. *Econometric Theory*, 19:1–31, 2003.

[29] Whitney K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, 79:147–168, 1997.

[30] M. I. Parzen, L. J. Wei, and Z. Ying. A resampling method based on pivotal estimating functions. *Biometrika*, 81(2):341–350, 1994.

[31] David Pollard. *A User's Guide to Measure Theoretic Probability*. Cambridge Series in Statistical and Probabilistic Mathematics, 2001.

[32] James L. Powell. Least absolute deviations estimation for the censored regression model. *Journal of Econometrics*, 25(3):303–325, 1984.

[33] R Development Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.

[34] Richard Schmalensee and Thomas M. Stoker. Household gasoline demand in the united states. *Econometrica*, 67(3):pp. 645–662, 1999.

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

[36] Sara van de Geer. M-estimation using penalties or sieves. *Journal of Statistical Planning and Inference*, 108(1-2):55–69, 2002.

[37] A. W. van der Vaart and J. A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics, 1996.

[38] Halbert L. White. Nonparametric estimation of conditional quantiles using neural networks. In *Proceedings of the Symposium on the Interface*, pages 190–199, 1992.

[39] Adonis Yatchew and Joungyeo Angela No. Household gasoline demand in canada. *Econometrica*, 69(6):pp. 1697–1709, 2001.