INFERENCE FOR BEST LINEAR APPROXIMATIONS TO SET IDENTIFIED FUNCTIONS

ARUN CHANDRASEKHAR, VICTOR CHERNOZHUKOV, FRANCESCA MOLINARI, AND PAUL SCHRIMPF

Abstract. This paper provides inference methods for best linear approximations to functions which are known to lie within a band. It extends the partial identification literature by allowing the upper and lower functions defining the band to be any functions, including ones carrying an index, which can be estimated parametrically or non-parametrically. The identification region of the parameters of the best linear approximation is characterized via its support function, and limit theory is developed for the latter. We prove that the support function approximately converges to a Gaussian process and establish validity of the Bayesian bootstrap. The paper nests as special cases the canonical examples in the literature: mean regression with interval valued outcome data and interval valued regressor data. Because the bounds may carry an index, the paper covers problems beyond mean regression; the framework is extremely versatile. Applications include quantile and distribution regression with interval valued data, sample selection problems, as well as mean, quantile, and distribution treatment effects. Moreover, the framework can account for the availability of instruments. An application is carried out, studying female labor force participation along the lines of Mulligan and Rubinstein (2008).

JEL Classification Codes: C13, C31.
AMS Classification Codes: 62G08, 62G09, 62P20.
First draft presented on Dec. 4, 2008.
This version: December 27, 2012

Key words and phrases. Set identified function; best linear approximation; partial identification; support function; bayesian bootstrap; convex set.

1. Introduction

This paper contributes to the literature on estimation and inference for best linear approximations to set identified functions. Specifically, we work with a family of functions $f(x, \alpha)$ indexed by some parameter $\alpha \in \mathcal{A}$, that is known to satisfy $\theta_0(x, \alpha) \leq f(x, \alpha) \leq \theta_1(x, \alpha)$ $x - a.s.$, with $x \in \mathbb{R}^d$ a vector of regressors. Econometric frameworks yielding such restriction are ubiquitous in economics and in the social sciences, as illustrated by Manski (2003, 2007). Cases explicitly analyzed in this paper include: (1) mean regression; (2) quantile regression; and (3) distribution and duration regression, in the presence of interval valued data, including hazard models with interval-valued failure times; (4) sample selection problems; (5) mean treatment effects; (6) quantile treatment effects; and (7) distribution treatment effects, see Section 3 for details.[1] Yet, the methodology that we propose can be applied to virtually any of the frameworks discussed in Manski (2003, 2007). In fact, our results below also allow for exclusion restrictions that yield intersection bounds of the form $\sup_{v \in \mathcal{V}} \theta_0(x, v, \alpha) \equiv \theta_0(x, \alpha) \leq f(x, \alpha) \leq \theta_1(x, \alpha) \equiv \inf_{v \in \mathcal{V}} \theta_1(x, v, \alpha)$ $x - a.s.$, with v an instrumental variable taking values in a set $\mathcal{V}$. The bounding functions $\theta_0(x, \alpha)$ and $\theta_1(x, \alpha)$ may be indexed by a parameter $\alpha \in \mathcal{A}$ and may be *any* estimable function of x.

Our method appears to be the first and currently only method available in the literature for performing inference on best linear approximations to set identified functions when the bounding functions $\theta_0(x, \alpha)$ and $\theta_1(x, \alpha)$ need to be estimated. Moreover, we allow for the functions to be estimated both parametrically as well as non-parametrically via series estimators. Previous closely related contributions by Beresteanu and Molinari (2008) and Bontemps, Magnac, and Maurin (2010) provided inference methods for best linear approximations to conditional expectations in the presence of interval outcome data. In that environment, the bounding functions do not need to be estimated, as the set of best linear approximations can be characterized directly through functions of moments of the observable variables. Hence, our paper builds upon and significantly generalizes their results. These generalizations are our main contribution and are imperative for many empirically relevant applications.

Our interest in best linear approximations is motivated by the fact that when the restriction $\theta_0(x, \alpha) \leq f(x, \alpha) \leq \theta_1(x, \alpha)$ $x - a.s.$ summarizes all the information available to the researcher, the sharp identification region for $f(\cdot, \alpha)$ is given by the set of functions

$$\mathfrak{F}(\alpha) = \{\phi(\cdot, \alpha) : \theta_0(x, \alpha) \leq \phi(x, \alpha) \leq \theta_1(x, \alpha) \; x - a.s.\}$$

The set $\mathfrak{F}(\alpha)$, while sharp, can be difficult to interpret and report, especially when x is multi-dimensional. Similar considerations apply to related sets, such as for example the set of marginal effects of components of x on $f(x, \alpha)$. Consequently, in this paper we focus on the sharp set of parameters characterizing best linear approximations to the functions comprising $\mathfrak{F}(\alpha)$. This set is of great interest in empirical work because of its tractability. Moreover, when the set identified function is a conditional expectation, the corresponding

[1]For example, one may be interested in the α-conditional quantile of a random variable y given x, denoted $Q_y(\alpha|x)$, but only observe interval data $[y_0, y_1]$ which contain y with probability one. In this case, $f(x, \alpha) \equiv Q_y(\alpha|x)$ and $\theta_\ell(x, \alpha) \equiv Q_\ell(\alpha|x)$, $\ell = 0, 1$, the conditional quantiles of properly specified random variables.

set of best linear approximations is robust to model misspecification (Ponomareva and Tamer (2010)).

In practice, we propose to estimate the sharp set of parameter vectors, denoted $B(\alpha)$, corresponding to the set of best linear approximations. Simple linear transformations applied to $B(\alpha)$ yield the set of best linear approximations to $f(x,\alpha)$, the set of linear combinations of components of $b \in B(\alpha)$, bounds on each single coefficient, etc. The set $B(\alpha)$ is especially tractable because it is a transformation, through linear operators, of the random interval $[\theta_0(x,\alpha), \theta_1(x,\alpha)]$, and therefore is convex. Hence, inference on $B(\alpha)$ can be carried out using its *support function* $\sigma(q,\alpha) \equiv \sup_{b\in B(\alpha)} q'b$, where $q \in \mathcal{S}^{d-1}$ is a direction in the unit sphere in d dimensions.[2] Beresteanu and Molinari (2008) and Bontemps, Magnac, and Maurin (2010) previously proposed the use of the support function as a key tool to conduct inference in best linear approximations to conditional expectation functions. An application of their results gives that the support function of $B(\alpha)$ is equal to the expectation of a function of $(\theta_0(x,\alpha), \theta_1(x,\alpha), x, \mathrm{E}(xx'))$. Hence, an application of the analogy principle suggests to estimate $\sigma(q,\alpha)$ through a sample average of the same function, where $\theta_0(x,\alpha)$ and $\theta_1(x,\alpha)$ are replaced by parametric or non-parametric estimators, and $E(xx')$ is replaced by its sample analog. The resulting estimator, denoted $\widehat{\sigma}(q,\alpha)$, yields an estimator for $B(\alpha)$ through the characterization in equation (2.2) below. We show that $\widehat{\sigma}(q,\alpha)$ is a consistent estimator for $\sigma(q,\alpha)$, uniformly over $q,\alpha \in \mathcal{S}^{d-1} \times \mathcal{A}$. We then establish the approximate asymptotic Gaussianity of our set estimator. Specifically, we prove that when properly recentered and normalized, $\widehat{\sigma}(q,\alpha)$ approximately converges to a Gaussian process on $\mathcal{S}^{d-1} \times \mathcal{A}$ (we explain below what we mean by "approximately"). The covariance function of this process is quite complicated, so we propose the use of a Bayesian bootstrap procedure for practical inference, and we prove consistency of this bootstrap procedure.

Because the support function process converges on $\mathcal{S}^{d-1} \times \mathcal{A}$, our asymptotic results also allow us to perform inference on statistics that involve a continuum of values for q and/or α. For example, for best linear approximations to conditional quantile functions in the presence of interval outcome data, we are able to test whether a given regressor x_j has a positive effect on the conditional α-quantile for any $\alpha \in \mathcal{A}$.

In providing a methodology for inference, our paper overcomes significant technical complications, thereby making contributions of independent interest. First, we allow for the possibility that some of the regressors in x have a discrete distribution. In order to conduct test of hypothesis and make confidence statements, both Beresteanu and Molinari (2008) and Bontemps, Magnac, and Maurin (2010) had explicitly ruled out discrete regressors, as their presence greatly complicates the derivation of the limiting distribution of the support function process. By using a simple data-jittering technique, we show that these complications completely disappear, albeit at the cost of basing statistical inference on a slightly conservative confidence set.

[2]"The support function (of a nonempty closed convex set B in direction q) $\sigma^B(q)$ is the signed distance of the support plane to B with exterior normal vector q from the origin; the distance is negative if and only if q points into the open half space containing the origin," Schneider (1993, page 37). See Rockafellar (1970, Chapter 13) or Schneider (1993, Section 1.7) for a thorough discussion of the support function of a closed convex set and its properties.

Second, when $\theta_0(x, \alpha)$ and $\theta_1(x, \alpha)$ are non-parametrically estimated through series estimators, we show that the support function process is approximated by a Gaussian process that may not necessarily converge as the number of series functions increases to infinity. To solve this difficulty, we use a strong approximation argument and show that each subsequence has a further subsequence converging to a tight Gaussian process with a uniformly equicontinuous and non-degenerate covariance function. We can then conduct inference using the properties of the covariance function.

To illustrate the use of our estimator, we revisit the analysis of Mulligan and Rubinstein (2008). The literature studying female labor force participation has argued that the gender wage gap has shrunk between 1975 and 2001. Mulligan and Rubinstein (2008) suggest that women's wages may have grown less than men's wages between 1975 and 2001, had their behavior been held constant, but a selection effect induces the data to show the gender wage gap contracting. They point out that a growing wage inequality within gender induces women to invest more in their market productivity. In turn, this would differentially pull high skilled women into the workplace and the selection effect may make it appear as if cross-gender wage inequality had declined.

To test this conjecture they employ a Heckman selection model to correct married women's conditional mean wages for selectivity and investment biases. Using CPS repeated cross-sections from 1975-2001 they argue that the selection of women into the labor market has changed sign, from negative to positive, or at least that positive selectivity bias has come to overwhelm investment bias. Specifically, they find that the gender wage gap measured by OLS decreased from -0.419 in 1975-1979 to -0.256 in 1995-1999. After correcting for selection using the classic Heckman selection model, they find that the wage gap was -0.379 in 1975-1979 and -0.358 in 1995-1999, thereby concluding that correcting for selection, the gender wage gap may have not shrunk at all. Because it is well known that without a strong exclusion restriction results of the normal selection model can be unreliable, Mulligan and Rubinstein conduct a sensitivity analysis which corroborates their findings.

We provide an alternative approach. We use our method to estimate bounds on the quantile gender wage gap without assuming a parametric form of selection or a strong exclusion restriction. We are unable to reject that the gender wage gap declined over the period in question. This suggests that the instruments may not be sufficiently strong to yield tight bounds and that there may not be enough information in the data to conclude that the gender gap has or has not declined from 1975 to 1999 without strong functional form assumptions.

Related Literature. This paper contributes to a growing literature on inference on set-identified parameters. Important examples in the literature include Andrews and Jia (2008), Andrews and Shi (2009), Andrews and Soares (2010), Beresteanu and Molinari (2008), Bontemps, Magnac, and Maurin (2010), Bugni (2010), Canay (2010), Chernozhukov, Hong, and Tamer (2007), Chernozhukov, Lee, and Rosen (2009), Galichon and Henry (2009), Kaido (2010), Romano and Shaikh (2008), Romano and Shaikh (2010), and Rosen (2008), among others. Beresteanu and Molinari (2008) propose an approach for estimation and inference for a class of partially identified models with convex identification region based on results from random set theory. Specifically, they consider models where the identification region is equal to the Aumann expectation of a properly defined random set that can be constructed from observable random variables. Extending the analogy principle, Beresteanu and Molinari suggest to estimate the Aumann expectation using a Minkowski average of random sets. Building on the fundamental insight in random set theory that convex compact sets can be represented via their support functions (see, e.g., Artstein and Vitale (1975)), Beresteanu and Molinari accordingly derive asymptotic properties of set estimators using limit theory for stochastic processes. Bontemps, Magnac, and Maurin (2010) extend the results of Beresteanu and Molinari (2008) in important directions, by allowing for incomplete linear moment restrictions where the number of restrictions exceeds the number of parameters to be estimated, and extend the familiar Sargan test for overidentifying restrictions to partially identified models. Kaido (2010) establishes a duality between the criterion function approach proposed by Chernozhukov, Hong, and Tamer (2007), and the support function approach proposed by Beresteanu and Molinari (2008).

Concurrently and independently of our work, Kline and Santos (2010) study the sensitivity of empirical conclusions about conditional quantile functions to the presence of missing outcome data, when the Kolmogorov-Smirnov distance between the conditional distribution of observed outcomes and the conditional distribution of missing outcomes is bounded by some constant k across all values of the covariates. Under these assumptions, Kline and Santos show that the conditional quantile function is sandwiched between a lower and an upper band, indexed by the level of the quantile and the constant k. To conduct inference, they assume that the support of the covariates is finite, so that the lower and upper bands can be estimated at parametric rates. Kline and Santos' framework is a special case of our sample selection example listed above. Hence, our results significantly extend the scope of Kline and Santos' analysis, by allowing for continuous regressors. Moreover, the method proposed in this paper allows for the upper and lower bounds to be non-parametrically estimated by series estimators, and allows the researcher to utilize instruments. While technically challenging, allowing for non-parametric estimates of the bounding functions and for intersection bounds considerably expands the domain of applicability of our results.

Structure of the Paper. This paper is organized as follows. In Section 2 we develop our framework, and in Section 3 we demonstrate its versatility by applying it to quantile regression, distribution regression, sample selection problems, and treatment effects. Section 4 provides an overview of our theoretical results and describes the estimation and inference procedures. Section 5 gives the empirical example. Section 6 concludes. All proofs are in the Appendix.

2. The General Framework

We aim at conducting inference for best linear approximations to the set of functions

$$\mathfrak{F}(\alpha) = \{\phi(\cdot,\alpha) : \theta_0(x,\alpha) \le \phi(x,\alpha) \le \theta_1(x,\alpha) \; x-a.s.\}$$

Here, $\alpha \in \mathcal{A}$ is some index with $\mathcal{A}$ a compact set, and x is a column vector in $\mathbb{R}^d$. For example, in quantile regression α denotes a quantile; in duration regression α denotes a failure time. For each x, $\theta_0(x,\alpha)$ and $\theta_1(x,\alpha)$ are point-identified lower and upper bounds on a true but non-point-identified function of interest $f(x,\alpha)$. If $f(x,\alpha)$ were point identified, we could approximate it with a linear function by choosing coefficients $\beta(\alpha)$ to minimize the expected squared prediction error $\mathrm{E}[(f(x,\alpha)-x'\beta(\alpha))^2]$. Because $f(x,\alpha)$ is only known to lie in $\mathfrak{F}(\alpha)$, performing this operation for each admissible function $\phi(\cdot,\alpha) \in \mathfrak{F}(\alpha)$ yields a set of observationally equivalent parameter vectors, denoted $B(\alpha)$:

$$\begin{aligned} B(\alpha) =& \{\beta \in \arg\min_b \mathrm{E}[(\phi(x,\alpha)-x'b)^2] : \mathrm{P}(\theta_0(x,\alpha) \le \phi(x,\alpha) \le \theta_1(x,\alpha)) = 1\} \\ =& \{\beta = \mathrm{E}[xx']^{-1}\mathrm{E}[x\phi(x,\alpha)] : \mathrm{P}(\theta_0(x,\alpha) \le \phi(x,\alpha) \le \theta_1(x,\alpha)) = 1\}. \end{aligned} \tag{2.1}$$

It is easy to see that the set $B(\alpha)$ is almost surely non-empty, compact, and convex valued, because it is obtained by applying linear operators to the (random) almost surely non-empty interval $[\theta_0(x,\alpha), \theta_1(x,\alpha)]$, see Beresteanu and Molinari (2008, Section 4) for a discussion. Hence, $B(\alpha)$ can be characterized quite easily through its support function

$$\sigma(q,\alpha) := \sup_{\beta(\alpha)\in B(\alpha)} q'\beta(\alpha),$$

which takes on almost surely finite values $\forall q \in \mathcal{S}^{d-1} := \left\{q \in \mathbb{R}^d : \|q\| = 1\right\}$, $d = \dim\beta$. In fact,

$$B(\alpha) = \bigcap_{q\in\mathcal{S}^{d-1}} \left\{b : q'b \le \sigma(q,\alpha)\right\}, \tag{2.2}$$

see Rockafellar (1970, Chapter 13). Note also that $[-\sigma(-q,\alpha), \sigma(q,\alpha)]$ gives sharp bounds on the linear combination of $\beta(\alpha)$'s components obtained using weighting vector q.

More generally, if the criterion for "best" linear approximation is to minimize $\mathrm{E}[(f(x,\alpha)-x'\beta(\alpha))\tilde{z}'W\tilde{z}(f(x,\alpha)-x'\beta(\alpha))]$, where W is a $j\times j$ weight matrix and $\tilde{z}$ a $j\times 1$ vector of instruments, then we have

$$B(\alpha) = \{\beta = \mathrm{E}[x\tilde{z}'W\tilde{z}x']^{-1}\mathrm{E}[x\tilde{z}'W\tilde{z}\phi(x,\alpha)] : \mathrm{P}(\theta_0(x,\alpha) \le \phi(x,\alpha) \le \theta_1(x,\alpha)) = 1\}.$$

As in Bontemps, Magnac, and Maurin (2010), Magnac and Maurin (2008), and Beresteanu and Molinari (2008, p. 807) the support function of $B(\alpha)$ can be shown to be

$$\sigma(q,\alpha) = \mathrm{E}[z_q w_q]$$

where

$$\begin{aligned} z &= x\tilde{z}'W\tilde{z},\ z_q = q'\mathrm{E}[xz']^{-1}z, \\ w_q &= \theta_1(x,\alpha)1(z_q > 0) + \theta_0(x,\alpha)1(z_q \le 0). \end{aligned}$$

We estimate the support function by plugging in estimates of θ_ℓ, $\ell = 0,1$, and taking empirical expectations:

$$\widehat{\sigma}(q,\alpha) = \mathbb{E}_n\left[q'\left(\mathbb{E}_n\left[x_i z_i'\right]\right)^{-1} z_i\left(\widehat{\theta}_1(x_i,\alpha)1(\widehat{z}_{iq} > 0) + \widehat{\theta}_0(x_i,\alpha)1(\widehat{z}_{iq} \le 0)\right)\right],$$

where $\mathbb{E}_n$ denotes the empirical expectation, $\widehat{z}_{iq} = q'\left(\mathbb{E}_n\left[x_i z_i'\right]\right)^{-1} z$, and $\widehat{\theta}_\ell(x,\alpha)$ are the estimators of $\theta_\ell(x,\alpha)$, $\ell = 0,1$.

3. Motivating Examples

3.1. **Interval valued data.** Analysis of regression with interval valued data has become a canonical example in the partial identification literature. Interest in this example stems from the fact that dealing with interval data is a commonplace problem in empirical work. Due to concerns for privacy, survey data often come in bracketed form. For example, public use tax data are recorded as the number of tax payers which belong to each of a finite number of cells, as seen in Picketty (2005). The Health and Retirement Study provides a finite number of income brackets to each of its respondents, with degenerate brackets for individuals who opt to fully reveal their income level; see Juster and Suzman (1995) for a description. The Occupational Employment Statistics (OES) program at the Bureau of Labor Statistics collects wage data from employers as intervals, and uses these data to construct estimates for wage and salary workers in 22 major occupational groups and 801 detailed occupations.[3]

3.1.1. *Interval valued y.* Beresteanu and Molinari (2008) and Bontemps, Magnac, and Maurin (2010), among others, have focused on estimation of best linear approximations to conditional expectation functions with interval valued outcome data. Our framework covers the conditional expectation case, as well as an extension to quantile regression wherein we set identify $\beta(\alpha)$ across all quantiles $\alpha \in \mathcal{A}$. To avoid redundancy with the related literature, here we describe the setup for quantile regression. Let the α-th conditional quantile of $y|x$ be denoted $Q_y(\alpha|x)$. We are interested in a linear approximation $x'\beta(\alpha)$ to this function. However, we do not observe y. Instead we observe y_0 and y_1, with $\mathrm{P}(y_0 \leq y \leq y_1) = 1$. It is immediate that

$$Q_{y_0}(\alpha|x) \leq Q_y(\alpha|x) \leq Q_{y_1}(\alpha|x) \quad x - a.s.,$$

where $Q_{y_\ell}(\alpha|x)$ is the α-th conditional quantile of $y_\ell|x$, $\ell = 0, 1$. Hence, the identification region $B(\alpha)$ is as in equation (2.1), with $\theta_\ell(\alpha, x) = Q_{y_\ell}(\alpha|x)$.

3.1.2. *Interval valued x_i.* Suppose now that one is interested in the conditional expectation $\mathrm{E}(y|x)$, but only observes y and variables x_0, x_1 such that $\mathrm{P}(x_0 \leq x \leq x_1) = 1$. This may occur, for example, when education data is binned into categories such as primary school, secondary school, college, and graduate education, while the researcher may be interested in the return to each year of schooling. It also happens when a researcher is interested in a model in which wealth is a covariate, but the available survey data report it by intervals.

Our approach applies to the framework of Manski and Tamer (2002) for conditional expectation with interval regressors, and extends it to the case of quantile regression.[4] Following Manski and Tamer, we assume that the conditional expectation of $y|x$ is (weakly) monotonic in x, say nondecreasing, and mean independent of x_0, x_1 conditional on x. Manski

[3]See Manski and Tamer (2002) and Bontemps, Magnac, and Maurin (2010) for more examples.

[4]Our methods also apply to the framework of Magnac and Maurin (2008), who study identification in semi-parametric binary regression models with regressors that are either discrete or measured by intervals. Compared to Manski and Tamer (2002), Magnac and Maurin's analysis requires an uncorrelated error assumption, a conditional independence assumption between error and interval/discrete valued regressor, and a finite support assumption.

and Tamer show that

$$\sup_{x_1 \leq x} \mathrm{E}\left(y|x_0, x_1\right) \leq \mathrm{E}\left(y|x\right) \leq \inf_{x_0 \geq x} \mathrm{E}\left(y|x_0, x_1\right).$$

Hence, the identification region $B(\alpha)$ is as in equation (2.1), with $\theta_0(\alpha, x) = \sup_{x_1 \leq x} \mathrm{E}\left(y|x_0, x_1\right)$ and $\theta_1(\alpha, x) = \inf_{x_0 \geq x} \mathrm{E}\left(y|x_0, x_1\right)$.

Next, suppose that the α-th conditional quantile of $y|x$ is monotonic in x, say nondecreasing, and that $Q_y(\alpha|x, x_0, x_1) = Q_y(\alpha|x)$. By the same reasoning as above,

$$\sup_{x_1 \leq x} Q_y\left(\alpha|x_0, x_1\right) \leq Q_y(\alpha|x) \leq \inf_{x_0 \geq x} Q_y\left(\alpha|x_0, x_1\right).$$

Hence, the identification region $B(\alpha)$ is as in equation (2.1), with $\theta_0(\alpha, x) = \sup_{x_1 \leq x} Q_y\left(\alpha|x_0, x_1\right)$ and $\theta_1(\alpha, x) = \inf_{x_0 \geq x} Q_y\left(\alpha|x_0, x_1\right)$.

3.2. **Distribution and duration regression with interval outcome data.** Researchers may also be interested in distribution regression with interval valued data. For instance, a proportional hazard model with time varying coefficients where the probability of failure conditional on survival may be dependent on covariates and coefficients that are indexed by time. More generally, we can consider models in which the conditional distribution of $y|x$ is given by

$$\mathrm{P}\left(y \leq \alpha|x\right) \equiv F_{y|x}(\alpha|x) = \Phi\left(f(\alpha, x)\right)$$

where $\Phi(.)$ is a known one-to-one link function. A special case of this class of models is the duration model, wherein we have $f(\alpha, x) = g(\alpha) + \gamma(x)$, where $g(.)$ is a monotonic function.

As in the quantile regression example, assume that we observe (y_0, y_1, x) with $\mathrm{P}\left(y_0 \leq y \leq y_1\right) = 1$. Then

$$\Phi^{-1}\left(F_{y_1|x}(\alpha|x)\right) \leq f(\alpha, x) \leq \Phi^{-1}\left(F_{y_0|x}(\alpha|x)\right).$$

Hence, the identification region $B(\alpha)$ is as in equation (2.1), with $\theta_\ell\left(\alpha, x\right) = \Phi^{-1}\left(F_{y_{1-\ell}|x}(\alpha|x)\right)$, $\ell = 0, 1$. A leading example, following Han and Hausman (1990) and Foresi and Peracchi (1995), would include Φ as a probit or logit link function.

3.3. **Sample Selection.** Sample selection is a well known first-order concern in the empirical analysis of important economic phenomena. Examples include labor force participation (see, e.g., Mulligan and Rubinstein (2008)), skill composition of immigrants (see, e.g., Jasso and Rosenzweig (2008)), returns to education (e.g., Card (1999)), program evaluation (e.g., Imbens and Wooldridge (2009)), productivity estimation (e.g., Olley and Pakes (1996)), insurance (e.g., Einav, Finkelstein, Ryan, Schrimpf, and Cullen (2011)), models with occupational choice and financial intermediation (e.g., Townsend and Urzua (2009)). In Section 5 we revisit the analysis of Mulligan and Rubinstein (2008) who confront selection in the context of female labor supply.

Consider a standard sample selection model. We are interested in the behavior of y conditional on x; however, we only observe y when $u = 1$. Manski (1989) observes that the following bound on the conditional distribution of y given x can be constructed:

$$F(y|x, u = 1)\mathrm{P}(u = 1|x) \leq F(y|x) \leq F(y|x, u = 1)\mathrm{P}(u = 1|x) + \mathrm{P}(u = 0|x).$$

The inverse image of these distribution bounds gives bounds on the conditional quantile function of y

$$Q_0(\alpha|x) = \begin{cases} Q_y\left(\frac{\alpha - \mathrm{P}(u=0|x)}{\mathrm{P}(u=1|x)}\middle| x, u=1\right) & \text{if } \alpha \geq \mathrm{P}(u=0|x) \\ y_0 & \text{otherwise} \end{cases}$$

$$Q_1(\alpha|x) = \begin{cases} Q_y\left(\frac{\alpha}{\mathrm{P}(u=1|x)}\middle| x, u=1\right) & \text{if } \alpha \leq \mathrm{P}(u=1|x) \\ y_1 & \text{otherwise} \end{cases}$$

where y_0 is the smallest possible value that y can take (possibly $-\infty$) and y_1 is the largest possible value that y can take (possibly $+\infty$). Thus, we obtain

$$Q_0(\alpha|x) \leq Q_y(\alpha|x) \leq Q_1(\alpha|x).$$

and the corresponding set of coefficients of linear approximations to $Q_y(\alpha|x)$ is as in equation (2.1), with $\theta_\ell(\alpha, x) = Q_\ell(\alpha|x)$, $\ell = 0, 1$.

3.3.1. *Alternative Form for the Bounds.* As written above, the expressions for $Q_0(\alpha|x)$ and $Q_1(\alpha|x)$ involve the propensity score, $\mathrm{P}(u|x)$ and several different conditional quantiles of $y|u = 1$. Estimating these objects might be computationally intensive. However, we show that Q_0 and Q_1 can also be written in terms of the α-th conditional quantile of a different random variable, thereby providing computational simplifications. Define

$$\tilde{y}_0 = y1\{u=1\} + y_0 1\{u=0\}, \qquad \tilde{y}_1 = y1\{u=1\} + y_1 1\{u=0\}. \tag{3.3}$$

Then one can easily verify that $Q_{\tilde{y}_0}(\alpha|x) = Q_0(\alpha|x)$, and $Q_{\tilde{y}_1}(\alpha|x) = Q_1(\alpha|x)$, and therefore the bounds on the conditional quantile function can be obtained without calculating the propensity score.

3.3.2. *Sample Selection With an Exclusion Restriction.* Notice that the above bounds let $F(y|x)$ and $\mathrm{P}(u = 1|x)$ depend on x arbitrarily. However, often when facing selection problems researchers impose exclusion restrictions. That is, the researcher assumes that there are some components of x that affect $\mathrm{P}(u = 1|x)$, but not $F(y|x)$. Availability of such an instrument, denoted v, can help shrink the bounds on $Q_y(\alpha|x)$. For concreteness, we replace x with (x, v) and suppose that $F(y|x, v) = F(y|x)$ $\forall v \in supp(v|x)$. By the same reasoning as above, for each $v \in supp(v|x)$ we have the following bounds on the conditional distribution function:

$$F(y|x, v, u=1)\mathrm{P}(u=1|x,v) \leq F(y|x) \leq F(y|x,v,u=1)\mathrm{P}(u=1|x,v) + \mathrm{P}(u=0|x,v).$$

This implies that the conditional quantile function satisfies:

$$Q_0(\alpha|x, v) \leq Q_y(\alpha|x) \leq Q_1(\alpha|x, v) \ \forall v \in supp(v|x),$$

and therefore

$$\sup_{v \in supp(v|x)} Q_0(\alpha|x,v) \leq Q_y(\alpha|x) \leq \inf_{v \in supp(v|x)} Q_1(\alpha|x,v)$$

where $Q_\ell(\alpha|x, v)$, $\ell = 0, 1$, are defined similarly to the previous section with x replaced by (x, v). Once again, we can avoid computing the propensity score by constructing the variables $\tilde{y}_\ell$, $\ell = 0, 1$ as in equation (3.3). Then $Q_{\tilde{y}_\ell}(\alpha|x, v) = Q_\ell(\alpha|x, v)$. Inspecting the formulae for $Q_\ell(\alpha|x, v)$, $\ell = 0, 1$, reveals that $Q_0(\alpha|x, v)$ can only be greater than y_0 when

$1-\mathrm{P}(u=1|x,v)<\alpha$, and $Q_1(\alpha|x,v)$ can only be smaller than y_1 when $\alpha<\mathrm{P}(u=1|x,v)$. Thus, both bounds are informative only when

$$1-\mathrm{P}(u=1|x,v)<\alpha<\mathrm{P}(u=1|x,v).$$

From this we see that the bounds are more informative for central quantiles than extreme ones. Also, the greater the probability of being selected conditional on x, the more informative the bounds are. If $P(u=1|x,v)=1$ for some $v\in supp(v|x)$ then $Q_y(\alpha|x)$ is point-identified. This is the large-support condition required to show non-parametric identification in selection models. If $\mathrm{P}(u=1|x,v)<1/2$ the upper and lower bounds cannot both be informative.

It is important to note that $Q_\ell(\alpha|x,v)$, $\ell=0,1$ depend on the quantiles of y conditional on both x and v. Moreover, $Q_y(\alpha|x,v,u=1)$ is generally not linear in x and v, even in the special cases where $Q_y(\alpha|x)$ is linear in x. Therefore, in practice, it is important to estimate $Q_y(\alpha|x,v,u=1)$ flexibly. Accordingly, our asymptotic results allow for series estimation of the bounding functions.

3.4. **Average, Quantile, and Distribution Treatment Effects.** Researchers are often interested in mean, quantile, and distributional treatment effects. Our framework easily accommodates these examples. Let y_i^C denote the outcome for person i if she does not receive treatment, and y_i^T denote the outcome for person i if she receives treatment. The methods discussed in the preceding section yield bounds on the conditional quantiles of these outcomes. In turn, these bounds can be used to obtain bounds on the quantile treatment effect as follows:

$$\begin{aligned}\sup_{v\in supp(v|x)} Q_0^T(\alpha|x,v) \;&-\; \inf_{v\in supp(v|x)} Q_1^C(\alpha|x,v)\\ &\leq\; Q_{Y^T}(\alpha|x)-Q_{Y^C}(\alpha|x)\\ &\leq\; \inf_{v\in supp(v|x)} Q_1^T(\alpha|x,v)-\sup_{v\in supp(v|x)} Q_0^C(\alpha|x,v).\end{aligned}$$

As we discussed in the previous section, $Q_0^T(\alpha|x,v)$ and $Q_1^T(\alpha|x,v)$ are both informative only when

$$1-P(u=1|x,v)<\alpha<P(u=1|x,v). \tag{3.4}$$

Similarly, $Q_0^C(\alpha|x,v)$ and $Q_1^C(\alpha|x,v)$ are both informative only when

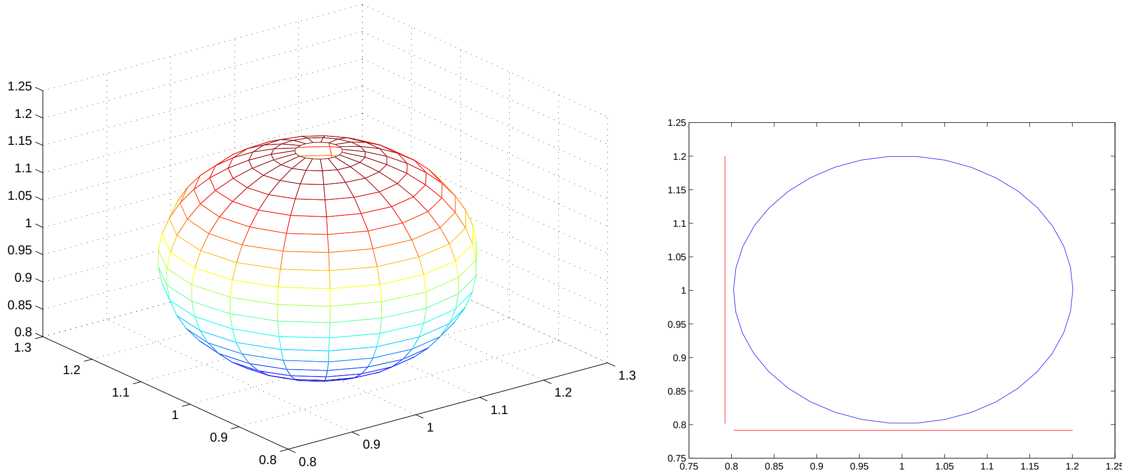
$$P(u=1|x,v)<\alpha<1-P(u=1|x,v). \tag{3.5}$$

Note that inequalities (3.4) and (3.5) cannot both hold. Thus, we cannot obtain informative bounds on the quantile treatment effect without an exclusion restriction. Moreover, to have an informative upper and lower bound on $Q_{Y^T}(\alpha|x)-Q_{Y^C}(\alpha|x)$, the excluded variables, v, must shift the probability of treatment, $P(u=1|x,v)$ sufficiently for both (3.4) and (3.5) to hold at x (for different values of v).

Analogous bounds apply for the distribution treatment effect and the mean treatment effect.[5]

[5]Interval regressors can also be accommodated, by merging the results in this Section with those in Section 3.1.

Figure 1. Identification region and its projection



4. Estimation and Inference

4.1. **Overview of the Results.** This section provides a less technically demanding overview of our results, and explains how these can be applied in practice. Throughout, we use sample selection as our primary illustrative example. As described in section 2, our goal is to estimate the support function, $\sigma(q,\alpha)$. The support function provides a convenient way to compute projections of the identified set. These can be used to report upper and lower bounds on individual coefficients and draw two-dimensional identification regions for pairs of coefficients. For example, the bound for the kth component of $\beta(\alpha)$ is $[-\sigma(-e_k,\alpha),\sigma(e_k,\alpha)]$, where e_k is the kth standard basis vector (a vector of all zeros, except for a one in the kth position). Similarly, the bound for a linear combination of the coefficients, $q'\beta(\alpha)$, is $[-\sigma(-q,\alpha),\sigma(q,\alpha)]$. Figures 1 provides an illustration. In this example, β is three dimensional. The left panel shows the entire identified set. The right panel shows the joint identification region for β_1 and β_2. The identified intervals for β_1 and β_2 are also marked in red on the right panel.

Suppose that $x=[1;x_1]$, with x_1 a scalar random variable, so $\beta(\alpha)=\begin{bmatrix}\beta_0(\alpha) & \beta_1(\alpha)\end{bmatrix}$ is two-dimensional. To simplify notation, let $z=x$. In most applications, $\beta_1(\alpha)$ is the primary object of interest. Stoye (2007), Beresteanu and Molinari (2008) and Bontemps, Magnac, and Maurin (2010) give explicit formulae for the upper and lower bound of $\beta_1(\alpha)$. Recall that the support function is given by:

$$\sigma(q)=q'\mathrm{E}[xx']^{-1}\mathrm{E}\left[x\left(\theta_1(x,\alpha)1\left\{q'\mathrm{E}[xx']^{-1}x>0\right\}+\theta_0(x,\alpha)1\left\{q'\mathrm{E}[xx']^{-1}x<0\right\}\right)\right]$$

Setting $q=\begin{pmatrix}0 & \pm1\end{pmatrix}$ yields these bounds as follows:

$$\underline{\beta}_1(\alpha)=\frac{\mathrm{E}\left[(x_{1i}-\mathrm{E}[x_{1i}])\left(\theta_{1i}1\left\{x_{1i}<\mathrm{E}[x_{1i}]\right\}+\theta_{0i}1\left\{x_{1i}>\mathrm{E}[x_{1i}]\right\}\right)\right]}{\mathrm{E}[x_{1i}^2]-\mathrm{E}[x_{1i}]^2}$$

$$\overline{\beta}_1(\alpha)=\frac{\mathrm{E}\left[(x_{1i}-\mathrm{E}[x_{1i}])\left(\theta_{1i}1\left\{x_{1i}>\mathrm{E}[x_{1i}]\right\}+\theta_{0i}1\left\{x_{1i}<\mathrm{E}[x_{1i}]\right\}\right)\right]}{\mathrm{E}[x_{1i}^2]-\mathrm{E}[x_{1i}]^2}$$

where $\theta_{0i} = \theta_0(x_i, \alpha)$ and $\theta_{1i} = \theta_1(x_i, \alpha)$.[6]

4.1.1. *Use of asymptotic results.* We develop limit theory that allows us to derive the asymptotic distribution of the support function process and provide inferential procedures, as well as to establish validity of the Bayesian bootstrap. Bootstrapping is especially important for practitioners, because of the potential complexity of the covariance functions involved in the limiting distributions.

First, our limit theory shows that the support function process $S_n(t) := \sqrt{n}\left(\widehat{\sigma}(t) - \sigma_0(t)\right)$ for $t \in \mathcal{S}^{d-1} \times \mathcal{A}$ is approximately distributed as a Gaussian process on $\mathcal{S}^{d-1} \times \mathcal{A}$. Specifically, we have that

$$S_n(t) = \mathbb{G}\left[h_k\left(t\right)\right] + o_{\mathrm{P}}\left(1\right)$$

in $\ell^\infty\left(T\right)$, where k denotes the number of series terms in our non-parametric estimator of $\theta_\ell(x, \alpha)$, $\ell = 0, 1$, $\ell^\infty\left(T\right)$ denotes the set of all uniformly bounded real functions on T, and $h_k\left(t\right)$ denotes a stochastic process carefully defined in Section 4.3. Here, $\mathbb{G}\left[h_k\left(t\right)\right]$ is a tight P-Brownian bridge with covariance function $\Omega_k\left(t, t'\right) = \mathrm{E}\left[h_k\left(t\right) h_k\left(t'\right)\right] - \mathrm{E}\left[h_k\left(t\right)\right]\mathrm{E}\left[h_k\left(t'\right)\right]$. By "approximately distributed" we mean that the sequence $\mathbb{G}[h_k(t)]$ does not necessarily converge weakly when $k \to \infty$; however, each subsequence has a further subsequence converging to a tight Gaussian process in $\ell^\infty(T)$ with a non-degenerate covariance function.

Second, we show that inference is possible by using the quantiles of the limiting distribution $\mathbb{G}\left[h_k\left(t\right)\right]$. Specifically, if we have a continuous function f that satisfies certain (non-restrictive) conditions detailed in Section 4.3,[7] and $\widehat{c}_n\left(1-\tau\right) = c_n\left(1-\tau\right) + o_{\mathrm{P}}\left(1\right)$ is a consistent estimator of the $(1-\tau)$-quantile of $f\left(\mathbb{G}\left[h_k\left(t\right)\right]\right)$, given by $c_n\left(1-\tau\right)$, then

$$\mathrm{P}\{f(S_n) \leq \widehat{c}_n(1-\tau)\} \to 1-\tau.$$

Finally, we consider the limiting distribution of the Bayesian bootstrap version of the support function process, denoted $\tilde{S}_n(t) := \sqrt{n}\left(\tilde{\sigma}(t) - \widehat{\sigma}(t)\right)$, and show that, conditional on the data, it admits an approximation

$$\tilde{S}_n\left(t\right) = \widetilde{\mathbb{G}\left[h_k\left(t\right)\right]} + o_{\mathrm{P}^e}\left(1\right)$$

where $\widetilde{\mathbb{G}\left[h_k\left(t\right)\right]}$ has the same distribution as $\mathbb{G}\left[h_k\left(t\right)\right]$ and is independent of $\mathbb{G}\left[h_k\left(t\right)\right]$. Since the bootstrap distribution is asymptotically close to the true distribution of interest, this allows us to perform many standard and some less standard inferential tasks.

[6] As one would expect from the definition of the support function and the properties of linear projection,

$$\underline{\beta}_1(\alpha) = \inf_{f_i \in [\theta_{i0}, \theta_{i1}]} \frac{\mathrm{cov}(x_{1i}, f_i)}{\mathrm{var}(x_{i1})},$$
$$\overline{\beta}_1(\alpha) = \sup_{f_i \in [\theta_{i0}, \theta_{i1}]} \frac{\mathrm{cov}(x_{1i}, f_i)}{\mathrm{var}(x_{i1})}.$$

[7] For example, functions yielding test statistics based on the directed Hausdorff distance and on the Hausdorff distance (see, e.g., Beresteanu and Molinari (2008)) belong to this class.

Pointwise asymptotics. Suppose we want to form a confidence interval for $q'\beta(\alpha)$ for some fixed q and α. Since our estimator converges to a Gaussian process, we know that

$$\sqrt{n}\begin{pmatrix} -\widehat{\sigma}(-q,\alpha)+\sigma_0(-q,\alpha) \\ \widehat{\sigma}(q,\alpha)-\sigma_0(q,\alpha) \end{pmatrix} \approx_d N\left(0,\Omega(q,\alpha)\right).$$

To form a confidence interval that covers the bound on $q'\beta(\alpha)$ with probability $1-\tau$ we can take[8]

$$-\widehat{\sigma}(-q,\alpha)+n^{1/2}\widehat{C}_{\tau/2}(q,\alpha) \leq q'\beta(\alpha) \leq \widehat{\sigma}(q,\alpha)+n^{-1/2}\widehat{C}_{1-\tau/2}(q,\alpha)$$

where the critical values, $\widehat{C}_{\tau/2}(q,\alpha)$ and $\widehat{C}_{1-\tau/2}(q,\alpha)$, are such that if $\begin{pmatrix} x_1 & x_2 \end{pmatrix}' \sim N(0,\Omega)$, then

$$\mathrm{P}\left(x_1 \geq \widehat{C}_{\tau/2}(q,\alpha)\,,\, x_2 \leq \widehat{C}_{1-\tau/2}(q,\alpha)\right) = 1-\tau+o_p(1)$$

If we had a consistent estimate of $\Omega(q,\alpha)$, then we could take

$$\begin{pmatrix} \widehat{C}_{\tau/2}(q,\alpha) \\ \widehat{C}_{1-\tau/2}(q,\alpha) \end{pmatrix} = \widehat{\Omega}^{1/2}(q,\alpha)\begin{pmatrix} -\Phi^{-1}(\sqrt{1-\tau}) \\ \Phi^{-1}(\sqrt{1-\tau}) \end{pmatrix}$$

where $\Phi^{-1}(\cdot)$ is the inverse normal distribution function. However, the formula for $\Omega(q,\alpha)$ is complicated and it can be difficult to estimate. Therefore, we recommend and provide theoretical justification for using a Bayesian bootstrap procedure to estimate the critical values. See section 4.1.3 for details.

Functional asymptotics. Since our asymptotic results show the functional convergence of $S_n(q,\alpha)$, we can also perform inference on statistics that involve a continuum of values of q and/or α. For example, in our application to quantile regression with selectively observed data, we might be interested in whether a particular variable has a positive affect on the outcome distribution. That is, we may want to test

$$H_0 : 0 \in [-\sigma_0(-q,\alpha),\sigma_0(q,\alpha)]\ \forall \alpha \in \mathcal{A},$$

with $q = e_j$. A natural family of test statistics is

$$T_n = \sqrt{n}\sup_{\alpha\in\mathcal{A}}\begin{pmatrix} 1\{-\widehat{\sigma}(-q,\alpha)>0\}|\widehat{\sigma}(-q,\alpha)|\rho(-q,\alpha)\vee \\ \vee\{\widehat{\sigma}(q,\alpha)<0\}|\widehat{\sigma}(q,\alpha)|\rho(q,\alpha) \end{pmatrix}$$

where $\rho(q,\alpha) \geq 0$ is some weighting function which can be chosen to maximize weighted power against some family of alternatives. There are many values of $\sigma_0(q,\alpha)$ consistent with the null hypothesis, but the one for which it will be hardest to control size is $-\sigma_0(-q,\cdot) = \sigma_0(q,\cdot) = 0$. In this case, we know that $S_n(t) = \sqrt{n}\widehat{\sigma}(t)$, $t = (q,\alpha) \in \mathcal{S}^{d-1}\times\mathcal{A}$, is well approximated by a Gaussian process, $\mathbb{G}[h_k(t)]$. Moreover, the quantiles of any functional of $S_n(t)$ converge to the quantiles of the same functional applied to $\mathbb{G}[h_k(t)]$. Thus, we could calculate a τ critical value for T_n by repeatedly simulating a realization of $\mathbb{G}[h_k(q,\cdot)]$, computing $T_n(\mathbb{G}[h_k(q,\cdot)])$, and then taking the $(1-\tau)$-quantile of the simulated values of

[8]Instead, if one believes there is some true value, $q'\beta_0(\alpha)$, in the identified set, and one wants to cover this true value (uniformly) with asymptotic probability $1-\tau$, then one can adopt the procedures of Imbens and Manski (2004) and Stoye (2009), see also Bontemps, Magnac, and Maurin (2010) for related applications.

$T_n(\mathbb{G}[h_k(q,\cdot)])$.[9] Simulating $\mathbb{G}[h_k(t)]$ requires estimating the covariance function. As stated above, the formula for this function is complicated and it can be difficult to estimate. Therefore, we recommend using the Bayesian bootstrap to compute the critical values. Theorem 4 proves that this bootstrap procedure yields consistent inference. Section 4.1.3 gives a more detailed outline of implementing this bootstrap. Similar reasoning can be used to test hypotheses involving a set of values of q and construct confidence sets that are uniform in q and/or α.

4.1.2. *Estimation.* The first step in estimating the support function is to estimate $\theta_0(x,\alpha)$ and $\theta_1(x,\alpha)$. Since economic theory often provides even less guidance about the functional form of these bounding functions than it might about the function of interest, our asymptotic results are written to accommodate non-parametric estimates of $\theta_0(x,\alpha)$ and $\theta_1(x,\alpha)$. In particular, we allow for series estimators of these functions. In this section we briefly review this approach. Parametric estimation follows as a special case where the number of series terms is fixed. Note that while the method of series estimation described here satisfies the conditions of theorems 1 and 2 below, there might be other suitable methods of estimation for the bounding functions.

In each of the examples in section 3, except for the case of sample selection with an exclusion restriction, series estimates of the bounding functions can be formed as follows. Suppose there is some $y_{i\ell}$, a known function of the data for observation i, and a known function $m(y,\theta(x,\alpha),\alpha)$ such that

$$\theta_\ell(\cdot,\alpha) = \underset{\theta\in\mathcal{L}^2(X,\mathrm{P})}{\arg\min}\ \mathrm{E}\left[m\left(y_{i\ell},\theta(x_i,\alpha),\alpha\right)\right],$$

where X denotes the support of x and $\mathcal{L}^2(X,\mathrm{P})$ denotes the space of real-valued functions g such that $\int_X |g(x)|^2\, d\mathrm{P}(x) < \infty$. Then we can form an estimate of the function $\theta_\ell(\cdot,\alpha)$ by replacing it with its series expansion and taking the empirical expectation in the equation above. That is, obtaining the coefficients

$$\widehat{\vartheta}_\ell(\alpha) = \arg\min_{\vartheta}\mathbb{E}_n\left[m\left(y_{i\ell}, p_k(x_i)'\vartheta,\alpha\right)\right],$$

and setting

$$\widehat{\theta}_l(x_i,\alpha) = p_k(x_i)'\widehat{\vartheta}_\ell(\alpha).$$

Here, $p_k(x_i)$ is a $k\times 1$ vector of k series functions evaluated at x_i. These could be any set of functions that span the space in which $\theta_\ell(x,\alpha)$ is contained. Typical examples include polynomials, splines, and trigonometric functions, see Chen (2007). Both the properties of $m(\cdot)$ and the choice of approximating functions affect the rate at which k can grow. We discuss this issue in more detail after stating our regularity conditions in section 4.2.

In the case of sample selection with an exclusion, one can proceed as follows. First, estimate $Q_{\tilde{y}_\ell}(\alpha|x,v)$, $\ell=0,1$, using the method described above. Next, set $\widehat{\theta}_l(x_i,\alpha) = \min_{v\in supp(v|x)} Q_{\tilde{y}_\ell}(\alpha|x,v)$. Below we establish the validity of our results also for this case.

[9]This procedure yields a test with correct asymptotic size. However, it might have poor power properties in some cases. In particular, when $\sigma_0(-q,\alpha)\neq\sigma_0(q,\alpha)$, the critical values might be too conservative. One can improve the power properties of the test by applying the generalized moment selection procedure proposed by Andrews and Shi (2009) to our framework.

4.1.3. *Bayesian Bootstrap* . We suggest using the Bayesian Bootstrap to conduct inference. In particular, we propose the following algorithm.

Procedure for Bayesian Bootstrap Estimation.

(1) Simulate each bootstrap draw of $\tilde{\sigma}(q,\alpha)$:
   (a) Draw $e_i \sim \exp(1)$, $i = 1, ..., n$, $\bar{e} = \mathbb{E}_n[e_i]$
   (b) Estimate:

$$\begin{aligned}
\tilde{\vartheta}_\ell &= \arg\min_{\vartheta} \mathbb{E}_n \left[\frac{e_i}{\bar{e}} m\left(y_{i\ell}, p_k(x_i)'\vartheta, \alpha\right)\right], \\
\tilde{\theta}_\ell(x,\alpha) &= p_k(x)'\tilde{\vartheta}_\ell, \\
\tilde{\Sigma} &= \mathbb{E}_n\left[\frac{e_i}{\bar{e}} x_i z_i'\right]^{-1}, \\
\tilde{w}_{i,q'\tilde{\Sigma}} &= \tilde{\theta}_1(x,\alpha)1(q'\tilde{\Sigma}z > 0) + \tilde{\theta}_0(x,\alpha)1(q'\tilde{\Sigma}z \le 0), \\
\tilde{\sigma}(q,\alpha) &= \mathbb{E}_n\left[\frac{e_i}{\bar{e}} q'\tilde{\Sigma} z_i \tilde{w}_{i,q'\tilde{\Sigma}}\right].
\end{aligned}$$

(2) Denote the bootstrap draws as $\tilde{\sigma}^{(b)}$, $b = 1, ..., B$, and let $\tilde{S}_n^{(b)} = \sqrt{n}(\tilde{\sigma}^{(b)} - \widehat{\sigma})$. To estimate the $1-\tau$ quantile of $\mathcal{T}(S_n)$ use the empirical $1-\tau$ quantile of the sample $\mathcal{T}(\tilde{S}_n^{(b)})$, $b = 1, ..., B$
(3) Confidence intervals for linear combinations of coefficients can be obtained as outlined in Section 4.1.1. Inference on statistics that involve a continuum of values of q and/or α can be obtained as outlined in Section 4.1.1.

4.2. **Regularity Conditions.** In what follows, we state the assumptions that we maintain to obtain our main results. We then discuss these conditions, and verify them for the examples in Section 3.

C1 (Smoothness of Covariate Distribution)**.** The covariates z_i have a sufficiently smooth distribution, namely for some $0 < m \le 1$, we have that $\mathrm{P}\left(|q'\Sigma z_i/\|z_i\|| < \delta\right)/\delta^m \lesssim 1$ as $\delta \searrow 0$ uniformly in $q \in \mathcal{S}^{d-1}$, with d the dimension of x. The matrix $\Sigma = (\mathrm{E}[x_i z_i])^{-1}$ is finite and invertible.

C2 (Linearization for the Estimator of Bounding Functions)**.** Let $\bar{\theta}$ denote either the unweighted estimator $\widehat{\theta}$ or the weighted estimator $\tilde{\theta}$, and let $v_i = 1$ for the case of the unweighted estimator, and $v_i = e_i$ for the case of the weighted estimator. We assume that for each $\ell = 0, 1$ the estimator $\bar{\theta}_\ell$ admits a linearization of the form:

$$\sqrt{n}\left(\bar{\theta}_\ell(x,\alpha) - \theta_\ell(x,\alpha)\right) = p_k(x)'J_\ell^{-1}(\alpha)\mathbb{G}_n[v_i p_i \varphi_{i\ell}(\alpha)] + \bar{R}_\ell(x,\alpha) \tag{4.6}$$

where $p_i = p_k(x_i)$, $\sup_{\alpha\in\mathcal{A}} \|\bar{R}_\ell(x_i,\alpha)\|_{\mathbb{P}_n,2} \to_P 0$, and $(x_i, z_i, \varphi_{i\ell})$ are i.i.d. random elements.

C3 (Design Conditions)**.** The score function $\varphi_{i\ell}(\alpha)$ is mean zero conditional on x_i, z_i and has uniformly bounded fourth moment conditional on x_i, z_i. The score function is smooth in mean-quartic sense: $\mathrm{E}\left[(\varphi_{i\ell}(\alpha) - \varphi_{i\ell}(\tilde{\alpha}))^4 |x_i, z_i\right]^{1/2} \le C\|\alpha - \tilde{\alpha}\|^{\gamma_\varphi}$ for some constants C and $\gamma_\varphi > 0$. Matrices $J_\ell(\alpha)$ exist and are uniformly Lipschitz over $\alpha \in \mathcal{A}$, a bounded and compact subset of $\mathbb{R}^l$, and $\sup_{\alpha\in\mathcal{A}} \|J^{-1}(\alpha)\|$ as well as the operator norms of matrices $\mathrm{E}[z_i z_i']$, $\mathrm{E}[z_i p_i']$, and $\mathrm{E}[\|p_i p_i'\|^2]$ are uniformly bounded in k. $\mathrm{E}[\|z_i\|^6]$ and $\mathrm{E}[\|x_i\|^6]$ are finite.

$\mathrm{E}\left[\|\theta_\ell(x_i,\alpha)\|^6\right]$ is uniformly bounded in α, and $\mathrm{E}[|\varphi_{i\ell}(\alpha)^4|x_i,z_i]$ is uniformly bounded in α, x, and z. The functions $\theta_\ell(x,\alpha)$ are smooth, namely $|\theta_\ell(x,\alpha)-\theta_\ell(x,\tilde{\alpha})| \le L(x)\,\|\alpha-\tilde{\alpha}\|^{\gamma_\theta}$ for some constant $\gamma_\theta > 0$ and some function $L(x)$ with $\mathrm{E}\left[L(x)^4\right]$ bounded.

C4 (Growth Restrictions)**.** When $k \to \infty$, $\sup_{x\in X}\|p_k(x)\| \le \xi_k$, and the following growth condition holds on the number of series terms:

$$\log^2 n \left(n^{-m/4} + \sqrt{(k/n)\cdot \log n}\cdot \max_i \|z_i\| \wedge \xi_k\right)\left|\sqrt{\max_{i\le n,} F_1^4}\right| \to_{\mathrm{P}} 0, \quad \xi_k^2 \log^2 n/n \to 0,$$

where $\mathcal{F}_1$ is defined in Condition C5 below.

C5 (Complexity of Relevant Function Classes)**.** The function set $\mathcal{F}_1 = \{\varphi_{i\ell}(\alpha), \alpha \in \mathcal{A}, \ell = 0,1\}$ has a square P-integrable envelope F_1 and has a uniform covering L_2 entropy equivalent to that of a VC class. The function class $\mathcal{F}_2 \supseteq \{\theta_{i\ell}(\alpha), \alpha \in \mathcal{A}, \ell = 0,1\}$ has a square P-integrable envelope F_2 for the case of fixed k and bounded envelope F_2 for the case of increasing k, and has a uniform covering L_2 entropy equivalent to that of a VC class.

4.2.1. *Discussion and verification of conditions.* Condition C1 requires that the covariates z_i be continuously distributed, which in turn assures that the support function is everywhere differentiable in $q \in \mathcal{S}^{d-1}$, see Beresteanu and Molinari (2008, Lemma A.8) and Lemma 3 in the Appendix. With discrete covariates, the identified set has exposed faces and therefore its support function is not differentiable in directions q orthogonal to these exposed faces, see e.g., Bontemps, Magnac, and Maurin (2010, Section 3.1). In this case, Condition C1 can be met by adding to the discrete covariates a small amount of smoothly distributed noise. Adding noise gives "curvature" to the exposed faces, thereby guaranteeing that the identified set intersects its supporting hyperplane in a given direction at only one point, and is therefore differentiable, see Schneider (1993, Corollary 1.7.3). Lemma 7 in the Appendix shows that the distance between the true identified set and the set resulting from jittered covariates can be made arbitrarily small. Therefore, in the presence of discrete covariates one can apply our method obtaining arbitrarily slightly conservative inference.

Assumptions C3 and C5 are common regularity conditions and they can be verified using standard arguments. Condition C2 requires the estimates of the bounding functions to be asymptotically linear. In addition, it requires that the number of series terms grows fast enough for the remainder term to disappear. This requirement must be reconciled with Condition C4, which limits the rate at which the number of series terms can increase. We show below how to verify these two conditions in each of the examples of Section 3.

Example (Mean regression, continued)**.** We begin with the simplest case of mean regression with interval valued outcome data. In this case, we have $\widehat{\theta}_\ell(\cdot,\alpha) = p_k(\cdot)'\widehat{\vartheta}_\ell$ with $\widehat{\vartheta}_\ell = (P'P)^{-1}P'y_\ell$ and $P = [p_k(x_1),...,p_k(x_n)]'$. Let ϑ_k be the coefficients of a projection of $\mathrm{E}[y_\ell|x_i]$ on P, or pseudo-true values, so that $\vartheta_k = (P'P)^{-1}P'\mathrm{E}[y_\ell|x_i]$. We have the following linearization for $\widehat{\theta}_\ell(\cdot,\alpha)$

$$\sqrt{n}\left(\widehat{\theta}_\ell(x,\alpha)-\theta_\ell(x,\alpha)\right) = \sqrt{n}p_k(x)(P'P)^{-1}P'(y_\ell - \mathrm{E}[y_\ell|x]) + \sqrt{n}\left(p_k(x)'\vartheta_k - \theta_\ell(x,\alpha)\right).$$

This is in the form of (4.6) with $J_\ell(\alpha) = P'P$, $\varphi_{i\ell}(\alpha) = (y_{i\ell} - \mathrm{E}[y_\ell|x_i])$, and $R_\ell(x,\alpha) = \sqrt{n}\left(p_k(x)'\vartheta_k - \theta_\ell(x,\alpha)\right)$. The remainder term is simply approximation error. Many results on the rate of approximation error are available in the literature. This rate depends on the choice of approximating functions, smoothness of $\theta_\ell(x,\alpha)$, and dimension of x. When using polynomials as approximating function, if $\theta_\ell(x,\alpha) = \mathrm{E}[y_\ell|x_i]$ is s times differentiable with respect to x, and x is d dimensional, then (see e.g. Newey (1997) or Lorentz (1986))

$$\sup_x |p_k(x)'\vartheta_k - \theta_\ell(x,\alpha)| = O(k^{-s/d}).$$

In this case C2 requires that $n^{1/2}k^{-s/d} \to 0$, or that k grows faster than $n^{\frac{d}{2s}}$. Assumption C4 limits the rate at which k can grow. This assumption involves ξ_k and $\sup_{i,\alpha}|\varphi_i(\alpha)|$. The behavior of these terms depends on the choice of approximating functions and some auxiliary assumptions. With polynomials as approximating functions and the support of x compact with density bounded away from zero, $\xi_k = O(k)$. If $y_{i\ell} - \mathrm{E}[y_\ell|x_i]$ has exponential tails, then $\sup_{i,\alpha}|\varphi_i(\alpha)| = O(2(\log n)^{1/2})$. In this case, a sufficient condition to meet C4 is that $k = o(n^{1/3}\log^{-6} n)$. Thus, we can satisfy both C2 and C4 by setting $k \propto n^\gamma$ for any γ in the interval connecting $\frac{d}{2s}$ and $\frac{1}{3}$. Notice that as usual in semiparametric problems, we require undersmoothing compared to the rate that minimizes mean-squared error, which is $\gamma = \frac{d}{d+2s}$. Also, our assumption requires increasing amounts of smoothness as the dimension of x increases.

We now discuss how to satisfy assumptions C2 and C4 more generally. Recall that in our examples, the series estimates of the bounding functions solve

$$\widehat{\theta}_\ell(\cdot,\alpha) = \underset{\theta_\ell \in \mathcal{L}^2(X,\mathrm{P})}{\arg\min}\ \mathbb{E}_n\left[m(y_{i\ell}, \theta_\ell(x_i,\alpha),\alpha)\right]$$

or $\widehat{\theta}_\ell(\cdot,\alpha) = p_k(\cdot)'\widehat{\vartheta}_\ell$ with $\widehat{\vartheta}_\ell = \arg\min_\vartheta \mathbb{E}_n\left[m(y_{i\ell}, p_k(x_i)'\vartheta,\alpha)\right]$. As above, let ϑ_k be the solution to $\vartheta_k = \arg\min_\vartheta \mathrm{E}\left[m(y_{i\ell}, p_k(x_i)'\vartheta,\alpha)\right]$. We show that the linearization in C2 holds by writing

$$\sqrt{n}\left(\widehat{\theta}(x,\alpha) - \theta_\ell(x,\alpha)\right) = \sqrt{n}p_k(x)\left(\widehat{\vartheta} - \vartheta_k\right) + \sqrt{n}\left(p_k(x)'\vartheta_k - \theta_\ell(x,\alpha)\right). \tag{4.7}$$

The first term in (4.7) is estimation error. We can use the results of He and Shao (2000) to show that

$$\left(\widehat{\vartheta} - \vartheta_k\right) = \mathbb{E}_n\left[J_\ell^{-1}p_i\psi_i\right] + o_p(n^{-1/2}),$$

where ψ denotes the derivative of $m(y_{i\ell}, p_k(x_i)'\vartheta,\alpha)$ with respect to ϑ.

The second term in (4.7) is approximation error. Standard results from approximation theory as stated in e.g. Chen (2007) or Newey (1997) give the rate at which the error from the best L_2-approximation to θ_ℓ disappears. When m is a least squares objective function, these results can be applied directly. In other cases, such as quantile or distribution regression, further work must be done.

Example (Quantile regression with interval valued data, continued)**.** The results of Belloni, Chernozhukov, and Fernandez-Val (2011) can be used to verify our conditions for quantile regression. In particular, Lemma 1 from Appendix B of Belloni, Chernozhukov, and Fernandez-Val (2011) gives the rate at which the approximation error vanishes, and Theorem 2 from Belloni, Chernozhukov, and Fernandez-Val (2011) shows that the linearization condition (C2) holds. The conditions required for these results are as follows.

(Q.1) The data $\{(y_{i0}, y_{i1}, x_i), 1 \leq i \leq n\}$ are an i.i.d. sequence of real $(2+d)$-vectors.
(Q.2) For $\ell = \{0,1\}$, the conditional density of y_ℓ given x is bounded above by $\overline{f}$, below by $\underline{f}$, and its derivative is bounded above by $\overline{f'}$ uniformly in y_ℓ and x. $f_{y_\ell|X}\left(Q_{y_\ell|x}(\alpha|x)|x\right)$ is bounded away from zero uniformly in $\alpha \in \mathcal{A}$ and $x \in \mathcal{X}$.
(Q.3) For all k, the eigenvalues of $\mathrm{E}[p_i p_i']$ are uniformly bounded above and away from zero.
(Q.4) $\xi_k = O(k^a)$. $Q_{y_\ell|x}$ is s times continuously differentiable with $s > (a+1)d$. The series functions are such that

$$\inf_\vartheta \mathrm{E}\left[\|p_k\vartheta - \theta_\ell(x,\alpha)\|_2\right] = O(k^{-s/d})$$

and

$$\inf_\vartheta \|p_k\vartheta - \theta_\ell(x,\alpha)\|_\infty = O(k^{-s/d}).$$

(Q.5) k is chosen such that $k^{3+6a}(\log n)^7 = o(n)$.

Condition Q.4 is satisfied by polynomials with $a = 1$ and by splines or trigonometric series with $a = 1/2$. Under these assumptions, Lemma 1 of appendix B from Belloni, Chernozhukov, and Fernandez-Val (2011) shows that the approximation error satisfies

$$\sup_{x\in\mathcal{X},\alpha\in\mathcal{A}} \left|p_k(x)'\vartheta_k(\alpha) - \theta_\ell(x,\alpha)\right| \lesssim k^{\frac{ad-s}{d}}.$$

Theorem 2 of Belloni, Chernozhukov, and Fernandez-Val (2011) then shows that C2 holds. Condition C4 also holds because for quantile regression ψ_i is bounded, so C4 only requires $k^{1+2a}(\log n)^2 = o(n)$, which is implied by Q.5.

Example (Distribution regression, continued)**.** As described above, the estimator solves $\widehat{\vartheta} = \arg\min_\vartheta \mathbb{E}_n\left[m(y_i, p_k(x_i)'\vartheta, \alpha)\right]$ with

$$m(y_i, p_k(x_i)'\vartheta, \alpha) = -1\{y < \alpha\}\log\Phi\left(p_k(x_i)'\vartheta\right) - 1\{y \geq \alpha\}\log\left(1 - \Phi\left(p_k(x_i)'\vartheta\right)\right)$$

for some known distribution function Φ. We must show that estimation error, $\widehat{\vartheta} - \vartheta_k$, can be linearized and that the bias, $p_k(x)\vartheta_k - \theta_\ell(x,\alpha)$, is $o(n^{-1/2})$. We first verify the conditions of He and Shao (2000) to show that $(\widehat{\vartheta} - \vartheta_k)$ can be linearized. Adopting their notation, in this example we have that the derivative of $m(y_i, p_k(x_i)'\vartheta, \alpha)$ with respect to ϑ is

$$\psi(y_i, x_i, \vartheta) = -\left(\frac{1\{y < \alpha\}}{\Phi\left(p_k(x_i)'\vartheta\right)} - \frac{1\{y \geq \alpha\}}{1 - \Phi\left(p_k(x_i)'\vartheta\right)}\right)\phi\left(p_k(x_i)'\vartheta\right)p_k(x_i),$$

where ϕ is the pdf associated with Φ. Because $m(y_i, p_k(x_i)'\vartheta, \alpha)$ is a smooth function of ϑ, $\mathbb{E}_n\psi(y_i, x_i, \widehat{\vartheta}) = 0$, and conditions C.0 and C.2 in He and Shao (2000) hold. If ϕ is differentiable with a bounded derivative, then ψ is Lipschitz in ϑ, and we have the bound

$$\|\eta_i(\vartheta,\tau)\|^2 \lesssim \|p_k(x_i)\|^2\|\tau - \vartheta\|^2,$$

where $\eta_i(\vartheta,\tau) = \psi(y_i,x_i,\vartheta) - \psi(y_i,x_i,\tau) - E\psi(y_i,x_i,\vartheta) + E\psi(y_i,x_i,\tau)$. If we assume

$$\max_{i\leq n}\|p_k(x_i)\| = O(k^a),$$

as would be true for polynomials with $a = 1$ or splines with $a = 1/2$, and k is of order less than or equal to $n^{1/a}$ then condition C.1 in He and Shao (2000) holds. Differentiability of

ϕ and C3 are sufficient for C.3 in He and Shao (2000). Finally, conditions C.4 and C.5 hold with $A(n,k)=k$ because

$$\left|s'\eta_i(\vartheta,\tau)\right|^2 \lesssim \left|s'p_k(x_i)\right|^2 \|\tau-\vartheta\|^2,$$

and $\mathrm{E}\left[|s'p_k(x_i)|^2\right]$ is uniformly bounded for $s\in S^k$ for all k when the series functions are orthonormal. Applying Theorem 2.2 of He and Shao (2000), we obtain the desired linearization if $k=o\left((n/\log n)^{1/2}\right)$.

The results of Hirano, Imbens, and Ridder (2003) can be used to show that the approximation bias is sufficiently small. Lemma 1 from Hirano, Imbens, and Ridder (2003) shows that for the logistic distribution regression,

$$\begin{aligned}\|\vartheta_k-\vartheta_k^*\| =& O(k^{-s/(2d)})\\ \sup_x |\Phi\left(p_k(x)\vartheta_k\right)-\Phi\left(\theta_\ell(x,\alpha)\right)| =& O(k^{-s/(2d)}\xi_k),\end{aligned}$$

which implies that

$$\sup_{x\in\mathcal{X},\alpha\in\mathcal{A}} |p_k(x)\vartheta_k-\theta_\ell(x,\alpha)| = O(k^{-s/(2d)}\xi_k).$$

This result is only for the logistic link function, but it can easily be adapted for any link function with first derivative bounded from above and second derivative bounded away from zero. We need the approximation error to be $o(n^{-1/2})$. For this, it suffices to have

$$k^{-s/(2d)}\xi_k n^{1/2}=o(1).$$

Letting $\xi_k=O(k^a)$ as above, it suffices to have $k\propto n^\gamma$ for $\gamma>\frac{d}{s-2ad}$.

To summarize, condition C2 can be met by having $k\propto n^\gamma$ for any $\gamma\in\left(\frac{d}{s-2ad},\frac{1}{2}\right)$. Finally, as in the mean and quantile regression examples above, condition C4 will be met if $\gamma<\frac{1}{1+2a}$.

4.3. **Theoretical Results.** In order to state the result we define

$$\begin{aligned}h_k(t) := & \quad q'\Sigma\mathrm{E}[z_ip_i'1\{q'\Sigma z_i>0\}]J_1^{-1}(\alpha)p_i\varphi_{i1}(\alpha)\\ & + \; q'\Sigma\mathrm{E}[z_ip_i'1\{q'\Sigma z_i<0\}]J_0^{-1}(\alpha)p_i\varphi_{i0}(\alpha)\\ & - \; q'\Sigma x_iz_i'\Sigma\mathrm{E}\left[z_iw_{i,q'\Sigma}(\alpha)\right]\\ & + \; q'\Sigma z_iw_{i,q'\Sigma}(\alpha).\end{aligned}$$

Theorem 1 (Limit Theory for Support Function Process)**.** *The support function process* $S_n(t)=\sqrt{n}\left(\widehat{\sigma}_{\widehat{\theta},\widehat{\Sigma}}-\sigma_{\theta,\Sigma}\right)(t)$, *where* $t=(q,\alpha)\in\mathcal{S}^{d-1}\times\mathcal{A}$, *admits the approximation* $S_n(t)=\mathbb{G}_n[h_k(t)]+o_{\mathrm{P}}(1)$ *in* $\ell^\infty(T)$. *Moreover, the support function process admits an approximation*

$$S_n(t)=\mathbb{G}[h_k(t)]+o_{\mathrm{P}}(1) \text{ in } \ell^\infty(T),$$

where the process $\mathbb{G}[h_k(t)]$ *is a tight P-Brownian bridge in* $\ell^\infty(T)$ *with covariance function* $\Omega_k(t,t')=\mathrm{E}[h_k(t)h_k(t')]-\mathrm{E}[h_k(t)]\mathrm{E}[h_k(t')]$ *that is uniformly Holder on* $T\times T$ *uniformly in* k, *and is uniformly non-degenerate in* k. *These bridges are stochastically equicontinuous with respect to the* L_2 *pseudo-metric* $\rho_2(t,t')=[\mathrm{E}[h(t)-h(t')]^2]^{1/2}\lesssim\|t-t'\|^c$ *for some* $c>0$ *uniformly in* k. *The sequence* $\mathbb{G}[h_k(t)]$ *does not necessarily converge weakly under*

$k \to \infty$; however, each subsequence has a further convergent subsequence converging to a tight Gaussian process in $\ell^\infty(T)$ with a non-degenerate covariance function. Furthermore, the canonical distance between the law of the support function process $S_n(t)$ and the law of $\mathbb{G}[h_k(t)]$ in $\ell^\infty(T)$ approaches zero, namely $\sup_{g \in BL_1(\ell^\infty(T),[0,1])} |\mathrm{E}[g(S_n)] - \mathrm{E}[g(\mathbb{G}[h_k])]| \to 0$.

Next we consider the behavior of various statistics based on the support function process. Formally, we consider these statistics as mappings $f : \ell^\infty(T) \to \mathbb{R}$ from the possible values s of the support function process S_n to the real line. Examples include:

- a support function evaluated at $t \in T$, $f(s) = s(t)$,
- a Kolmogorov statistic, $f(s) = \sup_{t \in T_0} |s(t)|/\varpi(t)$,
- a directed Kolmogorov statistic, $f(s) = \sup_{t \in T_0} \{-s(t)\}_+ /\varpi(t)$,
- a Cramer-Von-Mises statistic, $f(s) = \int_T s^2(t)/\varpi(t) d\nu(t)$,

where T_0 is a subset of T, ϖ is a continuous and uniformly positive weighting function, and ν is a probability measure over T whose support is T.[10] More generally we can consider any continuous function f such that $f(Z)$ (a) has a continuous distribution function when Z is a tight Gaussian process with non-degenerate covariance function and (b) $f(\xi_n + c) - f(\xi_n) = o(1)$ for any $c = o(1)$ and any $\|\xi_n\| = O(1)$. Denote the class of such functions $\mathcal{F}_c$ and note that the examples mentioned above belong to this class by the results of Davydov, Lifshits, and Smorodina (1998).

Theorem 2 (Limit Inference on Support Function Process)**.** *Furthermore, the canonical distance between the law of the support function process $S_n(t)$ and the law of $\mathbb{G}[h_k(t)]$ in $\ell^\infty(T)$ approaches zero, namely* $\sup_{g \in BL_1(\ell^\infty(T),[0,1])} |\mathrm{E}[g(S_n)] - \mathrm{E}[g(\mathbb{G}[h_k])]| \to 0$. *For any $\widehat{c}_n = c_n + o_{\mathrm{P}}(1)$ and $c_n = O_{\mathrm{P}}(1)$ and $f \in \mathcal{F}_c$ we have*

$$\mathrm{P}\{f(S_n) \le \widehat{c}_n\} - \mathrm{P}\{f(\mathbb{G}[h_k]) \le c_n\} \to 0.$$

If $c_n(1-\tau)$ is the $(1-\tau)$-quantile of $f(\mathbb{G}[h_k])$ and $\widehat{c}_n(1-\tau) = c_n(1-\tau) + o_{\mathrm{P}}(1)$ is any consistent estimate of this quantile, then

$$\mathrm{P}\{f(S_n) \le \widehat{c}_n(1-\tau)\} \to 1 - \tau.$$

Let $e_i^o = e_i - 1$, $h_k^o = h_k - \mathrm{E}[h_k]$, and let P^e denote the probability measure conditional on the data.

Theorem 3 (Limit Theory for the Bootstrap Support Function Process)**.** *The bootstrap support function process $\widetilde{S}_n(t) = \sqrt{n}(\widetilde{\sigma}_{\widetilde{\theta},\widetilde{\Sigma}} - \widehat{\sigma}_{\widehat{\theta},\widehat{\Sigma}})(t)$, where $t = (q, \alpha) \in \mathcal{S}^{d-1} \times \mathcal{A}$, admits the following approximation conditional on the data: $\widetilde{S}_n(t) = \mathbb{G}_n[e_i^o h_k^o(t)] + o_{\mathrm{P}^e}(1)$ in $\ell^\infty(T)$ in probability P. Moreover, the bootstrap support function process admits an approximation conditional on the data:*

$$\widetilde{S}_n(t) = \widetilde{\mathbb{G}[h_k(t)]} + o_{\mathrm{P}^e}(1) \text{ in } \ell^\infty(T), \text{ in probability } \mathrm{P},$$

where $\widetilde{\mathbb{G}[h_k]}$ is a sequence of tight P-Brownian bridges in $\ell^\infty(T)$ with the same distributions as the processes $\mathbb{G}[h_k]$ defined in Theorem 1, and independent of $\mathbb{G}[h_k]$. Furthermore, the canonical distance between the law of the bootstrap support function process

[10] Observe that test statistics based on the (directed) Hausdorff distance (see, e.g., Beresteanu and Molinari (2008)) are special cases of the (directed) Kolmogorov statistics above.

$\widetilde{S}_n(t)$ *conditional on the data and the law of* $\mathbb{G}[h_k(t)]$ *in* $\ell^\infty(T)$ *approaches zero, namely* $\sup_{g \in BL_1(\ell^\infty(T),[0,1])} |\mathrm{E}_{\mathrm{P}^e}[g(\widetilde{S}_n)] - \mathrm{E}[g(\mathbb{G}[h_k])]| \to_P 0$.

Theorem 4 (Bootstrap Inference on the Support Function Process)**.** *For any* $c_n = O_{\mathrm{P}}(1)$ *and* $f \in \mathcal{F}_c$ *we have*

$$\mathrm{P}\{f(S_n) \leq c_n\} - \mathrm{P}^e\{f(\widetilde{S}_n) \leq c_n\} \to_P 0.$$

In particular, if $\widetilde{c}_n(1-\tau)$ *is the* $(1-\tau)$*-quantile of* $f(\widetilde{S}_n)$ *under* P^e*, then*

$$\mathrm{P}\{f(S_n) \leq \widetilde{c}_n(1-\tau)\} \to_P 1-\tau.$$

5. Application: the gender wage gap and selection

An important question in labor economics is whether the gender wage gap is shrinking over time. Blau and Kahn (1997) and Card and DiNardo (2002), among others, have noted the coincidence between a rise in within-gender inequality and a fall in the gender wage gap over the last 40 years. Mulligan and Rubinstein (2008) observe that the growing wage inequality within gender should induce females to invest more in productivity. In turn, able females should differentially be pulled into the workforce. Motivated by this observation, they use Heckman's two-step estimator on repeated Current Population Survey cross-sections in order to compute relative wages for women since 1970, holding skill composition constant. They find that in the 1970s selection into the female workforce was negative, while in the 1990s it was positive. Moreover, they argue that the majority of the reduction in the gender gap can be attributed to the changes in the female workforce composition. In particular, the OLS estimates of the log-wage gap has fallen from -0.419 in the 1970s to -0.256 in the 1990s, though the Heckman two step estimates suggest that once one controls for skill composition, the wage gap is -0.379 in the 1970s and -0.358 in the 1990s. Based on these results, Mulligan and Rubinstein (2008) conclude that the wage gap has not shrunk over the last 40 years. Rather, the behavior of the OLS estimates can be explained by a switch from negative to positive selection into female labor force participation.

In what follows, we address the same question as Mulligan and Rubinstein (2008), but use our method to estimate bounds on the quantile gender wage gap without assuming a parametric form of selection or a strong exclusion restriction.[11] We follow their approach of comparing conditional quantiles that ignore the selection effect, with the bounds on these quantiles that one obtains when taking selection into account.

Our results show that we are unable to reject that the gender wage gap declined over the period in question. This suggests that the instruments may not be sufficiently strong to yield tight bounds and that there may not be enough information in the data to conclude that the gender gap has or has not declined from 1975 to 1999 without strong functional form assumptions.

5.1. **Setup.**

The Mulligan and Rubinstein (2008) setup relates log-wage to covariates in a linear model as follows:

$$\log w = x'\beta + \varepsilon,$$

[11] We use the same data, the same variables and the same instruments as in their paper.

Table 1. Gender wage gap estimates

	OLS	2-step	QR(0.5)	Low	High
1975-1979	-0.408	-0.360	-0.522	-1.242	0.588
	(0.003)	(0.013)	(0.003)	(0.016)	(0.061)
1995-1999	-0.268	-0.379	-0.355	-0.623	0.014
	(0.003)	(0.013)	(0.003)	(0.012)	(0.010)

This table shows estimates of the gender wage gap (female − male) conditional on having average characteristics. The first column shows OLS estimates of the average gender gap. The second column shows Heckman two-step estimate. The third column shows quantile regression estimates of the median gender wage gap. The fourth and fifth columns show estimates of bounds on the median wage gap that account for selection. Standard errors are shown in parentheses. The standard errors were calculated using the reweighting bootstrap described above.

wherein x includes marital status, years of education, potential experience, potential experience squared, and region dummies, as well as their interactions with an indicator for gender which takes the value 1 if the individual is female, and zero otherwise. They model selection as in the following equation:

$$u = 1\left\{z'\gamma > 0\right\},$$

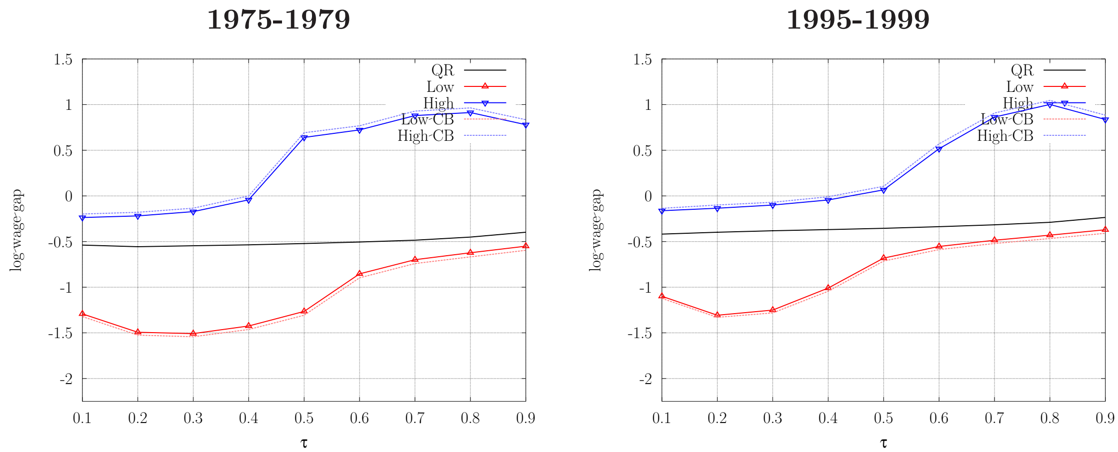
where $z = [x \; \tilde{z}]$ and $\tilde{z}$ is marital status interacted with indicators for having zero, one, two, or more than two children.

For each quantile, we estimate bounds for the gender wage gap utilizing our method. The bound equations we use are given by $\theta_\ell(x, v, \alpha) = p_k(x, v)'\vartheta_\ell^{xv}(\alpha)$, where $p_k(x, v) = \begin{bmatrix} x & v & w \end{bmatrix}$, v are indicators for the number of children, and w consists of years of education squared, potential experience cubed, and education × potential experience, and v interacted with marital status. After taking the intersection of the bounds over the excluded variables v, our estimated bounding functions are simply the minimum or maximum over v of $p_k(x, v)'\vartheta_\ell^{xv}(\alpha)$.

5.2. **Results.** Let $\bar{x}_f$ be a female with average (unconditional on gender) characteristics and $\bar{x}_m$ be a male with average (unconditional on gender or year) characteristics. In what follows, we report the predicted gender wage gap for someone with average characteristics, $(\bar{x}_f - \bar{x}_m)\,\beta(\alpha)$. The first two columns of table 1 reproduce the results of Mulligan and Rubinstein (2008). The first column shows the gender wage gap estimated by ordinary least squares. The second column shows estimates from Heckman's two-step selection correction. The OLS estimates show a decrease in the wage gap, while the Heckman selection estimates show no change. The third column shows estimates of the median gender wage gap from quantile regression. Like OLS, quantile regression shows a decrease in the gender wage gap. The final two columns show bounds on the median gender wage gap that account for selection. The bounds are wide, especially in the 1970s. In both periods, the bounds do not preclude a negative nor a positive gender wage gap. The bounds let us say very little about the change in the gender wage gap.

Figure 2 shows the estimated quantile gender wage gaps in the 1970s and 1990s. The solid black line shows the quantile gender wage gap when selection is ignored. In both the

FIGURE 2. Bounds at Quantiles for full sample



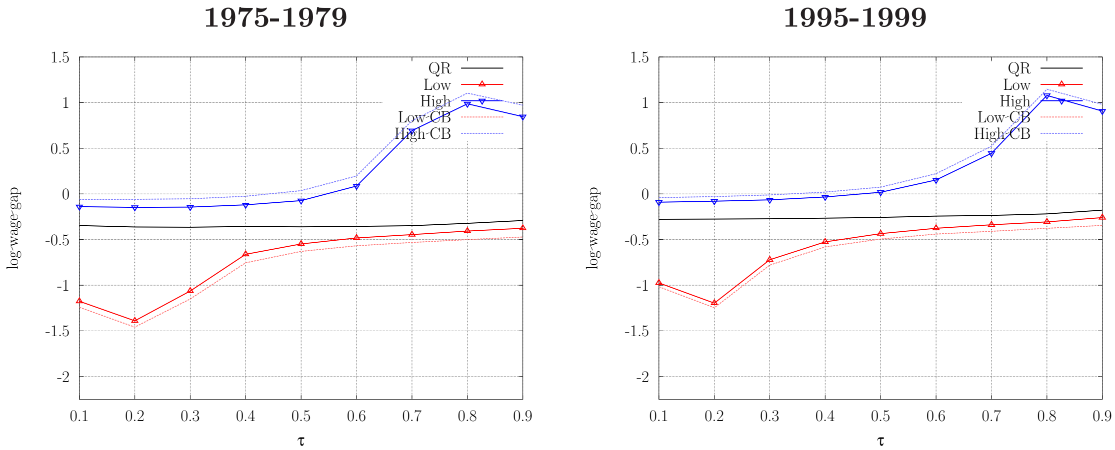
This figure shows the estimated quantile gender wage gap (female − male) conditional on having average characteristics. The solid black line shows the quantile gender wage gap when selection is ignored. The blue and red lines with upward and downward pointing triangles show upper and lower bounds that account for employment selection for females. The dashed lines represent a uniform 90% confidence region for the bounds.

1970s and 1990s, the gender wage gap is larger for lower quantiles. At all quantiles the gap in the 1990s is about smaller 40% smaller than in the 1970s. However, this result should be interpreted with caution because it ignores selection into the labor force.

The blue line with downward pointing triangles and the red line with upward pointing triangles show our estimated bounds on the gender wage gap after accounting for selection. The dashed lines represent a uniform 90% confidence region. In both the 1970s and 1990s, the upper bound lies below zero for low quantiles. This means that the low quantiles of the distribution of wages conditional on having average characteristics are lower for a woman than for a man. This difference exists even if we allow for the most extreme form of selection (subject to our exclusion restriction) into the labor force for women. For quantiles at or above the median, our estimated upper bound lies above zero and our lower bound lies below zero. Thus, high quantiles of the distribution of wages conditional on average characteristics could be either higher or lower for women than for men, depending on the true pattern of selection. For all quantiles, there is considerable overlap between the bounded region in the 1970s and in the 1990s. Therefore, we can essentially say nothing about the change in the gender wage gap. It may have decreases, as suggested by least squares or quantile regression estimates that ignore selection. The gap may also have stayed the same, as suggested by Heckman selection estimates. In fact, we cannot even rule out the possibility that the gap increased.

The bounds in figure 2 are tighter in the 1990s than in the 1970s. This reflects higher female labor force participation in the 1990s. To find even tighter bounds, we can repeat the estimation focusing only on subgroups with higher labor force attachment. Figures 3-6 show the estimated quantile gender wage gap conditional on being in certain subgroups. That is,

FIGURE 3. Quantile bounds for singles

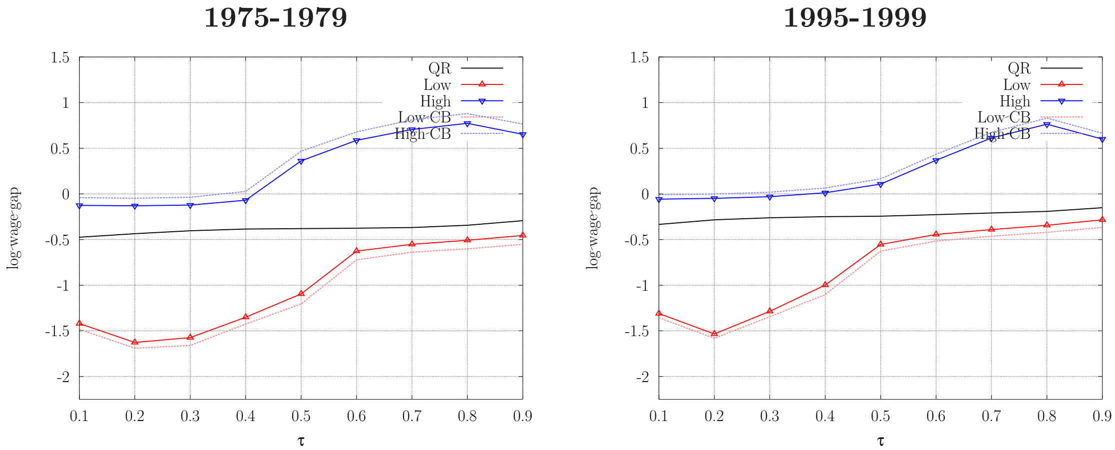


This figure shows the estimated quantile gender wage gap (female − male) conditional on being single with average characteristics. The solid black line shows the quantile gender wage gap when selection is ignored. The blue and red lines with upward and downward pointing triangles show upper and lower bounds that account for employment selection for females. The dashed lines represent a uniform 90% confidence region for the bounds.

rather than reporting the gender wage gap for someone with average characteristics, these figures show the gender wage gap for someone with average subgroup characteristics (e.g., unconditional on gender or year, conditional on subgroup: marital status and educaiton level). To generate these figures, the entire model was re-estimated using only observations within each subgroup.

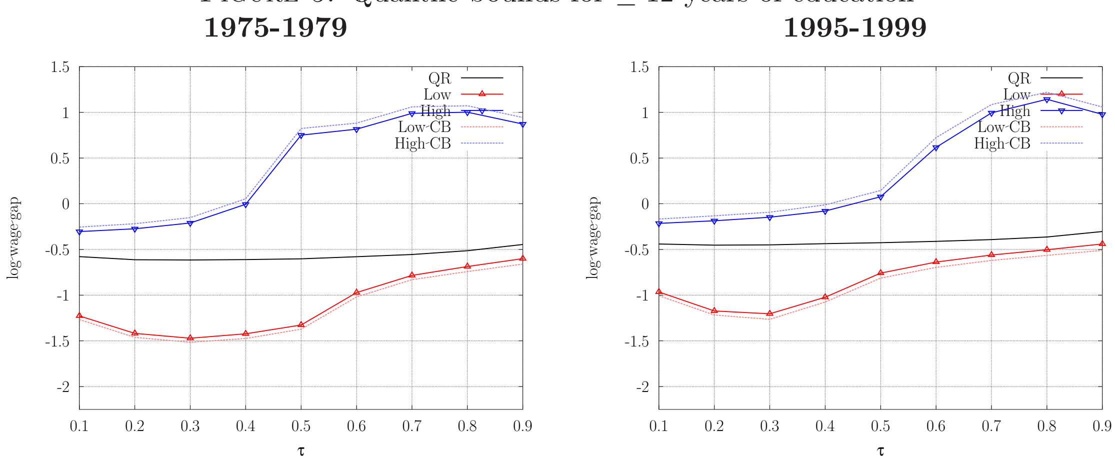
Figures 3 and 4 show the results for singles and people with at least 16 years of education. The results are broadly similar to the results for the full sample. There is robust evidence of a gap at low quantiles, although it is only marginally significant for the highly educated in the 1990s. As expected, the bounds are tighter than the full sample bounds. Nonetheless, little can be said about the gap at higher quantiles or the change in the gap. For comparison, figure 5 shows the results for people with no more than a high school education. These bounds are slightly wider than the full sample bounds, but otherwise very similar. Figure 6 shows results for singles with at least a college degree. These bounds are the tighter than all others, but still do not allow us to say anything about the change in the gender wage gap. Also, there is no longer robust evidence of a gap at low quantiles. A gap is possible, but we cannot reject the null hypothesis of zero gap at all quantiles at the 10% level.

FIGURE 4. Quantile bounds for ≥ 16 years of education



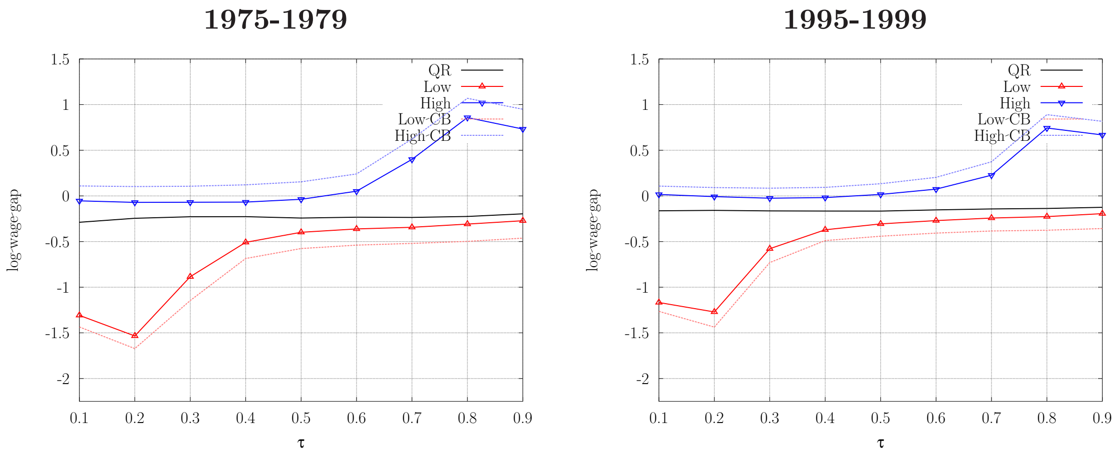
This figure shows the estimated quantile gender wage gap (female − male) conditional on having at least 16 years of education with average characteristics. The solid black line shows the quantile gender wage gap when selection is ignored. The blue and red lines with upward and downward pointing triangles show upper and lower bounds that account for employment selection for females. The dashed lines represent a uniform 90% confidence region for the bounds.

FIGURE 5. Quantile bounds for ≤ 12 years of education



This figure shows the estimated quantile gender wage gap (female − male) conditional on having 12 or fewer years of education with average characteristics. The solid black line shows the quantile gender wage gap when selection is ignored. The blue and red lines with upward and downward pointing triangles show upper and lower bounds that account for employment selection for females. The dashed lines represent a uniform 90% confidence region for the bounds.

Figure 6. Quantile bounds for ≥ 16 years of education and single



This figure shows the estimated quantile gender wage gap (female − male) conditional on being single with at least 16 years of education and average characteristics. The solid black line shows the quantile gender wage gap when selection is ignored. The blue and red lines with upward and downward pointing triangles show upper and lower bounds that account for employment selection for females. The dashed lines represent a uniform 90% confidence region for the bounds.

5.2.1. *With restrictions on selection.* Blundell, Gosling, Ichimura, and Meghir (2007) study changes in the distribution of wages in the UK. Like us, they allow for selection by estimating quantile bounds. Also, like us, Blundell, Gosling, Ichimura, and Meghir (2007) find that the estimated bounds are quite wide. As a result, they explore various restrictions to tighten the bound. One restriction is to assume that the wages of the employed stochastically dominates the distribution of wages for those not working. This implies that the observed quantiles of wages conditional on employment are an upper bound the quantiles of wages not conditional on employment.

Figure 7 shows results imposing stochastic dominance for the full sample and for highly educated singles. Stochastic dominance implies that the upper bound coincides with the quantile regression estimate. With stochastic dominance there is robust evidence of a gender wage gap at all quantiles in both the 1970s and 1990s, for both the full sample and the single highly educated subsample. The bounds with stochastic dominance are much tighter than without. In fact, it appears that they may be tight enough to say something about the change in the gender wage gap. Accordingly, figure 8 shows the estimated bounds for the change in the gender wage gap. It shows results for both the full sample and the single high education subsample. For the full sample, the estimated bounds include zero at low and moderate quantiles. At the 0.6 and higher quantiles, there is significant evidence that the gender wage gap decreased by approximately 0.15 log dollars. For highly educated singles, the change in the gender wage gap is not significantly different from zero for any quantiles.

The assumption of positive selection into employment is not innocuous. It may be violated if there is a strong positive correlation between potential wages and reservation wages. This may be the case if there is positive assortative matching in the marriage market. Women with high potential wages could marry men with high wages, making these high potential wage women less likely to work. Also, the conclusion of Mulligan and Rubinstein (2008) that there was a switch from adverse selection into the labor market in the 1970s to advantageous selection in the 1990s implies that stochastic dominance did not hold in the 1970s. Accordingly, we also explore some weaker restrictions. Blundell, Gosling, Ichimura, and Meghir (2007) propose a median restriction — that the median wage offer for those not working is less than or equal to the median observed wage. This restriction implies the following bounds on the distribution of wages

$$\begin{aligned} F(y|x,u=1)\mathrm{P}(u=1|x) + 1\{y \geq Q_y(0.5|x,u=1)\}0.5\mathrm{P}(u=0|x) \leq \\ \leq F(y|x) \leq F(y|x,u=1)\mathrm{P}(u=1|x) + \mathrm{P}(u=0|x), \end{aligned}$$

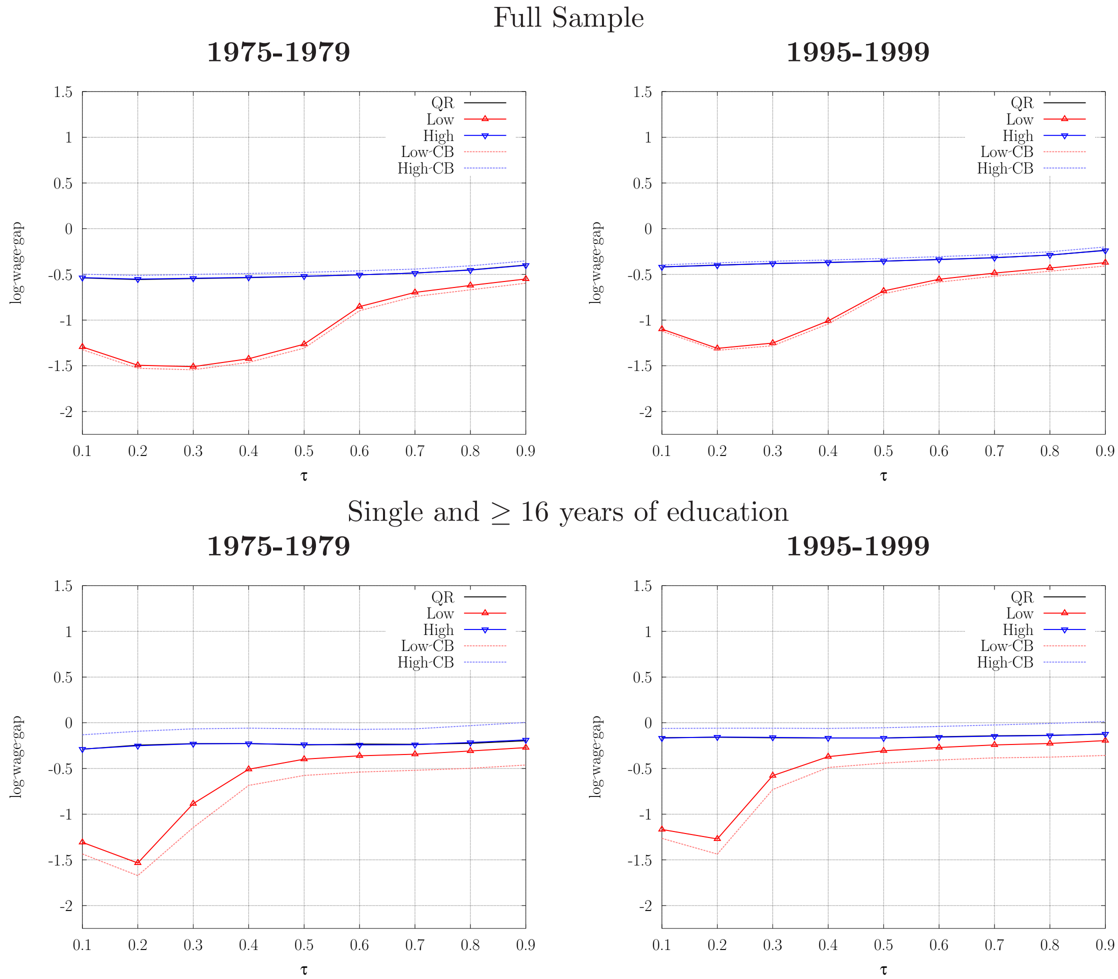
where y is wage and $u=1$ indicates employment. Transforming these into bounds on the conditional quantiles yields

$$Q_0\left(\alpha|x\right) \leq Q_y\left(\alpha|x\right) \leq Q_1\left(\alpha|x\right),$$

where

$$Q_0(\alpha|x) = \begin{cases} Q_y\left(\frac{\alpha-\mathrm{P}(u=0|x)}{\mathrm{P}(u=1|x)}\middle| x, u=1\right) & \text{if } \alpha \geq \mathrm{P}(u=0|x) \\ y_0 & \text{otherwise} \end{cases},$$

Figure 7. Quantile bounds for full sample imposing stochastic dominance



This figure shows the estimated quantile gender wage (female − male) conditional on average characteristics. The solid black line shows the quantile gender wage when selection is ignored. The blue and red lines with upward and downward pointing triangles show upper and lower bounds that account for employment selection for females. The dashed lines represent a uniform 90% confidence region for the bounds.

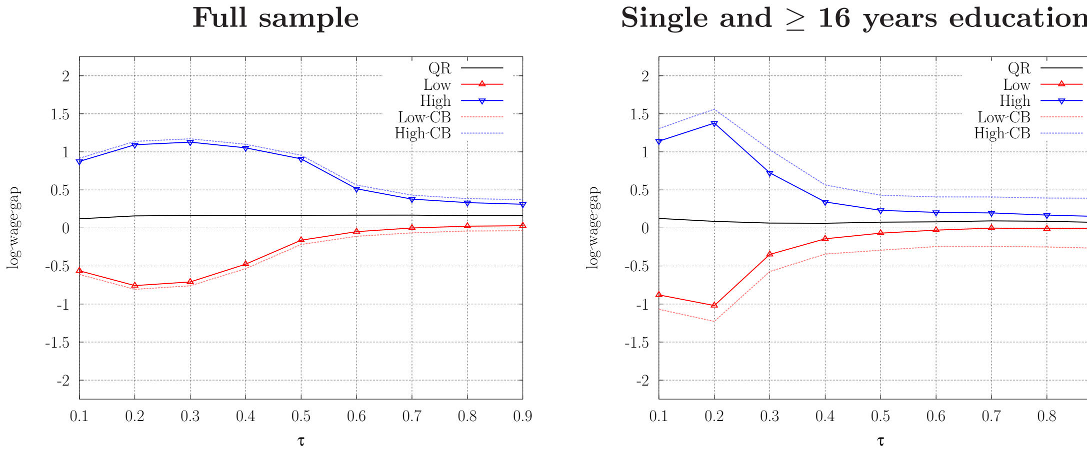
and

$$Q_1(\alpha|x) = \begin{cases} Q_y\left(\frac{\alpha}{\mathrm{P}(u=1|x)}\middle| x, u=1\right) & \text{if } \alpha < 0.5 \ \&\ \alpha \leq \mathrm{P}(u=1|x) \\ Q_y\left(\frac{\alpha - 0.5\mathrm{P}(u=0|x)}{\mathrm{P}(u=1|x)}\middle| x, u=1\right) & \text{if } \alpha \geq 0.5 \ \&\ \alpha \leq \frac{1+\mathrm{P}(u=1|x)}{2} \\ y_1 & \text{otherwise} \end{cases}.$$

As above, we can also express $Q_0(\alpha|x)$ and $Q_1(\alpha|x)$ as the α conditional quantiles of $\tilde{y}_0$ and $\tilde{y}_1$ where

$$\tilde{y}_0 = y1\{u=1\} + y_0 1\{u=0\}$$

FIGURE 8. Quantile bounds for the change in the gender wage gap imposing stochastic dominance



This figure shows the estimated change (1990s − 1970s) in the quantile gender wage gap (female − male) conditional on having average characteristics. The solid black line shows the quantile gender wage gap when selection is ignored. The blue and red lines with upward and downward pointing triangles show upper and lower bounds that account for employment selection for females. The dashed lines represent a uniform 90% confidence region for the bounds.

and

$$\tilde{y}_1 = \begin{cases} y1\{u=1\} + y_1 1\{u=0\} & \text{with probability } 0.5 \\ y1\{u=1\} + Q_y(0.5|x, u=1)1\{u=0\} & \text{with probability } 0.5 \end{cases}.$$

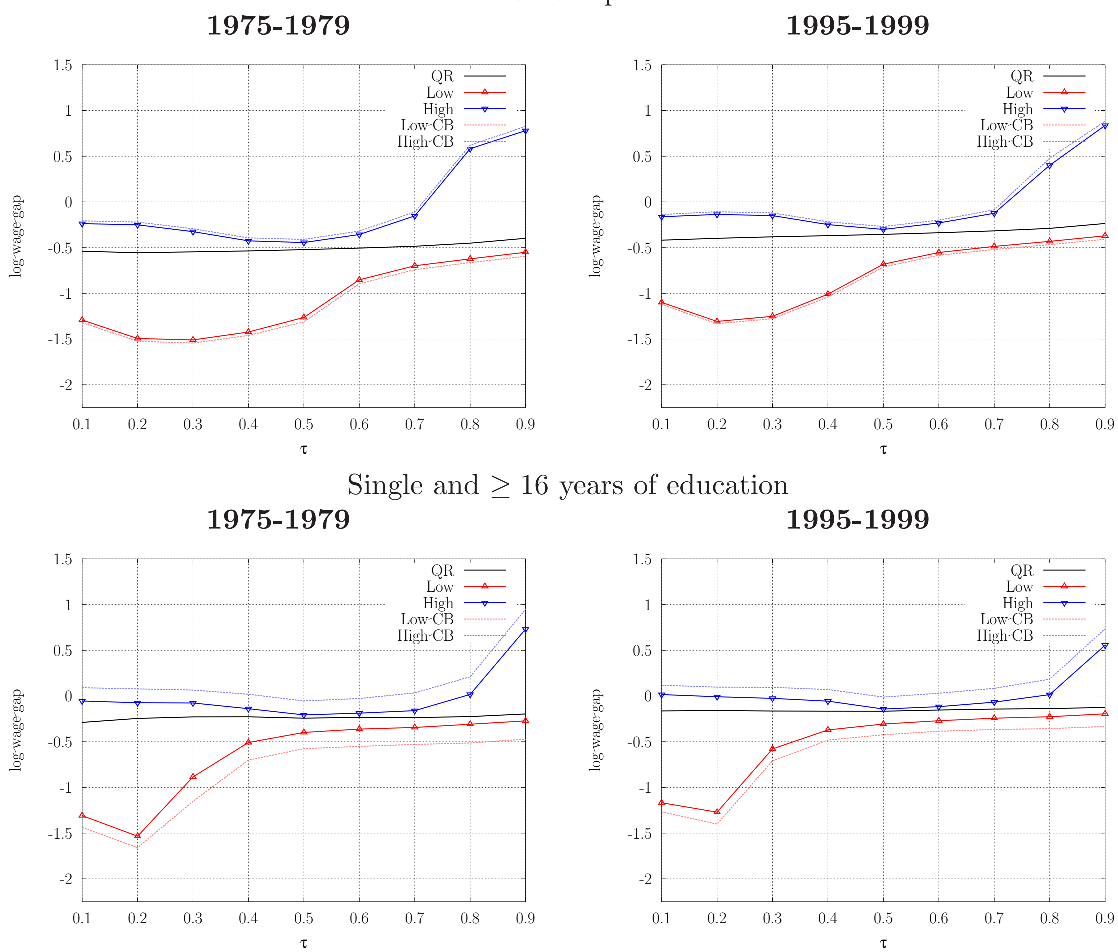
We can easily generalize this median restriction by assuming the α_1 quantile of wages conditional on working is greater than or equal to the α_0 quantile of wages conditional on not working. In that case, the bounds can still be expressed as α conditional quantiles of $\tilde{y}_0$ and $\tilde{y}_1$ with $\tilde{y}_0$ as defined above and

$$\tilde{y}_1 = \begin{cases} y1\{u=1\} + y_1 1\{u=0\} & \text{with probability } (1-\alpha_0) \\ y1\{u=1\} + Q_y(\alpha_1|x, u=1)1\{u=0\} & \text{with probability } \alpha_0 \end{cases}.$$

We can even impose a set of these restrictions for $(\alpha_1, \alpha_0) \in \mathcal{R} \subseteq \mathcal{A} \times \mathcal{A}$. Stochastic dominance is equivalent to imposing this restriction for $\alpha_1 = \alpha_0$ for all $\alpha_1 \in [0, 1]$.

Figure 9 show estimates of the gender wage gap with the median restriction. The results are qualitatively similar to the results without the restriction. As without the restriction, we obtain robust evidence of a gender wage gap at low quantiles in both the 1970s and 1990s, and there is substantial overlap in the bounds between the two periods, so we cannot say much about the change in the gender wage gap. The main difference with the median restriction is that there is also robust evidence of a gender gap at quantiles 0.4-0.7, as well as at lower quantiles.

Figure 9. Quantile bounds imposing the median restriction
Full sample



This figure shows the estimated quantile gender wage (female − male) conditional on average characteristics. The solid black line shows the quantile gender wage when selection is ignored. The blue and red lines with upward and downward pointing triangles show upper and lower bounds that account for employment selection for females. The dashed lines represent a uniform 90% confidence region for the bounds.

6. Conclusion

This paper provides a novel method for inference on best linear approximations to functions which are known to lie within a band. It advances the literature by allowing for bounding functions that may be estimated parametrically or non-parametrically by series estimators, and that may carry an index. Our focus on best linear approximations is motivated by the difficulty to work directly with the sharp identification region of the functions of interest, especially when the analysis is conditioned upon a large number of covariates. By contrast, best linear approximations are tractable and easy to interpret. In particular, the sharp identification region for the parameters characterizing the best linear approximation is convex, and as such can be equivalently described via its support function. The support function can in turn be estimated with a plug-in method, that replaces moments of the data with their sample analogs, and the bounding functions with their estimators. We show that the support function process approximately converges to a Gaussian process. By "approximately" we mean that while the process may not converge weakly as the number of series terms increases to infinity, each subsequence contains a further subsequence that converges weakly to a tight Gaussian process with a uniformly equicontinuous and non-degenerate covariance function. We establish validity of the Bayesian bootstrap for practical inference, and verify our regularity conditions for a large number of empirically relevant problems, including mean regression with interval valued outcome data and interval valued regressor data; quantile and distribution regression with interval valued data; sample selection problems; and mean, quantile, and distribution treatment effects.

Appendix A. Notation

$$
\begin{array}{rcl}
\mathcal{S}^{d-1} & : & =\left\{q\in\mathbb{R}^d:\|q\|=1\right\};\\
\mathbb{G}_n[h(t)] & : & =\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(h_i(t)-\mathrm{E}h(t));\\
\mathbb{G}[h_k],\ \widetilde{\mathbb{G}[h_k]} & : & =\text{ P-Brownian bridge processes, independent of each other, and}\\
 & & \quad\text{with identical distributions;}\\
\mathcal{L}^2(X,\mathrm{P}) & : & =\left\{g:X\longrightarrow\mathbb{R}\text{ s.t. }\int_X|g(x)|^2\,d\mathrm{P}(x)<\infty\right\};\\
\ell^\infty(T) & : & \text{set of all uniformly bounded real functions on } T;\\
BL_1(\ell^\infty(T),[0,1]) & : & \text{set of real functions on } \ell^\infty(T) \text{ with Lipschitz norm bounded by 1;}\\
 & \lesssim & \text{left side bounded by a constant times the right side;}\\
f^o & : & =f-\mathrm{E}f.
\end{array}
$$

Appendix B. Proof of the Results

Throughout this Appendix, we impose Conditions C1-C5.

B.1. **Proof of Theorems 1 and 2. Step 1.** We can write the difference between the estimated and true support function as the sum of three differences.

$$\widehat{\sigma}_{\widehat{\theta},\widehat{\Sigma}}-\sigma_{\theta,\Sigma}=\left(\widehat{\sigma}_{\widehat{\theta},\widehat{\Sigma}}-\widehat{\sigma}_{\theta,\widehat{\Sigma}}\right)+\left(\widehat{\sigma}_{\theta,\widehat{\Sigma}}-\widehat{\sigma}_{\theta,\Sigma}\right)+\left(\widehat{\sigma}_{\theta,\Sigma}-\sigma_{\theta,\Sigma}\right)$$

where $t\in T:=\mathcal{S}^{d-1}\times\mathcal{A}$. Let $\mu:=q'\Sigma$ and

$$w_{i,\mu}(\alpha):=\left(\theta_0(x,\alpha)1(\mu z_i<0)+\theta_1(x,\alpha)1(\mu z_i\geq 0)\right).$$

We define

$$\widehat{\sigma}_{\theta,\widehat{\Sigma}}:=\mathbb{E}_n\left[q'\widehat{\Sigma}z_iw_{i,q'\widehat{\Sigma}}(\alpha)\right]\ \text{ and }\ \widehat{\sigma}_{\theta,\Sigma}:=\mathbb{E}_n\left[q'\Sigma z_iw_{i,q'\Sigma}(\alpha)\right].$$

By Lemma 1 uniformly in $t\in T$

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\sigma}_{\widehat{\theta},\widehat{\Sigma}}-\widehat{\sigma}_{\theta,\widehat{\Sigma}}\right)(t) &= q'\Sigma\mathrm{E}[z_ip_i'1\{q'\Sigma z_i>0\}]J_1^{-1}(\alpha)\mathbb{G}_n[p_i\varphi_{i1}(\alpha)]\\
&+ q'\Sigma\mathrm{E}[z_ip_i'1\{q'\Sigma z_i<0\}]J_0^{-1}(\alpha)\mathbb{G}_n[p_i\varphi_{i0}(\alpha)]+o_{\mathrm{P}}(1).
\end{aligned}
$$

By Lemma 2 uniformly in $t\in T$

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\sigma}_{\theta,\widehat{\Sigma}}-\widehat{\sigma}_{\theta,\Sigma}\right)(t) &= \sqrt{n}q'\left(\widehat{\Sigma}-\Sigma\right)\mathrm{E}\left[z_iw_{i,q'\Sigma}(\alpha)\right]+o_{\mathrm{P}}(1)\\
&= -q'\widehat{\Sigma}\mathbb{G}_n[x_iz_i']\Sigma\mathrm{E}\left[z_iw_{i,q'\Sigma}(\alpha)\right]+o_{\mathrm{P}}(1)\\
&= -q'\Sigma\mathbb{G}_n[x_iz_i']\Sigma\mathrm{E}\left[z_iw_{i,q'\Sigma}(\alpha)\right]+o_{\mathrm{P}}(1).
\end{aligned}
$$

By definition

$$\sqrt{n}\left(\widehat{\sigma}_{\theta,\Sigma}-\sigma_{\theta,\Sigma}\right)(t)=\mathbb{G}_n[q'\Sigma z_iw_{i,q'\Sigma}(u)].$$

Putting all the terms together uniformly in $t \in T$

$$\sqrt{n}(\widehat{\sigma}_{\widehat{\theta},\widehat{\Sigma}} - \sigma_{\theta,\Sigma})(t) = \mathbb{G}_n[h_k(t)] + o_{\mathrm{P}}(1),$$

where for $t := (q, \alpha) \in T = \mathcal{S}^{d-1} \times \mathcal{A}$

$$\begin{aligned} h_k(t) := &\quad q'\Sigma \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}] J_1^{-1}(\alpha) p_i \varphi_{i1}(\alpha) \\ &+ \quad q'\Sigma \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i < 0\}] J_0^{-1}(\alpha) p_i \varphi_{i0}(\alpha) \\ &- \quad q'\Sigma x_i z_i' \Sigma \mathrm{E}\left[z_i w_{i,q'\Sigma}(\alpha)\right] \\ &+ \quad q'\Sigma z_i w_{i,q'\Sigma}(\alpha) \\ := &\quad h_{1i}(t) + h_{2i}(t) + h_{3i}(t) + h_{4i}(t), \end{aligned} \tag{B.8}$$

where k indexes the number of series terms.

Step 2. (Finite k). This case follows from $\mathcal{H} = \{h_i(t), t \in T\}$ being a Donsker class with square-integrable envelopes. Indeed, $\mathcal{H}$ is formed as finite products and sums of VC classes or entropically equivalent classes, so we can apply Lemma 8. The result

$$\mathbb{G}_n[h_k(t)] \Rightarrow \mathbb{G}[h_k(t)] \text{ in } \ell^\infty(T),$$

follows, and the assertion that

$$\mathbb{G}_n[h_k(t)] =_d \mathbb{G}[h_k(t)] + o_{\mathrm{P}}(1) \text{ in } \ell^\infty(T)$$

follows from e.g., the Skorohod-Dudley-Whichura construction. (The $=_d$ can be replaced by $=$ as in Step 3, in which case $\mathbb{G}[h_k(t)]$ is a sequence of Gaussian processes indexed by n and identically distributed for each n.)

Step 3. (Case with growing k.) This case is considerably more difficult. The main issue here is that the uniform covering entropy of $\mathcal{H}_l = \{h_{li}(t), t \in T\}$, $l = 0, 1$, grows without bound, albeit at a very slow rate $\log n$. The envelope H_l of this class also grows in general, and so we can not rely on the usual uniform entropy-based arguments; for similar reasons we can not rely on the bracketing-based entropy arguments. Instead, we rely on a strong approximation argument, using ideas in Chernozhukov, Lee, and Rosen (2009) and Belloni and Chernozhukov (2009a), to show that $\mathbb{G}_n[h_k(t)]$ can be approximated by a tight sequence of Gaussian processes $\mathbb{G}[h(t)]$, implicitly indexed by k, where the latter sequence is very well-behaved. Even though it may not converge as $k \to \infty$, for every subsequence of k there is a further subsequence along which the Gaussian process converges to a well-behaved Gaussian process. The latter is sufficient for carrying out the usual inference.

Lemma 4 below establishes that

$$\mathbb{G}_n[h(t)] = \mathbb{G}[h(t)] + o_{\mathrm{P}}(1) \text{ in } \ell^\infty(T),$$

where $\mathbb{G}[h]$ is a sequence of P-Brownian bridges with the covariance function $\mathrm{E}[h(t)h(t')] - \mathrm{E}[h(t)]\mathrm{E}[h(t')]$. Lemma 6 below establishes that for some $0 < c \leq 1/2$

$$\rho_2(h(t), h(t')) = \left(\mathrm{E}[h(t) - h(t')]^2\right)^{1/2} \lesssim \rho(t, t') := \left\|t - t'\right\|^c,$$

and the function $\mathrm{E}[h(t)h(t')] - \mathrm{E}[h(t)]\mathrm{E}[h(t')]$ is equi-continuous on $T \times T$ uniformly in k. By assumption C3 we have that $\inf_{t \in T} \mathrm{var}[h(t)] > C > 0$, with Lemma 6 providing a sufficient condition for this.

An immediate consequence of the above result is that we also obtain the convergence in the bounded Lipschitz metric

$$\sup_{g\in BL_1(\ell^\infty(T),[0,1])} |\mathrm{E}\left[g(\mathbb{G}_n[h])\right] - \mathrm{E}\left[g(\mathbb{G}[h])\right]| \le \mathrm{E}\sup_{t\in T}|\mathbb{G}_n[h(t)] - \mathbb{G}[h(t)]| \wedge 1 \to 0.$$

Step 4. Let's recognize the fact that h depends on k by using the notation h_k in this step of the proof. Note that k itself is implicitly indexed by n. Let $F_k(c) := \mathrm{P}\{f(\mathbb{G}[h_k]) \le c\}$ and observe that by Step 3 and $f \in \mathcal{F}_c$

$$\begin{aligned} & |\mathrm{P}\{f(S_n) \le c_n + o_p(1)\} - \mathrm{P}\{f(\mathbb{G}[h_k]) \le c_n\}| \\ \le\ & |\mathrm{P}\{f(\mathbb{G}[h_k]) \le c_n + o_p(1)\} - \mathrm{P}\{f(\mathbb{G}[h_k]) \le c_n\}| \\ \le\ & \delta_n(o_p(1)) \to_P 0, \quad \text{for } \delta_n(\epsilon) := \sup_{c\in\mathbb{R}} |F_k(c+\epsilon) - F_k(c)|, \end{aligned}$$

where the last step follows by the Extended Continuous Mapping Theorem (Theorem 18.11 in van der Vaart (2000)) provided that we can show that for any $\epsilon_n \searrow 0$, $\delta_n(\epsilon_n) \to 0$. Suppose otherwise, then there is a subsequence along which $\delta_n(\epsilon_n) \to \delta \neq 0$. We can select a further subsequence say $\{n_j\}$ along which the covariance function of $\mathbb{G}_n[h_k]$, denoted $\Omega_{nk}(t,t')$ converges to a covariance function $\Omega_0(t,t')$ uniformly on $T\times T$. We can do so by the Arzelà-Ascoli theorem in view of the uniform equicontinuity in k of the sequence of the covariance functions $\Omega_{nk}(t,t')$ on $T \times T$. Moreover, $\inf_{t\in T}\Omega_0(t,t) > C > 0$ by our assumption on $\Omega_{nk}(t,t')$. But along this subsequence $\mathbb{G}[h_k]$ converges in $\ell^\infty(T)$ in probability to a tight Gaussian process, say Z_0. The latter happens because $\mathbb{G}[h_k]$ converges to Z_0 marginally by Gaussianity and by $\Omega_{nk}(t,t') \to \Omega_0(t,t')$ uniformly and hence pointwise on $T \times T$ and because $\mathbb{G}[h_k]$ is asymptotically equicontinuous as shown in the proof of Lemma 4. Thus, along this subsequence we have that

$$F_k(c) \to F_0(c) = \mathrm{P}\{f(Z_0) \le c\}, \text{ uniformly in } c \in \mathbb{R},$$

because we have pointwise convergence that implies uniform convergence by Polya's theorem, since F_0 is continuous by $f \in \mathcal{F}_c$ and by $\inf_{t\in T}\Omega_0(t,t) > C > 0$. This implies that along this subsequence $\delta_{n_j}(\epsilon_{n_j}) \to 0$, which gives a contradiction.

Step 5. Finally, we observe that $c(1-\tau) = O(1)$ holds by $\sup_{t\in T}\|\mathbb{G}[h_k(t)]\| = O_{\mathrm{P}}(1)$ as shown in the proof of Lemma 4, and the second part of Theorem 2 follows. □

B.2. **Proof of Theorems 3 and 4. Step 1.** We can write the difference between a bootstrap and true support function as the sum of three differences.

$$\widetilde{\sigma}_{\widetilde{\theta},\widetilde{\Sigma}} - \sigma_{\theta,\Sigma} = \left(\widetilde{\sigma}_{\widetilde{\theta},\widetilde{\Sigma}} - \widetilde{\sigma}_{\theta,\widetilde{\Sigma}}\right) + \left(\widetilde{\sigma}_{\theta,\widetilde{\Sigma}} - \widetilde{\sigma}_{\theta,\Sigma}\right) + \left(\widetilde{\sigma}_{\theta,\Sigma} - \sigma_{\theta,\Sigma}\right)$$

where for

$$w_{i,\mu}(\alpha) =: (\theta_0(x,\alpha)1(\mu z_i < 0) + \theta_1(x,\alpha)1(\mu z_i \ge 0))$$

we define

$$\widetilde{\sigma}_{\theta,\widetilde{\Sigma}} := \mathbb{E}_n\left[(e_i/\bar{e})q'\widetilde{\Sigma}'z_i w_{i,q'\widetilde{\Sigma}}(\alpha)\right] \text{ and } \widetilde{\sigma}_{\theta,\Sigma} =: \mathbb{E}_n\left[(e_i/\bar{e})q'\Sigma' z_i w_{i,q'\Sigma}(\alpha)\right],$$

where $\bar{e} = \mathbb{E}_n e_i \to_{\mathrm{P}} 1$.

By Lemma 1 uniformly in $t \in T$

$$\begin{aligned}\sqrt{n}\left(\widetilde{\sigma}_{\widetilde{\theta},\widetilde{\Sigma}} - \widetilde{\sigma}_{\theta,\widetilde{\Sigma}}\right)(t) &= q'\Sigma \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}] J_1^{-1}(\alpha)\mathbb{G}_n[e_i p_i \varphi_{i1}(\alpha)] \\ &+ q'\Sigma \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i < 0\}] J_0^{-1}(\alpha)\mathbb{G}_n[e_i p_i \varphi_{i0}(\alpha)] + o_{\mathrm{P}}(1).\end{aligned}$$

By Lemma 2 uniformly in $t \in T$

$$\begin{aligned}\sqrt{n}\left(\widetilde{\sigma}_{\theta,\widetilde{\Sigma}} - \widetilde{\sigma}_{\theta,\Sigma}\right)(t) &= \sqrt{n}q'\left(\widetilde{\Sigma} - \Sigma\right)\mathrm{E}\left[z_i w_{i,q'\Sigma}(\alpha)\right] + o_{\mathrm{P}}(1) \\ &= q'\widetilde{\Sigma}\mathbb{G}_n[(e_i/\bar{e})(x_i z_i')^o]\Sigma \mathrm{E}\left[z_i w_{i,q'\Sigma}(\alpha)\right] + o_{\mathrm{P}}(1) \\ &= q'\Sigma\mathbb{G}_n[e_i(x_i z_i')^o]\Sigma \mathrm{E}\left[z_i w_{i,q'\Sigma}(\alpha)\right] + o_{\mathrm{P}}(1).\end{aligned}$$

By definition

$$\sqrt{n}\left(\widetilde{\sigma}_{\theta,\Sigma} - \sigma_{\theta,\Sigma}\right)(t) = \mathbb{G}_n[e_i\left(q'\Sigma z_i w_{i,q'\Sigma}(\alpha)\right)^o]/\bar{e} = \mathbb{G}_n[e_i\left(q'\Sigma z_i w_{i,q'\Sigma}(\alpha)\right)^o](1 + o_p(1)).$$

Putting all the terms together uniformly in $t \in T$

$$\sqrt{n}(\widetilde{\sigma}_{\widetilde{\theta},\widetilde{\Sigma}} - \sigma_{\theta,\Sigma})(t) = \mathbb{G}_n[e_i h_i^o(t)] + o_{\mathrm{P}}(1).$$

Step 2. Combining conclusions of Theorems 1 and Step 1 above we obtain:

$$\begin{aligned}&\tilde{S}_n(t) = \sqrt{n}(\widetilde{\sigma}_{\widetilde{\theta},\widetilde{\Sigma}} - \widehat{\sigma}_{\widehat{\theta},\widehat{\Sigma}})(t) \\ =\ & \sqrt{n}(\widetilde{\sigma}_{\widetilde{\theta},\widetilde{\Sigma}} - \sigma_{\theta,\Sigma})(t) - \sqrt{n}(\widehat{\sigma}_{\widehat{\theta},\widehat{\Sigma}} - \sigma_{\theta,\Sigma})(t) \\ =\ & \mathbb{G}_n[e_i h_i^o(t)] - \mathbb{G}_n[h(t)] + o_{\mathrm{P}}(1) \\ =\ & \mathbb{G}_n[e_i^o h_i^o(t)] + o_{\mathrm{P}}(1).\end{aligned}$$

Observe that the bootstrap process $\mathbb{G}_n[e_i^o h_i^o(t)]$ has the unconditional covariance function

$$\mathrm{E}[h(t)h(t')] - \mathrm{E}[h(t)]\mathrm{E}[h(t')],$$

which is equal to the covariance function of the original process $\mathbb{G}_n[h_i]$. Conditional on data the covariance function of this process is

$$\mathbb{E}_n[h(t)h(t')] - \mathbb{E}_n[h(t)]\mathbb{E}_n[h(t')].$$

Comment B.1. Note that if a bootstrap random element Z_n taking values in a normed space $(E, \|\cdot\|)$ converges in probability P unconditionally, that is $Z_n = o_{\mathrm{P}}(1)$, then $Z_n = o_{\mathrm{P}^e}(1)$ in $L^1(P)$ sense and hence probability P, where P^e denotes the probability measure conditional on the data. In other words, Z_n also converges in probability conditionally on the data. This follows because $\mathrm{E}_{\mathrm{P}}|\mathrm{P}^e\{\|Z_n\| > \epsilon\}| = \mathrm{P}\{\|Z_n\| > \epsilon\} \to 0$, so that $\mathrm{P}^e\{\|Z_n\| > \epsilon\} \to 0$ in $L^1(P)$ sense and hence in probability P. Similarly, if $Z_n = O_{\mathrm{P}}(1)$, then $Z_n = O_{\mathrm{P}^e}(1)$ in probability P.

Step 3. (Finite k). This case follows from $\mathcal{H} = \{h_i(t), t \in T\}$ being a Donsker class with square-integrable envelopes. Indeed, $\mathcal{H}$ is formed as a Lipschitz composition of VC classes or entropically equivalent classes. Then by the Donsker theorem for exchangeable bootstraps, see e.g., van der Vaart and Wellner (1996), we have weak convergence conditional on the data

$$\mathbb{G}_n[e_i^o h_i^o(t)]/\bar{e} \Rightarrow \widetilde{\mathbb{G}[h(t)]} \text{ under } \mathrm{P}^e \text{ in } \ell^\infty(T) \text{ in probability P},$$

where $\widetilde{\mathbb{G}[h]}$ is a sequence of P-Brownian bridges independent of $\mathbb{G}[h]$ and with the same distribution as $\mathbb{G}[h]$. In particular, the covariance function of $\widetilde{\mathbb{G}[h]}$ is $\mathrm{E}[h(t)h(t')]-\mathrm{E}[h(t)]\mathrm{E}[h(t')]$. Since $\bar{e} \to_{\mathrm{P}^e} 1$, the above implies

$$\mathbb{G}_n[e_i^o h_i^o(t)] \Rightarrow \widetilde{\mathbb{G}[h(t)]} \text{ under } \mathrm{P}^e \text{ in } \ell^\infty(T) \text{ in probability P.}$$

The latter statement simply means

$$\sup_{g \in BL_1(\ell^\infty(T),[0,1])} |\mathrm{E}_{\mathrm{P}^e}[g(\mathbb{G}_n[h])] - \mathrm{E}[g(\widetilde{\mathbb{G}[h]})]| \to_{\mathrm{P}} 0.$$

This statement can be strengthened to a coupling statement as in Step 4.

Step 4. (Growing k.) By Lemma 4 below we can show that (on a suitably extended probability space) there exists a sequence of Gaussian processes $\widetilde{\mathbb{G}[h(t)]}$ such that

$$\mathbb{G}_n[e_i^o h_i^o(t)] = \widetilde{\mathbb{G}[h(t)]} + o_{\mathrm{P}}(1) \text{ in } \ell^\infty(T),$$

which implies by Remark B.1 that

$$\mathbb{G}_n[e_i^o h_i^o(t)] = \widetilde{\mathbb{G}[h(t)]} + o_{\mathrm{P}^e}(1) \text{ in } \ell^\infty(T) \text{ in probability.}$$

Here, as above, $\widetilde{\mathbb{G}[h]}$ is a sequence of P-Brownian bridges independent of $\mathbb{G}[h]$ and with the same distribution as $\mathbb{G}[h]$. In particular, the covariance function of $\widetilde{\mathbb{G}[h]}$ is $\mathrm{E}[h(t)h(t')] - \mathrm{E}[h(t)]\mathrm{E}[h(t')]$. Lemma 6 describes the properties of this covariance function, which in turn define the properties of this Gaussian process.

An immediate consequence of the above result is the convergence in bounded Lipschitz metric

$$\sup_{g \in BL_1(\ell^\infty(T),[0,1])} \left| \mathrm{E}_{\mathrm{P}^e}[g(\mathbb{G}_n[e_i^o h_i^o(t)])] - \mathrm{E}_{\mathrm{P}^e}[g(\widetilde{\mathbb{G}[h]})] \right| \leq \mathrm{E}_{\mathrm{P}^e} \sup_{t \in T} |\mathbb{G}_n[e_i^o h_i^o(t)] - \widetilde{\mathbb{G}[h(t)]}| \wedge 1 \to_{\mathrm{P}} 0.$$

Note that $\mathrm{E}_{\mathrm{P}^e}[g(\widetilde{\mathbb{G}[h]})] = \mathrm{E}_{\mathrm{P}}[g(\widetilde{\mathbb{G}[h]})]$, since the covariance function of $\widetilde{\mathbb{G}[h]}$ does not depend on the data. Therefore

$$\sup_{g \in BL_1(\ell^\infty(T),[0,1])} |\mathrm{E}_{\mathrm{P}^e}[g(\mathbb{G}_n[e_i^o h_i^o(t)])] - \mathrm{E}_{\mathrm{P}}[g(\widetilde{\mathbb{G}[h]})]| \to_{\mathrm{P}} 0.$$

Step 5. Let us recognize the fact that h depends on k by using the notation h_k in this step of the proof. Note that k itself is implicitly indexed by n. By the previous steps and Theorem 1 there exist $\epsilon_n \searrow 0$ such that $\pi_1 = \mathrm{P}^e\{|f(\widetilde{S}_n) - f(\widetilde{\mathbb{G}[h_k]})| > \epsilon_n\}$ and $\pi_2 = \mathrm{P}\{|f(\widetilde{S}_n) - f(\mathbb{G}[h_k])| > \epsilon_n\}$ obey $\mathrm{E}[\pi_1] \to_P 0$ and $\pi_2 \to 0$. Let

$$F(c) := \mathrm{P}\{f(\mathbb{G}[h_k]) \leq c\} = \mathrm{P}\{f(\widetilde{\mathbb{G}[h_k]}) \leq c\} = \mathrm{P}^e\{f(\mathbb{G}[h_k]) \leq c\},$$

where the equality holds because $\mathbb{G}[h_k]$ and $\widetilde{\mathbb{G}[h_k]}$ are P-Brownian bridges with the same covariance kernel, which in the case of the bootstrap does not depend on the data.

For any c_n which is a measurable function of the data,

$$\begin{aligned}
& \mathrm{E}|\mathrm{P}^e\{f(\widetilde{S}_n) \le c_n\} - \mathrm{P}\{f(S_n) \le c_n\}| \\
\le\ & \mathrm{E}[\mathrm{P}^e\{f(\widetilde{\mathbb{G}[h_k]}) \le c_n + \epsilon_n\} - \mathrm{P}\{f(\mathbb{G}[h_k]) \le c_n - \epsilon_n\} + \pi_1 + \pi_2] \\
=\ & \mathrm{E}F(c_n + \epsilon_n) - \mathrm{E}F(c_n - \epsilon_n) + o(1) \\
\le\ & \sup_{c \in \mathbb{R}} |F(c + \epsilon_n) - F(c - \epsilon_n)| + o(1) = o(1),
\end{aligned}$$

where the last step follows from the proof of Theorem 1. This proves the first claim of Theorem 4 by the Chebyshev inequality. The second claim of Theorem 4 follows similarly to Step 5 in the proof of Theorems 1-2. □

B.3. Main Lemmas for the Proofs of Theorems 1 and 2.

Lemma 1 (Linearization)**.** *1. (Sample) We have that uniformly in $t \in T$*

$$\begin{aligned}
\sqrt{n}\left(\widehat{\sigma}_{\widehat{\theta},\widehat{\Sigma}} - \widehat{\sigma}_{\theta,\widehat{\Sigma}}\right) &= q'\widehat{\Sigma}\sqrt{n}\left(\mathbb{E}_n z_i \left(\widehat{\theta}_{1,i}(\alpha) - \theta_{1,i}(\alpha)\right) 1\left\{q'\widehat{\Sigma} z_i > 0\right\}\right) \\
&+ q'\widehat{\Sigma}\sqrt{n}\left(\mathbb{E}_n z_i \left(\widehat{\theta}_{0,i}(\alpha) - \theta_{0,i}(\alpha)\right) 1\left\{q'\widehat{\Sigma} z_i < 0\right\}\right) \\
&= q'\Sigma \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}] J_1^{-1}(\alpha)\mathbb{G}_n[p_i \varphi_{i1}(\alpha)] \\
&+ q'\Sigma \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i < 0\}] J_0^{-1}(\alpha)\mathbb{G}_n[p_i \varphi_{i0}(\alpha)] + o_{\mathrm{P}}(1).
\end{aligned}$$

2. (Bootstrap) We have that uniformly in $t \in T$

$$\begin{aligned}
\sqrt{n}\left(\widetilde{\sigma}_{\widetilde{\theta},\widetilde{\Sigma}} - \widetilde{\sigma}_{\theta,\widetilde{\Sigma}}\right)(t) &= q'\widetilde{\Sigma}\sqrt{n}\left(\mathbb{E}_n (e_i/\bar{e}) z_i \left(\widetilde{\theta}_{1,i}(\alpha) - \theta_{1,i}(\alpha)\right) 1\left\{q'\widetilde{\Sigma} z_i > 0\right\}\right) \\
&+ q'\widetilde{\Sigma}\sqrt{n}\left(\mathbb{E}_n (e_i/\bar{e}) z_i \left(\widetilde{\theta}_{0,i}(\alpha) - \theta_{0,i}(\alpha)\right) 1\left\{q'\widetilde{\Sigma} z_i < 0\right\}\right) \\
&= q'\Sigma \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}] J_1^{-1}(\alpha)\mathbb{G}_n[e_i p_i \varphi_{i1}(\alpha)] \\
&+ q'\Sigma \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i < 0\}] J_0^{-1}(\alpha)\mathbb{G}_n[e_i p_i \varphi_{i0}(\alpha)] + o_{\mathrm{P}}(1).
\end{aligned}$$

Proof of Lemma 1. In order to cover both cases with one proof, we will use $\bar{\theta}$ to mean either the unweighted estimator $\widehat{\theta}$ or the weighted estimator $\tilde{\theta}$ and so on, and v_i to mean either 1 in the case of the unweighted estimator or exponential weights e_i in the case of the weighted estimator. We also observe that $\bar{\Sigma} \to_P \Sigma$ by the law of large numbers and the continuous mapping theorem.

Step 1. It will suffice to show that

$$\begin{aligned}
& q'\bar{\Sigma}\sqrt{n}\left(\mathbb{E}_n (v_i/\bar{v}) z_i \left(\bar{\theta}_{1,i}(\alpha) - \theta_{1,i}(\alpha)\right) 1\left\{q'\bar{\Sigma} z_i > 0\right\}\right) \\
=\ & q'\Sigma \mathrm{E}[z_i p_i' 1\{q\Sigma z_i > 0\}] J_1^{-1}(\alpha)\mathbb{G}_n[v_i p_i \varphi_{i1}(\alpha)] + o_{\mathrm{P}}(1)
\end{aligned}$$

and that

$$\begin{aligned}
& q'\bar{\Sigma}\sqrt{n}\left(\mathbb{E}_n (v_i/\bar{v}) z_i \left(\bar{\theta}_{0,i}(\alpha) - \theta_{0,i}(\alpha)\right) 1\left\{q'\bar{\Sigma} z_i < 0\right\}\right) \\
=\ & q'\Sigma \mathrm{E}[z_i p_i' 1\{q\Sigma z_i < 0\}] J_0^{-1}(\alpha)\mathbb{G}_n[v_i p_i \varphi_{i0}(\alpha)] + o_{\mathrm{P}}(1).
\end{aligned}$$

We show the argument for the first part; the argument for the second part is identical. We also drop the index $\ell = 1$ to ease notation. By Assumption C2 we write

$$\begin{aligned} q'\bar{\Sigma}\sqrt{n}\mathbb{E}_n(v_i/\bar{v})z_i\left(\bar{\theta}_i - \theta_i\right)1\left\{q'\bar{\Sigma}z_i > 0\right\} &= \left\{q'\bar{\Sigma}\mathbb{E}_n[(v_i z_i p_i' 1\{q'\bar{\Sigma}z_i > 0\}]J^{-1}(\alpha)\mathbb{G}_n\left[v_i p \varphi_i(\alpha)\right]\right. \\ &\quad \left. + q'\bar{\Sigma}\mathbb{E}_n[v_i z_i \bar{R}_i(\alpha)1\{q'\bar{\Sigma}z_i > 0\}]\right\}/\bar{v} \\ &=: (a(\alpha) + b(\alpha))/(1 + o_{\mathrm{P}}(1)). \end{aligned}$$

We have from the assumptions of the theorem

$$\sup_{\alpha\in\mathcal{A}}|b(\alpha)| \leq \|q'\bar{\Sigma}\| \cdot \|\|z_i\|\|_{\mathbb{P}_n,2}\|\|v_i\|\|_{\mathbb{P}_n,2} \cdot \sup_{\alpha\in\mathcal{A}}\|\bar{R}_i(\alpha)\|_{\mathbb{P}_n,2} = O_{\mathrm{P}}(1)O_{\mathrm{P}}(1)o_{\mathrm{P}}(1) = o_{\mathrm{P}}(1).$$

Write $a(\alpha) = c(\alpha) + d(\alpha)$, where

$$\begin{aligned} c(\alpha) &:= q'\Sigma\mathrm{E}[z_i p_i' 1\{q\Sigma z_i > 0\}]J^{-1}(\alpha)\mathbb{G}_n\left[v_i p_i \varphi_i(\alpha)\right] \\ d(\alpha) &:= \bar{\mu}' J^{-1}(\alpha)\mathbb{G}_n\left[v_i p_i \varphi_i(\alpha)\right] \\ \bar{\mu}' &:= q'\bar{\Sigma}\mathbb{E}_n[v_i z_i p_i' 1\{q'\bar{\Sigma}z_i > 0\}] - q'\Sigma\mathrm{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}] \end{aligned} \quad \text{(B.9)}$$

The claim follows after showing that $\sup_{\alpha\in\mathcal{A}}|d(\alpha)| = o_{\mathrm{P}}(1)$, which is shown in subsequent steps below.

Step 2. (Special case, with k fixed). This is the parametric case, which is trivial. In this step we have to show $\sup_{\alpha\in\mathcal{A}}|d(\alpha)| = o_{\mathrm{P}}(1)$. We can write

$$d(\alpha) = \mathbb{G}_n[\bar{\mu}' f_\alpha], \quad f_\alpha := (f_{\alpha j}, j = 1, ..., k)\,, \; f_{\alpha j} := J^{-1}(\alpha)v_i p_{ij}\varphi_i(\alpha)$$

and define the function class $\mathcal{F} := \{f_{\alpha j}, \alpha \in \mathcal{A}, j = 1, ..., k\}$. Since k is finite, and given the assumptions on $\mathcal{F}_1 = \{\varphi(\alpha), \alpha \in \mathcal{A}\}$, application of Lemmas 8 and 9-2(a) yields

$$\sup_Q \log N(\epsilon\|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \lesssim \log(1/\epsilon).$$

and the envelope is P-square integrable. Therefore, $\mathcal{F}$ is P-Donsker and

$$\sup_{\alpha\in\mathcal{A}}|\mathbb{G}_n[f_\alpha]| \lesssim_{\mathrm{P}} 1$$

and $\sup_{\alpha\in\mathcal{A}}|d(\alpha)| \lesssim_{\mathrm{P}} k\|\bar{\mu}\| \to_{\mathrm{P}} 0$.

Step 3. (General case, with $k \to \infty$). In this step we have to show $\sup_{\alpha\in\mathcal{A}}|d(\alpha)| = o_{\mathrm{P}}(1)$. The case of $k \to \infty$ is much more difficult if we want to impose rather weak conditions on the number of series terms. We can write

$$d(\alpha) = \mathbb{G}_n[f_{\alpha n}], \quad f_{\alpha n} := \bar{\mu}' J^{-1}(\alpha)v_i p_i \varphi_i(\alpha)$$

and define the function class $\mathcal{F}_3 := \{f_{\alpha n}, \alpha \in \mathcal{A}\}$, see equation (B.13) below. By Lemma 9 the random entropy of this function class obeys

$$\log N(\epsilon\|F_3\|_{\mathbb{P}_n,2}, \mathcal{F}_3, L_2(\mathbb{P}_n)) \lesssim_{\mathrm{P}} \log n + \log(1/\epsilon).$$

Therefore by Lemma 11, conditional on $X_n = (x_i, z_i, i = 1, ..., n)$, for each $\delta > 0$ there exists a constant K_δ, that does not depend on n, such that for all n:

$$\mathrm{P}\left\{\sup_{\alpha\in\mathcal{A}}|d(\alpha)| \geq K_\delta\sqrt{\log n}\left(\sup_{\alpha\in\mathcal{A}}\|f_{\alpha n}\|_{\mathbb{P}_n,2} \vee \sup_{\alpha\in\mathcal{A}}\|f_{\alpha n}\|_{\mathrm{P}|X_n,2}\right)\right\} \leq \delta,$$

where $\mathrm{P}|X_n$ denotes the probability measure conditional on X_n. The conclusion follows if we can demonstrate that $\sup_{\alpha\in\mathcal{A}} \|f_{\alpha n}\|_{\mathbb{P}_n,2} \vee \sup_{\alpha\in\mathcal{A}} \|f_{\alpha n}\|_{\mathrm{P}|X_n,2} \to_{\mathrm{P}} 0$. To show this note that

$$\sup_{\alpha\in\mathcal{A}} \|f_{\alpha n}\|_{\mathbb{P}_n,2} \leq \|\bar{\mu}\| \sup_{\alpha\in\mathcal{A}} \|J^{-1}(\alpha)\| \|\mathbb{E}_n p_i p_i'\| \cdot \sup_{i\leq n} v_i \sup_{i\leq n,\alpha\in\mathcal{A}} |\varphi_i(\alpha)| \to_{\mathrm{P}} 0,$$

where the convergence to zero in probability follows because

$$\|\bar{\mu}\| \lesssim_{\mathrm{P}} n^{-m/4} + \sqrt{(k/n)\cdot \log n} \cdot (\log n \max_i \|z_i\|) \wedge \xi_k, \quad \sup_{i\leq n} v_i \lesssim_{\mathrm{P}} \log n$$

by Step 4 below, $\sup_{\alpha\in\mathcal{A}} \|J^{-1}(\alpha)\| \lesssim 1$ by assumption C3, $\|\mathbb{E}_n p_i p_i'\| \lesssim_{\mathrm{P}} 1$ by Lemma 10, and

$$\log^2 n \left(n^{-m/4} + \sqrt{(k/n)\cdot \log n} \cdot \max_i \|z_i\| \wedge \xi_k \right) \sup_{i\leq n,\alpha\in\mathcal{A}} |\varphi_i(\alpha)| \to_{\mathrm{P}} 0$$

by assumption C4. Also note that

$$\sup_{\alpha\in\mathcal{A}} \|f_{\alpha n}\|_{\mathrm{P}|X_n,2} \leq \|\bar{\mu}\| \sup_{\alpha\in\mathcal{A}} \|J^{-1}(\alpha)\| \|\mathbb{E}_n p_i p_i'\| \cdot (\mathrm{E}[v_i^2])^{1/2} \cdot \sup_{i\leq n,\alpha\in\mathcal{A}} \left[\mathrm{E}[\varphi_i^2(u)|x_i,z_i]\right]^{1/2} \to_{\mathrm{P}} 0,$$

by the preceding argument and $\mathrm{E}[\varphi_i^2(u)|x_i,z_i]$ uniformly bounded in α and i by assumption C3.

Step 4. In this step we show that

$$\|\bar{\mu}\| \lesssim_{\mathrm{P}} n^{-m/4} + \sqrt{(k/n)\cdot \log n} \cdot (\log n \max_i \|z_i\|) \wedge \xi_k.$$

We can bound

$$\|\bar{\mu}\| \leq \|\Sigma - \bar{\Sigma}\| \|\mathrm{E}\left[z_i p_i 1\{q'\Sigma z_i > 0\}\right]\| + \|\bar{\Sigma}\|\mu_1 + \|\bar{\Sigma}\|\mu_2,$$

where

$$\begin{aligned} \mu_1 &= \|\mathbb{E}_n\left[v_i z_i p_i' 1\{q'\Sigma z_i > 0\}\right] - \mathrm{E}\left[z_i p_i' 1\{q'\Sigma z_i > 0\}\right]\| \\ \mu_2 &= \|\mathbb{E}_n[v_i z_i p_i' \{1\{q'\Sigma z_i > 0\} - 1\{q'\bar{\Sigma} z_i > 0\}\}]\|. \end{aligned}$$

By Lemma 10, $\|\Sigma - \bar{\Sigma}\| = o_{\mathrm{P}}(1)$, and from Assumption C3 $\|\mathrm{E}\left[z_i p_i 1\{q'\Sigma z_i > 0\}\right]\| \lesssim 1$.

By elementary inequalities

$$\bar{\mu}_2^2 \leq \mathbb{E}_n\|v_i\|^2 \mathbb{E}_n \|z_i\|^2 \|\mathbb{E}_n[p_i p_i']\| \mathbb{E}_n[\{1\{q'\Sigma z_i > 0\} - 1\{q'\bar{\Sigma} z_i > 0\}\}^2] \lesssim_{\mathrm{P}} n^{-m/2},$$

where we used the Chebyshev inequality along with $\mathrm{E}\|v_i\|^2 = 1$ and $\mathrm{E}[\|z_i\|^2] < \infty$, $\|\mathbb{E}_n[p_i p_i']\| \lesssim_{\mathrm{P}} 1$ by Lemma 10, and $\mathbb{E}_n[\{1\{q'\Sigma z_i > 0\} - 1\{q'\bar{\Sigma} z_i > 0\}\}^2] \lesssim_{\mathrm{P}} n^{-m/2}$ by Step 5 below.

We can write $\mu_1 = \sup_{g\in\mathcal{G}} |\mathbb{E}_n g - \mathrm{E} g|$, where $\mathcal{G} := \{v_i \gamma' z_i p_i' \eta 1\{q'\Sigma z_i > 0\}, \|\gamma\| = 1, \|\eta\| = 1\}$. The function class $\mathcal{G}$ obeys

$$\sup_Q \log N(\epsilon \|G\|_{Q,2}, \mathcal{G}, L_2(Q)) \lesssim (\dim(z_i) + \dim(p_i)) \log(1/\epsilon) \lesssim k \log(1/\epsilon)$$

for the envelope $G_i = v_i \|z\|_i \cdot \xi_k$ that obeys $\max_i \log G_i \lesssim_{\mathrm{P}} \log n$ by $E|v_i|^p < \infty$ for any $p > 0$, $\mathrm{E}\|z_i\|^2 < \infty$ and $\log \xi_k \lesssim_{\mathrm{P}} \log n$. Invoking Lemma 11 we obtain

$$\mu_1 \lesssim_{\mathrm{P}} \sqrt{(k/n)\cdot \log n} \times \sup_{g\in\mathcal{G}} \|g\|_{\mathbb{P}_n,2} \vee \sup_{g\in\mathcal{G}} \|g\|_{\mathrm{P},2},$$

where

$$\sup_{g\in\mathcal{G}} \|g\|_{\mathbb{P}_n,2} \lesssim_{\mathrm{P}} \left(\max_i \|z_i\| \max_i v_i \cdot \|\mathbb{E}_n[p_i p_i']\|\right) \wedge \left(\left[\mathbb{E}_n \|v_i z_i\|^2\right]^{1/2} \xi_k\right)$$
$$\lesssim_{\mathrm{P}} (\max_i v_i \max_i \|z_i\|) \wedge \xi_k \lesssim_{\mathrm{P}} (\log n \max_i \|z_i\|) \wedge \xi_k$$

by $\mathrm{E}\|z_i\|^2 < \infty$ and by $\mathbb{E}_n[p_i p_i'] \lesssim_{\mathrm{P}} 1$, $\max_i v_i \lesssim_{\mathrm{P}} \log n$ and $\sup_{g\in\mathcal{G}} \|g\|_{\mathrm{P},2} = \|E z_i p_i'\| \lesssim 1$ by Assumption C3. Thus

$$\mu_1 \lesssim_{\mathrm{P}} \sqrt{(k/n)\cdot \log n}(\max_i \|z_i\| \log n) \wedge \xi_k,$$

and the claim of the step follows.

Step 5. Here we show

$$\sup_{q\in\mathcal{S}^{d-1}} \mathbb{E}_n \left[\left(1(q'\Sigma z_i < 0) - 1(q'\bar{\Sigma} z_i < 0)\right)^2\right] \lesssim_{\mathrm{P}} n^{-m/2}.$$

Note $\left(1(q'\Sigma z_i < 0) - 1(q'\bar{\Sigma} z_i < 0)\right)^2 = 1(q'\Sigma z_i < 0 < q'\bar{\Sigma} z_i) + 1(q'\Sigma z_i > 0 > q'\bar{\Sigma} z_i)$. The set

$$\mathcal{F} = \left\{1(q'\Sigma z_i < 0 < q'\tilde{\Sigma} z_i) + 1(q'\Sigma z_i > 0 > q'\tilde{\Sigma} z_i), q \in \mathcal{S}^{d-1}, \|\Sigma\| \leq M, \|\tilde{\Sigma}\| \leq M\right\}$$

is P-Donsker because it is a VC class with a constant envelope. Therefore, $|\mathbb{E}_n f - \mathrm{E} f| \lesssim_{\mathrm{P}} n^{-1/2}$ uniformly on $f \in \mathcal{F}$. Hence uniformly in $q \in \mathcal{S}^{d-1}$, $\mathbb{E}_n[(1(q'\Sigma z_i < 0) - 1(q'\bar{\Sigma}' z_i < 0))^2]$ is equal to

$$\begin{aligned} & \mathrm{E}\left[1(q'\Sigma z_i < 0 < q'\bar{\Sigma}' z_i) + 1(q'\Sigma z_i > 0 > q'\bar{\Sigma}' z_i)\right] + O_{\mathrm{P}}\left(n^{-1/2}\right) \\ = \;& \mathrm{P}\left(\left|q'\Sigma z_i\right| < \left|q'\left(\Sigma - \bar{\Sigma}\right) z_i\right|\right) + O_{\mathrm{P}}\left(n^{-1/2}\right) \\ \leq \;& \|\Sigma - \bar{\Sigma}\|^m + O_{\mathrm{P}}\left(n^{-1/2}\right) \lesssim_{\mathrm{P}} n^{-m/2} + n^{-1/2} \lesssim_{\mathrm{P}} n^{-m/2} \end{aligned}$$

where we are using that for $0 < m \leq 1$

$$\mathrm{P}\left(\left|q'\Sigma z_i\right| < \left|q'\left(\Sigma - \bar{\Sigma}\right) z_i\right|\right) \leq \mathrm{P}\left(|q'\Sigma z_i/\|z_i\|| < \|q\| \|\Sigma - \bar{\Sigma}\|\right) \lesssim \|\bar{\Sigma} - \Sigma\|^m,$$

where the last inequality holds by Assumption C1, which gives that $\mathrm{P}\left(|q'\Sigma z_i/\|z_i\|| < \delta\right)/\delta^m \lesssim 1$. □

Lemma 2. *Let* $w_{i,\mu}(\alpha) =: (\theta_0(x,\alpha) 1(\mu z_i < 0) + \theta_1(x,\alpha) 1(\mu z_i \geq 0))$. *1. (Sample) Then uniformly in* $t \in T$

$$\sqrt{n}\left(\widehat{\sigma}_{\theta,\widehat{\Sigma}} - \widehat{\sigma}_{\theta,\Sigma}\right)(t) = \sqrt{n} q' \left(\widehat{\Sigma} - \Sigma\right) \mathrm{E}\left[z_i w_{i,q'\Sigma}(\alpha)\right] + o_{\mathrm{P}}(1)$$

2. (Bootstrap) Then uniformly in $t \in T$

$$\sqrt{n}\left(\tilde{\sigma}_{\theta,\tilde{\Sigma}} - \tilde{\sigma}_{\theta,\Sigma}\right)(t) = \sqrt{n} q' \left(\tilde{\Sigma} - \Sigma\right) \mathrm{E}\left[z_i w_{i,q'\Sigma}(\alpha)\right] + o_{\mathrm{P}}(1)$$

Proof of Lemma 2. In order to cover both cases with one proof, we will use $\bar{\theta}$ to mean either the unweighted estimator $\widehat{\theta}$ or the weighted estimator $\tilde{\theta}$ and so on, and v_i to mean either 1 in the case of the unweighted estimator or exponential weights e_i in the case of the weighted estimator. We also observe that $\bar{\Sigma} \to_{\mathrm{P}} \Sigma$ by the law of large numbers (Lemma 10) and the continuous mapping theorem.

Step 1. Define $\mathcal{F} = \left\{ q'\Sigma z_i w_{i,q'\Sigma}(t) : t \in T, \|\Sigma\| \le C \right\}$. We have that for $\bar{f}_i(t) = q'\bar{\Sigma} z_i w_{i,q'\bar{\Sigma}}(t)$ and $f_i(t) = q'\Sigma z_i w_{i,q'\Sigma}(t)$ by definition

$$\begin{aligned}
\sqrt{n}\big(\bar{\sigma}_{\theta,\bar{\Sigma}} - \bar{\sigma}_{\theta,\Sigma}\big)(t) &= \sqrt{n}\mathbb{E}_n[(v_i/\bar{v})(\bar{f}_i(t) - f_i(t))] \\
&= \sqrt{n}\mathrm{E}[\bar{f}_i(t) - f_i(t)] + \mathbb{G}_n[v_i(\bar{f}_i(t) - f_i(t))^o]/\bar{v} \\
&= \sqrt{n}\mathrm{E}[\bar{f}_i(t) - f_i(t)] + \mathbb{G}_n[v_i(\bar{f}_i(t) - f_i(t))^o]/(1 + o_{\mathrm{P}}(1)).
\end{aligned}$$

By intermediate value expansion and Lemma 3, uniformly in $\alpha \in \mathcal{A}$ and $q \in \mathcal{S}^{d-1}$

$$\sqrt{n}\left(\mathrm{E}[\bar{f}_i(t) - f_i(t)]\right) = \sqrt{n}(q'\bar{\Sigma} - q'\Sigma)\mathrm{E}[z_i w_{i,q'\Sigma^*(t)}(t)] = \sqrt{n}(q'\bar{\Sigma} - q'\Sigma)\mathrm{E}[z_i w_{i,q'\Sigma}(t)] + o_{\mathrm{P}}(1),$$

for $q\Sigma^*(t)$ on the line connecting $q'\Sigma$ and $q'\bar{\Sigma}$, where the last step follows by the uniform continuity of the mapping $(\alpha, q'\Sigma) \mapsto \mathrm{E}[z_i w_{i,q'\Sigma}(t)]$ and $q'\bar{\Sigma} - q'\Sigma \to_{\mathrm{P}} 0$. Furthermore $\sup_{t\in T} |\mathbb{G}_n[v_i(\bar{f}_i(t) - f_i(t))^o]| \to_{\mathrm{P}} 0$ by Step 2 below, proving the claim of the Lemma.

Step 2. It suffices to show that for any $t \in T$, we have that $\mathbb{G}_n[v_i\left[\bar{f}_i(t) - f_i(t)\right]^o] \to_{\mathrm{P}} 0$. By Lemma 19.24 from van der Vaart (2000) it follows that if $v_i\left[\bar{f}_i(t) - f_i(t)\right]^o \in \mathcal{G} = v_i((\mathcal{F} - \mathcal{F})^o)$ is such that

$$\left(\mathrm{E}\left[(v_i(\bar{f}_i(t) - f_i(t))^o)^2\right]\right)^{1/2} \le 2\left(\mathrm{E}\left[(v_i(\bar{f}_i(t) - f_i(t)))^2\right]\right)^{1/2} \to_{\mathrm{P}} 0,$$

and $\mathcal{G}$ is P-Donsker, then $\mathbb{G}_n[v_i(\bar{f}_i(t) - f_i(t))^o] \to_{\mathrm{P}} 0$. Here $\mathcal{G}$ is P-Donsker because $\mathcal{F}$ is a P-Donsker class formed by taking products of $\mathcal{F}_2 \supseteq \{\theta_{i\ell}(\alpha) : \alpha \in \mathcal{A}, \ell = 0, 1\}$, which possess a square-integrable envelope, with bounded VC classes $\{1(q'\Sigma z_i > 0), q \in \mathcal{S}^{d-1}, \|\Sigma\| \le C\}$ and $\{1(q'\Sigma z_i \le 0), q \in \mathcal{S}^{d-1}, \|\Sigma\| \le C\}$ and then summing followed by demeaning. The difference $(\mathcal{F} - \mathcal{F})^o$ is also P-Donsker, and its product with the independent square-integrable variable v_i is still a P-Donsker class with a square-integrable envelope. The functions class has a square-integrable envelope. Note that

$$\mathrm{E}[\bar{f}_i(t) - f_i(t)]^2 = \mathrm{E}\begin{pmatrix} (q'\bar{\Sigma} - q'\Sigma) z_i \theta_{0i}(\alpha) 1(q'\bar{\Sigma}' z_i < 0) 1\left(q'\Sigma z_i < 0\right) \\ +(q'\bar{\Sigma} - q'\Sigma) z_i \theta_{1i}(\alpha) 1\left(q'\bar{\Sigma} z_i > 0\right) 1\left(q'\Sigma z_i > 0\right) \\ +\left(q'\bar{\Sigma} z_i \theta_{0i}(\alpha) - q'\Sigma z_i \theta_{i1}(\alpha)\right) 1\left(q'\bar{\Sigma} z_i < 0 < q'\Sigma z_i\right) \\ +\left(q'\bar{\Sigma} z_i \theta_{1i}(\alpha) - q'\Sigma z_i \theta_{0i}(\alpha)\right) 1\left(q'\bar{\Sigma} z_i > 0 > q'\Sigma z_i\right) \end{pmatrix}^2$$

$$\begin{aligned}
&\lesssim \big\|\|\bar{\Sigma} - \Sigma\|^2\big\|_{\mathrm{P},2} \cdot \big\|\|z_i\|^2\big\|_{\mathrm{P},2} \max_{\alpha \in \mathcal{A}, \ell \in \{0,1\}} \|\theta^2_{\ell i}(\alpha)\|_{\mathrm{P},2} \\
&\quad + \left(\|\bar{\Sigma}\|^2_{\mathrm{P}} \vee \|\Sigma\|^2\right) \cdot \|\|z_i\|^2\|_{\mathrm{P},2} \max_{\alpha \in \mathcal{A}, \ell \in \{0,1\}} \|\theta^2_{\ell i}(\alpha)\|_{\mathrm{P},2} \cdot \sup_{q \in \mathcal{S}^{d-1}} \mathrm{P}\left[\left|q'\Sigma z_i\right| < \left|q'\left(\bar{\Sigma} - \Sigma\right) z\right|\right]^{1/2} \\
&\lesssim {}_{\mathrm{P}} \|\bar{\Sigma} - \Sigma\|^2 + \sup_{q \in \mathcal{S}^{d-1}} \mathrm{P}\left[\left|q'\Sigma z_i/\|z_i\|\right| < \|\bar{\Sigma} - \Sigma\|\right]^{1/2} \to 0,
\end{aligned}$$

where we invoked the moment and smoothness assumptions. □

Lemma 3 (A Uniform Derivative)**.** *Let $\sigma_{i,\mu}(\alpha) = \mu z_i\left(\theta_{0i}(\alpha) 1(\mu z_i < 0) + \theta_{1i}(\alpha) 1(\mu z_i > 0)\right)$. Uniformly in $\mu \in M = \{q'\Sigma : q \in \mathcal{S}^{d-1}, \|\Sigma\| \le C\}$ and $\alpha \in \mathcal{A}$*

$$\frac{\partial \mathrm{E}[\sigma_{i,\mu}(\alpha)]}{\partial \mu} = \mathrm{E}[z_i w_{i,\mu}(\alpha)],$$

where the right hand side is uniformly continuous in μ and α.

Proof: The continuity of the mapping $(\mu, \alpha) \mapsto \mathrm{E}[z_i w_{i,\mu}(u)]$ follows by an application of the dominated convergence theorem and stated assumptions on the envelopes.

Note that for any $\|\delta\| \to 0$

$$\frac{\mathrm{E}[(\mu+\delta) z_i w_{i,\mu+\delta}(\alpha)] - \mathrm{E}[\mu z_i w_{i,\mu}(\alpha)]}{\|\delta\|} = \frac{\delta}{\|\delta\|}\mathrm{E}[z_i w_{i,\mu}(\alpha)] + \frac{1}{\|\delta\|}\mathrm{E}[R_i(\delta,\mu,\alpha)],$$

where

$$\begin{aligned} R_i(\delta,\mu,\alpha) := \quad & (\mu+\delta)\, z_i\, (\theta_{1i}(\alpha) - \theta_{0i}(\alpha))\, 1\, (\mu z_i < 0 < (\mu+\delta)\, z_i) \\ + \quad & (\mu+\delta)\, z_i\, (\theta_{0i}(\alpha) - \theta_{1i}(\alpha))\, 1\, (\mu z_i > 0 > (\mu+\delta)\, z_i)\,. \end{aligned}$$

By Cauchy-Schwarz and the maintained assumptions

$$\begin{aligned} \sup_{\mu\in M, \alpha\in\mathcal{A}} \mathrm{E}|R_i(\delta,\mu,\alpha)| \quad & \lesssim \quad \|\delta z\|_{\mathrm{P},2} \cdot \sup_{\alpha\in\mathcal{A}, \ell\in\{0,1\}} \|\theta_{\ell i}(\alpha)\|_{\mathrm{P},2} \sup_{\mu\in M,\alpha\in\mathcal{A}} [\mathrm{P}\,(|\mu z| < |\delta z|)]^{1/2} \\ & \lesssim \quad \|\delta z\|_{\mathrm{P},2} \cdot 1 \cdot \delta^{m/2}. \end{aligned}$$

Therefore, as $\|\delta\| \to 0$

$$\sup_{\mu\in M,\alpha\in\mathcal{A}} \frac{1}{\|\delta\|}|\mathrm{E}[R_i(\delta,\mu,u)]| \leq \sup_{\mu\in M,\alpha\in\mathcal{A}} \frac{1}{\|\delta\|}\mathrm{E}|R_i(\delta,\mu,\alpha)| \lesssim \delta^{m/2} \to 0. \quad \square$$

Lemma 4 (Coupling Lemma)**.** *1. (Sample) We have that*

$$\mathbb{G}_n[h(t)] = \mathbb{G}[h(t)] + o_{\mathrm{P}}(1) \text{ in } \ell^\infty(T),$$

where $\mathbb{G}$ is a P-Brownian bridge with covariance function $\mathrm{E}[h(t)h(t')] - \mathrm{E}[h(t)]\mathrm{E}[h(t')]$.

2. (Bootstrap). We have that

$$\mathbb{G}_n[e^o h^o(t)] = \widetilde{\mathbb{G}[h(t)]} + o_{\mathrm{P}}(1) \text{ in } \ell^\infty(T),$$

where $\widetilde{\mathbb{G}}$ is a P-Brownian bridge with covariance function $\mathrm{E}[h(t)h(t')] - \mathrm{E}[h(t)]\mathrm{E}[h(t')]$.

Proof. The proof can be accomplished by using a single common notation. Specifically it will suffice to show that for either the case $g_i = 1$ or $g_i = e_i - 1$

$$\mathbb{G}_n[gh^o] = \mathbb{G}^g[h(t)] + o_{\mathrm{P}}(1) \text{ in } \ell^\infty(T),$$

where $\mathbb{G}$ is a P-Brownian bridge with covariance function $\mathrm{E}[h(t)h(t')] - \mathrm{E}[h(t)]\mathrm{E}[h(t')]$. The process $\mathbb{G}^g$ for the case of $g_i = 1$ is different (in fact independent) of the process $\mathbb{G}^g$ for the case of $g_i = e_i - 1$, but they both have identical distributions. Once we understand this, we can drop the index g for the process.

Within this proof, it will be convenient to define:

$$S_n(t) := \mathbb{G}_n[gh^o(t)] \;\text{ and }\; Z_n(t) := \mathbb{G}[h(t)].$$

Let $B_{jk}, j = 1, ..., p$ be a partition of T into sets of diameter at most j^{-1}. We need at most

$$p \lesssim j^d, \quad d = \dim(T)$$

such partition sets. Choose t_{jk} as arbitrary points in B_{jk}, for all $j = 1, ..., p$. We define the sequence of projections $\pi_j : T \to T$, $j = 0, 1, 2, \ldots, \infty$ by $\pi_j(t) = t_{jk}$ if $t \in B_{jk}$.

In what follows, given a process Z in $\ell^\infty(T)$ and its projection $Z \circ \pi_j$, whose paths are constant over the partition set, we shall identify the process $Z \circ \pi_j$ with a random vector $Z \circ \pi_j$ in $\mathbb{R}^p$, when convenient. Analogously, given a random vector Z in $\mathbb{R}^p$ we identify it with a process Z in $\ell^\infty(T)$, whose paths are constant over the elements of the partition sets.

The result follows from the following relations proven below:

1. Finite-Dimensional Approximation. As $j/\log n \to \infty$, then $\Delta_1 = \sup_{t \in T} \|S_n(t) - S_n \circ \pi_j(t)\| \to_{\mathrm{P}} 0$.

2. Coupling with a Normal Vector. There exists $\mathcal{N}_{nj} =_d N(0, \mathrm{var}[S_n \circ \pi_j])$ such that, if $p^5 \xi_k^2 / n \to 0$, then $\Delta_2 = \sup_j |\mathcal{N}_{nj} - S_n \circ \pi_j| \to_{\mathrm{P}} 0$.

3. Embedding a Normal Vector into a Gaussian Process. There exists a Gaussian process Z_n with the properties stated in the lemma such that $\mathcal{N}_{nj} = Z_n \circ \pi_j$ almost surely.

4. Infinite-Dimensional Approximation. if $j \to \infty$, then $\Delta_3 = \sup_{t \in T} |Z_n(t) - Z_n \circ \pi_j(t)| \to_{\mathrm{P}} 0$.

We can select the sequence $j = \log^2 n$ such that the conditions on j stated in relations (1)-(4) hold. We then conclude using the triangle inequality that

$$\sup_{t \in T} |S_n(t) - Z_n(t)| \leq \Delta_1 + \Delta_2 + \Delta_3 \to_{\mathrm{P}} 0.$$

Relation 1 follows from

$$\Delta_1 = \sup_{t \in T} |S_n(t) - S_n \circ \pi_j(t)| \leq \sup_{\|t - t'\| \leq j^{-1}} |S_n(t) - S_n(t')| \to_{\mathrm{P}} 0,$$

where the last inequality holds by Lemma 5.

Relation 2 follows from the use of Yurinskii's coupling (Pollard (2002, page 244)): Let $\zeta_1, \ldots, \zeta_n$ be independent p-vectors with $\mathrm{E}\zeta_i = 0$ for each i, and $\kappa := \sum_i \mathrm{E}\left[\|\zeta_i\|^3\right]$ finite. Let $S = \zeta_1 + \cdot + \zeta_n$. For each $\delta > 0$ there exists a random vector T with a $N(0, \mathrm{var}(S))$ distribution such that

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

for some universal constant C_0.

In order to apply the coupling, we collapse $S_n \circ \pi_j$ to a p-vector, and we let

$$\zeta_i = \zeta_{1i} + ... + \zeta_{4i} \in \mathbb{R}^p, \quad \zeta_{li} = g_i h_{li}^o \circ \pi \in \mathbb{R}^p,$$

where $h_{li}, l = 1, ..., 4$ are defined in (B.8), so that $S_n \circ \pi_j = \sum_{i=1}^n \zeta_i / \sqrt{n}$. Now note that since $\mathrm{E}[\|\zeta_i\|^3] \lesssim \max_{1 \leq l \leq 4} \mathrm{E}[\|\zeta_{li}\|^3]$ and

$$\begin{aligned} \mathrm{E}\|\zeta_{li}\|^3 &= p^{3/2} \mathrm{E}\left(\frac{1}{p}\sum_{k=1}^p |g_i h_{li}^o(t_{kj})|^2\right)^{3/2} \leq p^{3/2} \mathrm{E}\left(\frac{1}{p}\sum_{k=1}^p |g_i h_{li}^o(t_{kj})|^3\right) \\ &\leq p^{3/2} \sup_{t \in T} \mathrm{E}|h_{li}^o(t_{kj})|^3 \mathrm{E}|g_i|^3, \end{aligned}$$

where we use the independence of g_i, we have that

$$\mathrm{E}[\|\zeta_i\|^3] \lesssim p^{3/2} \max_{1\le l\le 4} \sup_{t\in T} \mathrm{E}|h^o_{li}(t)|^3 \mathrm{E}|g_i|^3.$$

Next we bound the right side of the display above for each l. First, for $A(t) := q'\Sigma\mathrm{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}] J_1^{-1}(\alpha)$

$$\begin{aligned}
\sup_{t\in T} \mathrm{E}|h^o_{1i}(t))|^3 &= \sup_{t\in T} \mathrm{E}\left|A(t) p_i \varphi_{i1}(\alpha)\right|^3 \le \sup_{t\in T} \|A(t)\|^3 \cdot \sup_{\|\delta\|=1} \mathrm{E}|\delta' p_i|^3 \sup_{\alpha\in\mathcal{A}, x\in X} \mathrm{E}[|\varphi_i(\alpha)|^3 | x_i = x] \\
&\lesssim \sup_{\|\delta\|=1} \mathrm{E}|\delta' p_i|^3 \sup_{\alpha\in\mathcal{A}, x\in X} \mathrm{E}[|\varphi_i(\alpha)|^3 | x_i = x] \\
&\lesssim \xi_k \sup_{\|\delta\|=1} \mathrm{E}|\delta' p_i|^2 \sup_{\alpha\in\mathcal{A}, x\in X} \mathrm{E}[|\varphi_i(\alpha)|^3 | x_i = x] \lesssim \xi_k,
\end{aligned}$$

where we used the assumption that $\sup_{\alpha\in\mathcal{A},x\in X} \mathrm{E}[|\varphi_i(\alpha)|^3|x_i = x] \lesssim 1$, $\|\mathrm{E}p_i p_i\| \lesssim 1$, and that

$$\sup_{t\in T} \|A(t)\| \le \sup_{\|\delta\|=1} [\mathrm{E}[z_i'\delta]^2]^{1/2} \sup_{\|\delta\|=1} [\mathrm{E}[p_i'\delta]^2]^{1/2} \sup_{\alpha\in\mathcal{A}} \|J^{-1}(\alpha)\| \lesssim 1,$$

where the last bound is true by assumption. Similarly $\mathrm{E}|h^o_{2i}(t))|^3 \lesssim \xi_k$. Next

$$\begin{aligned}
\sup_{t\in T} \mathrm{E}|h^o_{3i}(t)|^3 &= \sup_{t\in T} \mathrm{E}|q'\Sigma \left(x_i z_i'\right)^o \Sigma\mathrm{E}[z_i, w_{i,q'\Sigma}(\alpha)]|^3 \\
&\lesssim \mathrm{E}\| \left(x_i z_i'\right)^o \|^3 \sup_{t\in T} \| \left[\mathrm{E}[z_i w_{i,q'\Sigma}(\alpha)]\right] \|^3 \\
&\lesssim \left(\mathrm{E}\left\|\left(x_i z_i'\right)\right\|^3 + \left\|\mathrm{E}\left(x_i z_i'\right)\right\|^3\right) \left(\mathrm{E}\|z_i\|^2\right)^{3/2} \sup_{\alpha\in\mathcal{A}} \mathrm{E}\left[\left|\theta_{li}\left(\alpha\right)\right|^2\right]^{3/2} \\
&\lesssim 1, \qquad\qquad (\mathrm{B}.10)
\end{aligned}$$

where the last bound follows from assumptions C3. Finally,

$$\begin{aligned}
\sup_{t\in T} \left[\mathrm{E}|h^o_{4i}(t)|^3\right]^{1/3} &= \sup_{t\in T} \left[\mathrm{E}|q'\Sigma z_i w_{i,q'\Sigma}(\alpha)|^3\right]^{1/3} + \sup_{t\in T} |\mathrm{E}q'\Sigma z_i w_{i,q'\Sigma}(\alpha)| \\
&\le 2\sup_{t\in T} [\mathrm{E}|q'\Sigma z_i|^6]^{1/6} \left[\mathrm{E}|w_{i,q'\Sigma}(\alpha)|^6\right]^{1/6} \\
&\lesssim [\mathrm{E}|z_i|^6]^{1/6} \sup_{\alpha\in\mathcal{A}} \mathrm{E}[|\theta_{li}(\alpha)|^6]^{1/6} \lesssim 1,
\end{aligned}$$

where the last line follows from assumption C3.

Therefore, by Yurinskii's coupling, observing that in our case by the above arguments $\kappa = \frac{p^{3/2}\xi_k n}{(\sqrt{n})^3}$, for each $\delta > 0$ if $p^5 \xi_k^2/n \to 0$,

$$\mathrm{P}\left\{\left\|\frac{\sum_{i=1}^n \zeta_i}{\sqrt{n}} - \mathcal{N}_{n,j}\right\| \ge 3\delta\right\} \lesssim \frac{np p^{3/2}\xi_k}{(\delta\sqrt{n})^3} = \frac{p^{5/2}\xi_k}{(\delta^3 n^{1/2})} \to 0.$$

This verifies relation (2).

Relation (3) follows from the a.s. embedding of a finite-dimensional random normal vector into a path of a Gaussian process whose paths are continuous with respect to the standard metric ρ_2, defined in Lemma 6, which is shown e.g., in Belloni and Chernozhukov (2009). Moreover, since ρ_2 is continuous with respect to the Euclidian metric on T, as shown in part 2 of Lemma 6, the paths of the process are continuous with respect to the Euclidian metric as well.

Relation (4) follows from the inequality

$$\Delta_3 = \sup_{t\in T} |Z_n(t) - Z_n \circ \pi_j(t)| \leq \sup_{\|t-t'\|\leq j^{-1}} |Z_n(t) - Z_n(t')| \lesssim_{\mathrm{P}} (1/j)^c \log(1/j)^c \to 0,$$

where $0 < c \leq 1/2$ is defined in Lemma 6. This inequality follows from the entropy inequality for Gaussian processes (Corollary 2.2.8 of van der Vaart and Wellner (1996))

$$\mathrm{E} \sup_{\rho_2(t,t')\leq\delta} \left|Z_n(t) - Z_n(t'_0)\right| \leq \int_0^\delta \sqrt{\log N(\epsilon, T, \rho_2)}d\epsilon$$

and parts 2 and 3 of Lemma 6. From part 2 of Lemma 6 we first conclude that

$$\log N(\epsilon, T, \rho_2) \lesssim \log(1/\epsilon),$$

and second that $\|t - t'\| \leq (1/j)$ implies $\rho_2(t, t') \leq (1/j)^c$, so that

$$\mathrm{E} \sup_{\|t-t'\|\leq 1/j} |Z_n(t) - Z_n(t')| \leq (1/j)^c \log(1/j)^c \text{ as } j \to \infty.$$

The claimed inequality then follows by Markov inequality. □

Lemma 5 (Bounded Oscillations)**.** *1. (Sample) For $\epsilon_n = o((\log n)^{-1/(2c)})$, we have that*

$$\sup_{\|t-t'\|\leq\epsilon_n} |\mathbb{G}_n[h(t) - h(t')]| \to_{\mathrm{P}} 0.$$

2. (Bootstrap). For $\epsilon_n = o((\log n)^{-1/(2c)})$, we have that

$$\sup_{\|t-t'\|\leq\epsilon_n} |\mathbb{G}_n[(e_i - 1)(h^o(t) - h^o(t'))]| \to_{\mathrm{P}} 0.$$

Proof. To show both statements, it will suffice to show that for either the case $g_i = 1$ or $g_i = e_i - 1$, we have that

$$\sup_{\|t-t'\|\leq\epsilon_n} |\mathbb{G}_n[g_i(h^o(t) - h^o(t'))]| \to_{\mathrm{P}} 0.$$

Step 1. Since

$$\sup_{\|t-t'\|\leq\epsilon_n} |\mathbb{G}_n[g_i(h^o(t) - h^o(t'))]| \lesssim \max_{1\leq\ell\leq 4} \sup_{\|t-t'\|\leq\epsilon_n} |\mathbb{G}_n[g_i(h^o_\ell(t) - h^o_\ell(t'))]|,$$

we bound the latter for each ℓ. Using the results in Lemma 9 that bound the random entropy of $\mathcal{H}_1$ and $\mathcal{H}_2$ and the results in Lemma 11 we have that for $\ell = 1$ and 2

$$\Delta_{n\ell} = \sup_{\|t-t'\|\leq\epsilon_n} |\mathbb{G}_n[g_i(h^o_\ell(t) - h^o_\ell(t'))]| \lesssim_{\mathrm{P}} \sqrt{\log n} \sup_{\|t-t'\|\leq\epsilon_n} \max_{\mathbb{P}\in\{\mathrm{P},\mathbb{P}_n\}} \|g_i(h^o_\ell(t) - h^o_\ell(t'))\|_{\mathbb{P},2}.$$

By Lemma 9 that bounds the entropy of $g_i(\mathcal{H}^o_\ell - \mathcal{H}^o_\ell)^2$ and Lemma 11 we have that for $\ell = 1$ or $\ell = 2$,

$$\sup_{\|t-t'\|\leq\epsilon_n} \left|\|g(h^o_\ell(t)-h^o_\ell(t'))\|_{\mathbb{P}_n,2}-\|g(h^o_\ell(t)-h^o_\ell(t'))\|_{\mathrm{P},2}\right| \lesssim_{\mathrm{P}} \sqrt{\frac{\log n}{n}} \sup_{t\in T} \max_{\mathbb{P}\in\{\mathrm{P},\mathbb{P}_n\}} \|g^2 h^{o2}_\ell(t)\|_{\mathbb{P},2}.$$

By Step 2 below we have

$$\sup_{t\in T}\max_{\mathbb{P}\in\{\mathrm{P},\mathbb{P}_n\}}\|g^2h_\ell^{o2}(t)\|_{\mathbb{P},2}\lesssim_{\mathrm{P}}\sqrt{\xi_k^2\max_{i\le n}F_1^4\max_i|g_i|^4}\lesssim_{\mathrm{P}}\sqrt{\xi_k^2\max_{i\le n}F_1^4(\log n)^4},$$

and by Lemma 6, $\|g(h_\ell(t)-h_\ell(t'))\|_{\mathrm{P},2}\lesssim\|t-t'\|^c$. Putting the terms together we conclude

$$\Delta_{n\ell}\lesssim_{\mathrm{P}}\sqrt{\log n}\left(\epsilon_n^c+\sqrt{\frac{\log n}{n}}\xi_k\sqrt{\max_{i\le n}F_1^4}(\log n)^2\right)\to_{\mathrm{P}}0,$$

by assumption and the choice of ϵ_n.

For $\ell=3$ and $\ell=4$, by Lemma 9, $g(\mathcal{H}_3^o-\mathcal{H}_3^o)$ and $g(\mathcal{H}_4^o-\mathcal{H}_4^o)$ are P-Donsker, so that

$$\Delta_{n\ell}\le\sup_{\rho_2(t,t'_n)\le\epsilon_n^c}|\mathbb{G}_n[g(h_\ell^o(t)-h_\ell^o(t'))]|\to_{\mathrm{P}}0.\quad\square$$

Step 2. Since $\|g^2h_\ell^{o2}(t)\|_{\mathbb{P},2}\le 2\|g^2h_\ell^2(t)\|_{\mathbb{P},2}+2\|g^2E[h_\ell^o(t)]^2\|_{\mathbb{P},2}$, for $\mathbb{P}\in\{\mathrm{P},\mathbb{P}_n\}$, it suffices to bound each term separately.

Uniformly in $t\in T$ for $\ell=1,2$

$$\begin{aligned}\mathbb{E}_n[gh_\ell(t)]^4&\le\max_i g_i^4\cdot\|\Sigma\|^4\|\mathbb{E}_n[z_ip_i1\{q'\Sigma z_i<0\}]\|\|J_0^{-1}(\alpha)\|\cdot\sup_{\|\delta\|=1}\mathbb{E}_n[[\delta'p_i]^4\varphi_{i0}^4(\alpha)]\\&\lesssim_{\mathrm{P}}(\log n)^4\cdot 1\cdot\xi_k^2\|\mathbb{E}_n[p_ip_i']\|\max_{i\le n}F_1^4\lesssim_{\mathrm{P}}\xi_k^2\max_{i\le n}F_1^4(\log n)^4,\end{aligned}$$

where we used assumptions C3 and C5 and the fact that $\|\mathbb{E}_n[z_ip_i1\{q'\Sigma z_i<0\}]\|\lesssim_{\mathrm{P}}1$ and $\mathbb{E}_n[p_ip_i']\lesssim_{\mathrm{P}}1$ as shown in the proof of Lemma 1.

Uniformly in $t\in T$ for $\ell=1,2$

$$\begin{aligned}\mathrm{E}[gh_\ell(t)]^4&\le\mathrm{E}[g^4]\|\Sigma\|^4\|\mathrm{E}[z_ip_i1\{q'\Sigma z_i<0\}]\|\|J_0^{-1}(\alpha)\|\cdot\sup_{\|\delta\|=1}\mathrm{E}[[\delta'p_i]^4\varphi_{i0}^4(\alpha)]\\&\lesssim_{\mathrm{P}}1\cdot\xi_k^2\|\mathrm{E}[p_ip_i']\|\sup_{x\in X,\alpha\in\mathcal{A}}\mathrm{E}[\varphi_{i0}^4(\alpha)|x_i=x]\lesssim\xi_k^2,\end{aligned}$$

where we used assumption C3.

Uniformly in $t\in T$ for $\ell=1,2$

$$\mathbb{E}_n[g^4\mathrm{E}[h_\ell^o(t)]^4]\le\mathbb{E}_ng^4\mathrm{E}[h_\ell^o(t)]^4\lesssim_P 1\cdot\mathrm{E}[h_\ell^{o2}(t)]^2\lesssim 1,$$

and

$$\mathrm{E}[g^4\mathrm{E}[h_\ell^o(t)]^4]\le\mathrm{E}g^4\mathrm{E}[h_\ell^o(t)]^4\lesssim 1\cdot\mathrm{E}[h_\ell^{o2}(t)]^2\lesssim 1.$$

where the bound in $\mathrm{E}[h_\ell^{o2}(t)]^2$ follows from calculations given in the proof of Lemma 6. $\square$

Lemma 6 (Covariance Properties)**.** 1. *For some* $0<c\le 1/2$

$$\rho_2(h(t),h(t'))=\left(\mathrm{E}[h(t)-h(t')]^2\right)^{1/2}\lesssim\rho(t,t'):=\left\|t-t'\right\|^c$$

2. *The covariance function* $\mathrm{E}[h(t)h(t')]-\mathrm{E}[h(t)]\mathrm{E}[h(t')]$ *is equi-continuous on* $T\times T$ *uniformly in* k.

3. *A sufficient condition for the variance function to be bounded away from zero,* $\inf_{t\in T} \text{var}(h(t)) \geq L > 0$, *uniformly in* k *is that the following matrices have minimal eigenvalues bounded away from zero uniformly in* k*: var* $\left(\begin{bmatrix}\varphi_{i1}(\alpha) & \varphi_{i0}(\alpha)\end{bmatrix}' |x_i, z_i\right)$ $\text{E}[p_i p_i']$, $J_0^{-1}(\alpha)$, $J_1^{-1}(\alpha)$, $b_0'b_0$, *and* $b_1'b_1$, *where* $b_1 = \text{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}]$ *and* $b_0 = \text{E}[z_i p_i' 1\{q'\Sigma z_i < 0\}]$.

Comment B.2. We emphasize that claim 3 only gives sufficient conditions for $\text{var}\,(h(t))$ to be bounded away from zero. In particular, the assumption that

$$\text{mineig}\left(\text{var}\left(\begin{bmatrix}\varphi_{i1}(\alpha) & \varphi_{i0}(\alpha)\end{bmatrix}' |x_i, z_i\right)\right) \geq L$$

is not necessary, and does not hold in all relevant situations. For example, when the upper and lower bounds have first-order equivalent asymptotics, which can occur in the point-identified and local to point-identified cases, this condition fails. However, the result still follows from equation (B.11) under the assumption that

$$\text{var}\left(\varphi_{i1}(\alpha)|x_i, z_i\right) = \text{var}\left(\varphi_{i0}(\alpha)|x_i, z_i\right) \geq L$$

Proof. **Claim 1.** Observe that $\rho_2\left(h(t), h(\tilde{t})\right) \lesssim \max_j \rho_2\left(h_j(t), h_j(\tilde{t})\right)$. We will bound each of these four terms. For the first term, we have

$$\begin{aligned}
\rho_2(h_1(t), h_1(\tilde{t})) =& \text{E}\left[\left(\begin{array}{c} q'\Sigma \text{E}\left[z_i p_i' 1\left\{q'\Sigma z_i > 0\right\}\right] J_1^{-1}(\alpha)\, p_i \varphi_{i1}(\alpha) - \\ -\tilde{q}'\Sigma \text{E}\left[z_i p_i' 1\left\{\tilde{q}'\Sigma z_i > 0\right\}\right] J_1^{-1}(\tilde{\alpha})\, p_i \varphi_{i1}(\tilde{\alpha}) \end{array}\right)^2\right]^{1/2} \\
\leq& \text{E}\left[\left((q-\tilde{q})'\Sigma \text{E}\left[z_i p_i' 1\left\{q'\Sigma z_i > 0\right\}\right] J_1^{-1}(\alpha)\, p_i \varphi_{i1}(\alpha)\right)^2\right]^{1/2} + \\
&+ \text{E}\left[\left(\tilde{q}'\Sigma\left(\begin{array}{c} \text{E}\left[z_i p_i' 1\left\{q'\Sigma z_i > 0\right\}\right] - \\ -\text{E}\left[z_i p_i' 1\left\{\tilde{q}'\Sigma z_i > 0\right\}\right] \end{array}\right) J_1^{-1}(\alpha)\, p_i \varphi_{i1}(\alpha)\right)^2\right]^{1/2} + \\
&+ \text{E}\left[\left(\tilde{q}'\Sigma \text{E}\left[z_i p_i' 1\left\{\tilde{q}'\Sigma z_i > 0\right\}\right]\left(J_1^{-1}(\alpha) - J_1^{-1}(\tilde{\alpha})\right) p_i \varphi_{i1}(\alpha)\right)^2\right]^{1/2} + \\
&+ \text{E}\left[\left(\tilde{q}'\Sigma \text{E}\left[z_i p_i' 1\left\{\tilde{q}'\Sigma z_i > 0\right\}\right] J_1^{-1}(\tilde{\alpha})\, p_i\left(\varphi_{i1}(\alpha) - \varphi_{i1}(\tilde{\alpha})\right)\right)^2\right]^{1/2}
\end{aligned}$$

For the first term we have

$$\begin{aligned}
&\text{E}\left[\left((q-\tilde{q})'\Sigma \text{E}\left[z_i p_i' 1\left\{q'\Sigma z_i > 0\right\}\right] J_1^{-1}(\alpha)\, p_i \varphi_{i1}(\alpha)\right)^2\right]^{1/2} \leq \\
&\leq \|q-\tilde{q}\|\, \|\Sigma\|\, \left\|\text{E}\left[z_i p_i'\right]\right\|\, \left\|J_1^{-1}(\alpha)\right\| \text{E}[\left\|p_i p_i'\right\|^2]^{1/2} \sup_{x_i, z_i} \text{E}[\varphi_{i\ell}(\alpha)^4 | x_i, z_i]^{1/2}
\end{aligned}$$

By assumption C3, $\|\text{E}\left[z_i p_i'\right]\|$, $\left\|J_1^{-1}(\alpha)\right\|$, $\text{E}[\|p_i p_i'\|^2]$, and $\sup_{x_i, z_i} \text{E}[\varphi_{i1}(\alpha)^4 | x_i, z_i]$ are bounded uniformly in k and α. Therefore,

$$\text{E}\left[\left((q-\tilde{q})'\Sigma \text{E}\left[z_i p_i' 1\left\{q'\Sigma z_i > 0\right\}\right] J_1^{-1}(\alpha)\, p_i \varphi_{i1}(\alpha)\right)^2\right]^{1/2} \lesssim \|q-\tilde{q}\|$$

The same conditions give the following bound on the second term.

$$\begin{aligned}
&\mathrm{E}\left[\left(\tilde{q}'\Sigma\left(\begin{array}{c}\mathrm{E}\left[z_i p_i' 1\left\{q'\Sigma z_i > 0\right\}\right] - \\ -\mathrm{E}\left[z_i p_i' 1\left\{\tilde{q}'\Sigma z_i > 0\right\}\right]\end{array}\right) J_1^{-1}(\alpha)\, p_i \varphi_{i1}(\alpha)\right)^2\right]^{1/2} \\
&\lesssim \left\|\mathrm{E}\left[1\left\{q'\Sigma z_i > 0\right\} - 1\left\{\tilde{q}'\Sigma z_i > 0\right\}\right]\right\| \\
&\lesssim \mathrm{E}\left[\left(1\left\{q'\Sigma z_i > 0\right\} - 1\left\{\tilde{q}'\Sigma z_i > 0\right\}\right)^2\right]^{1/2}
\end{aligned}$$

As in step 5 of the proof of Lemma 1, the assumption that $\mathrm{P}\left(\left|q'\Sigma z_i/\|z_i\|\right| < \delta\right)/\delta^m \lesssim 1$ implies

$$\mathrm{E}\left[\left(1\left\{q'\Sigma z_i > 0\right\} - 1\left\{\tilde{q}'\Sigma z_i > 0\right\}\right)^2\right]^{1/2} \lesssim \|q - \tilde{q}\|^{m/2}$$

Similarly, the third term is bounded as follows:

$$\mathrm{E}\left[\left(\tilde{q}'\Sigma\mathrm{E}\left[z_i p_i' 1\left\{\tilde{q}'\Sigma z_i > 0\right\}\right]\left(J_1^{-1}(\alpha) - J_1^{-1}(\tilde{\alpha})\right) p_i \varphi_{i1}(\alpha)\right)^2\right]^{1/2} \lesssim \left\|J_1^{-1}(\alpha) - J_1^{-1}(\tilde{\alpha})\right\|.$$

Note that $J_1^{-1}(\alpha)$ is uniformly Lipschitz in $\alpha \in \mathcal{A}$ by assumption C3, so $\left\|J_1^{-1}(\alpha) - J_1^{-1}(\tilde{\alpha})\right\| \lesssim \|\alpha - \tilde{\alpha}\|$. Finally, the fourth term is bounded by

$$\begin{aligned}
&\mathrm{E}\left[\left(\tilde{q}'\Sigma\mathrm{E}\left[z_i p_i' 1\left\{\tilde{q}'\Sigma z_i > 0\right\}\right] J_1^{-1}(\tilde{\alpha})\, p_i\left(\varphi_{i1}(\alpha) - \varphi_{i1}(\tilde{\alpha})\right)\right)^2\right]^{1/2} \lesssim \\
&\lesssim \sup_{x_i, z_i} \mathrm{E}\left[\left(\varphi_{i1}(\alpha) - \varphi_{i1}(\tilde{\alpha})\right)^4 | x_i, z_i\right]^{1/2} \\
&\lesssim \|\alpha - \tilde{\alpha}\|^{\gamma_\varphi}
\end{aligned}$$

where we used the assumption that $\mathrm{E}\left[\left(\varphi_{i1}(\alpha) - \varphi_{i1}(\tilde{\alpha})\right)^4 | x_i, z_i\right]^{1/2}$ is uniformly γ_φ-Hölder continuous in α. Combining, we have

$$\begin{aligned}
\rho_2(h_1(t), h_1(\tilde{t})) &\lesssim \|q - \tilde{q}\| + \|q - \tilde{q}\|^{m/2} + \|\alpha - \tilde{\alpha}\| + \|\alpha - \tilde{\alpha}\|^{\gamma_\varphi} \\
&\lesssim \left\|t - t'\right\|^{1 \wedge m/2 \wedge \gamma_\varphi}
\end{aligned}$$

An identical argument shows that $\rho_2\left(h_2(t), h_2(t')\right) \lesssim \|t - t'\|^{1 \wedge m/2 \wedge \gamma_\varphi}$.

The third and fourth components of $h(t)$ can be bounded using similar arguments. For $h_3(t)$, we have

$$\begin{aligned}
\rho_2\left(h_3(t), h_3(\tilde{t})\right) &= \mathrm{E}\left[\left(q'\Sigma x_i z_i'\Sigma \mathrm{E}\left[z_i w_{i,q'\Sigma}(\alpha)\right] - \tilde{q}'\Sigma x_i z_i'\Sigma \mathrm{E}\left[z_i w_{i,\tilde{q}'\Sigma}(\tilde{\alpha})\right]\right)^2\right]^{1/2} \\
&\lesssim \mathrm{E}\left[\left((q-\tilde{q})'\Sigma x_i z_i'\Sigma \mathrm{E}\left[z_i w_{i,q'\Sigma}(\alpha)\right]\right)^2\right]^{1/2} + \\
&\quad + \mathrm{E}\left[\left(\tilde{q}'\Sigma x_i z_i'\Sigma\left(\mathrm{E}\left[z_i w_{i,q'\Sigma}(\alpha)\right] - \mathrm{E}\left[z_i w_{i,\tilde{q}'\Sigma}(\alpha)\right]\right)\right)^2\right]^{1/2} + \\
&\quad + \mathrm{E}\left[\left(\tilde{q}'\Sigma x_i z_i'\Sigma\left(\mathrm{E}\left[z_i w_{i,\tilde{q}'\Sigma}(\alpha)\right] - \mathrm{E}\left[z_i w_{i,\tilde{q}'\Sigma}(\tilde{\alpha})\right]\right)\right)^2\right]^{1/2} \\
&\lesssim \|q-\tilde{q}\| \|\Sigma\| \mathrm{E}[z_i' z_i \max_{\ell\in\{0,1\}} \theta_\ell(x_i,\alpha)^2]^{1/2} + \\
&\quad + \|\Sigma\| \mathrm{E}\left[z_i' z_i \left(\theta_1(x_i,\alpha) - \theta_0(x_i,\alpha)\right)^2 1\{\left|(q-\tilde{q})'\Sigma z_i\right| \geq |q'\Sigma z_i|\}\right]^{1/2} + \\
&\quad + \|\Sigma\| \mathrm{E}\left[z_i' z_i \max_{\ell\in\{0,1\}} \left(\theta_\ell\left(x_i,\alpha\right) - \theta_\ell\left(x_i,\tilde{\alpha}\right)\right)^2\right]^{1/2}
\end{aligned}$$

By assumption, $\mathrm{E}\left[z_i' z_i \theta_\ell(x_i,\alpha)^2\right] \leq \left(\mathrm{E}\left[\|z_i\|^4\right] \mathrm{E}\left[\theta_\ell(x_i,\alpha)^4\right]\right)^{1/2}$ is bounded uniformly in α. Also,

$$\begin{aligned}
&\mathrm{E}\left[z_i' z_i \left(\theta_1(x_i,\alpha) - \theta_0(x_i,\alpha)\right)^2 1\{|q'\Sigma z_i| / \|z_i\| \leq \|q-\tilde{q}\|\}\right] \lesssim \\
&\quad \lesssim \mathrm{E}\left[\|z_i\|^4\right]^{1/2} \left(\mathrm{E}\left[\theta_1(x_i,\alpha)^4\right]^{1/2} + \mathrm{E}\left[\theta_0(x_i,\alpha)^4\right]^{1/2}\right) \mathrm{E}\left[1\left\{|q'\Sigma z_i| / \|z_i\| \leq \|q-\tilde{q}\|\right\}\right]^{1/2} \\
&\quad \lesssim \|q-\tilde{q}\|^{m/2}
\end{aligned}$$

where we have used the smoothness condition (C1) and the fact that $\mathrm{E}[\|z_i\|^4] < \infty$ and $\mathrm{E}[\theta_\ell(x_i,\alpha)^4] < \infty$ uniformly in α.

By assumption, $\theta_\ell(x,\alpha)$ are Hölder continuous in α with coefficient $L(x)$, so

$$\begin{aligned}
\mathrm{E}\left[z_i' z_i \max_{\ell\in\{0,1\}} \left(\theta_\ell\left(x_i,\alpha\right) - \theta_\ell\left(x_i,\tilde{\alpha}\right)\right)^2\right]^{1/2} &\lesssim \mathrm{E}\left[\|z_i\|^4\right]^{1/2} \mathrm{E}\left[L(x_i)^4\right]^{1/2} \|\alpha-\tilde{\alpha}\|^{\gamma_\theta} \\
&\lesssim \|\alpha-\tilde{\alpha}\|^{\gamma_\theta}
\end{aligned}$$

Thus,

$$\begin{aligned}
\rho_2(h_3(t), h_3(\tilde{t})) &\lesssim \|q-\tilde{q}\| + \|q-\tilde{q}\|^{m/2} + \|\alpha-\tilde{\alpha}\|^{\gamma_\theta} \\
&\lesssim \left\|t-t'\right\|^{1\wedge m/2\wedge\gamma_\theta}
\end{aligned}$$

For h_4, we have

$$
\begin{aligned}
\rho_2\left(h_4(t), h_4(\tilde{t})\right) = & \mathrm{E}\left[\left(q'\Sigma z_i w_{i,q'\Sigma}(\alpha) - \tilde{q}'\Sigma z_i w_{i,\tilde{q}'\Sigma}(\tilde{\alpha})\right)^2\right]^{1/2} \\
\leq & \mathrm{E}\left[\left((q-\tilde{q})'\,\Sigma z_i w_{i,q'\Sigma}(\alpha)\right)^2\right]^{1/2} + \\
& + \mathrm{E}\left[\left(\tilde{q}'\Sigma z_i\left(w_{i,q'\Sigma}(\alpha) - w_{i,\tilde{q}'\Sigma}(\alpha)\right)\right)^2\right]^{1/2} + \\
& + \mathrm{E}\left[\left(\tilde{q}'\Sigma z_i\left(w_{i,\tilde{q}'\Sigma}(\alpha) - w_{i,\tilde{q}'\Sigma}(\tilde{\alpha})\right)\right)^2\right]^{1/2} \\
\lesssim & \|q-\tilde{q}\| + \|q-\tilde{q}\|^{m/2} + \|\alpha-\tilde{\alpha}\|^{\gamma_\theta}
\end{aligned}
$$

by the exact same arguments used for h_3.

Claim 2. It suffices to show that $\mathrm{E}\left[h_j(t)\right]$ for $j = 1, ..., 4$ and $\mathrm{E}\left[h_j(t)h_k(t')\right]$ for $j = 1, ..., 4$ and $k = 1, ..., 4$ are uniformly equicontinuous. Hölder continuity implies equicontinuity, so we show that each of these functions are uniformly Hölder continuous.

Jensen's inequality and the result in Part 1 show that $\mathrm{E}[h_j(t)]$ are uniformly Hölder.

$$
\left|\mathrm{E}[h_j(t)] - \mathrm{E}[h_j(t')]\right| \leq \mathrm{E}\left[\left(h_j(t) - h_j(t')\right)^2\right]^{1/2} \lesssim \left\|t - t'\right\|^c
$$

Given this, a simple calculation shows that $\mathrm{E}\left[h_j(t_1)h_k(t_2)\right]$ are uniformly Hölder as well.

$$
\begin{aligned}
\left|\mathrm{E}\left[h_j(t_1)h_k(t_2) - h_j(t_1')h_k(t_2')\right]\right| = & \left|\mathrm{E}\begin{bmatrix} \left(h_j(t_1) - h_j(t_1')\right)h_k(t_2) + \\ + h_j(t_1')\left(h_k(t_2) - h_k(t_2')\right)\end{bmatrix}\right| \\
\leq & \mathrm{E}\left[\left(h_j(t_1) - h_j(t_1')\right)^2\right]^{1/2}\mathrm{E}[h_k(t_2)^2]^{1/2} + \\
& + \mathrm{E}[h_j(t_1')^2]^{1/2}\mathrm{E}\left[\left(h_k(t_2) - h_k(t_2')\right)^2\right]^{1/2} \\
\lesssim & \left\|t_1 - t_1'\right\|^c \vee \left\|t_2 - t_2'\right\|^c
\end{aligned}
$$

Claim 3. By the law of total variance,

$$
\mathrm{var}(h(t)) = \mathrm{E}\left[\mathrm{var}\left(h(t)|x_i, z_i\right)\right] + \mathrm{var}\left(\mathrm{E}\left[h(t)|x_i, z_i\right]\right).
$$

Note that $h_3(t)$ and $h_4(t)$ are constant conditional on x_i, z_i, so

$$
\begin{aligned}
\mathrm{var}\left(h(t)|x_i, z_i\right) = & \mathrm{var}\left(h_1(t) + h_2(t)|x_i, z_i\right) \\
= & \begin{bmatrix} q'\Sigma\mathrm{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}]J_1^{-1}(\alpha)p_i \\ q'\Sigma\mathrm{E}[z_i p_i' 1\{q'\Sigma z_i < 0\}]J_0^{-1}(\alpha)p_i \end{bmatrix}' \mathrm{var}\left(\begin{bmatrix}\varphi_{i1}(\alpha) \\ \varphi_{i0}(\alpha)\end{bmatrix}|x_i, z_i\right) \times \\
& \times \begin{bmatrix} q'\Sigma\mathrm{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}]J_1^{-1}(\alpha)p_i \\ q'\Sigma\mathrm{E}[z_i p_i' 1\{q'\Sigma z_i < 0\}]J_0^{-1}(\alpha)p_i \end{bmatrix}
\end{aligned}
\tag{B.11}
$$

Recall that $b_1 = \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}]$ and $b_0 = \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i < 0\}]$. Let $\gamma_\ell = q'\Sigma b_\ell$, and $\mathrm{mineig}(M)$ denote the minimal eigenvalue of any matrix M. By assumption,

$$
\mathrm{mineig}\left(\mathrm{var}\left(\begin{bmatrix}\varphi_{i1}(\alpha) & \varphi_{i0}(\alpha)\end{bmatrix}'|x_i, z_i\right)\right) > L,
$$

so

$$\begin{aligned}
\mathrm{E}\left[\mathrm{var}\left(h(t)|x_i,z_i\right)\right] \gtrsim & \mathrm{E}\left[\left\|\left[\gamma_1 J_1^{-1}(\alpha)p_i \quad \gamma_0 J_0^{-1}(\alpha)p_i\right]\right\|^2\right] \\
\gtrsim & \mathrm{E}\left[\left\|\gamma_1 J_1^{-1}(\alpha)p_i\right\|^2 \vee \left\|\gamma_0 J_0^{-1}(\alpha)p_i\right\|^2\right] \\
\gtrsim & \mathrm{E}\left[\left\|\gamma_1 J_1^{-1}(\alpha)p_i\right\|^2\right] \vee \mathrm{E}\left[\left\|\gamma_0 J_0^{-1}(\alpha)p_i\right\|^2\right]
\end{aligned}$$

Repeated use of the inequality $\|xy\|^2 \geq \mathrm{mineig}(yy')\|x\|^2$ yields for $l = 0, 1$,

$$\begin{aligned}
\mathrm{E}\left[\left\|\gamma_\ell J_\ell^{-1}(\alpha)p_i\right\|^2\right] \geq & \ \mathrm{mineig}\left(\mathrm{E}\left[p_i p_i'\right]\right)\mathrm{mineig}\left(J_\ell^{-1}(\alpha)\right)^2\|\gamma_\ell\|^2 \\
\gtrsim & \ \mathrm{mineig}(b_\ell' b_\ell)\left\|q'\Sigma\right\|^2 \\
\gtrsim & \ b_\ell' b_\ell
\end{aligned}$$

where the last line follows from the fact that $b_\ell' b_\ell$ is a scalar. We now show that $b_\ell' b_\ell > 0$. Let $f_{1i} = z_i 1\{q'\Sigma z_i > 0\}$ and $f_{0i} = z_i 1\{q'\Sigma z_i < 0\}$. Observe that $z_i = f_{1i} + f_{0i}$ and $\mathrm{E}\left[f_{1i}' f_{0i}\right] = 0$, so

$$\mathrm{E}\left[f_{1i}' f_{1i}\right] \vee \mathrm{E}\left[f_{0i}' f_{0i}\right] \geq \frac{1}{2}\mathrm{E}\left[z_i' z_i\right] > 0$$

By the completeness of our series functions, we can represent f_{1i} and f_{0i} in terms of the series functions. Let

$$f_{1i} = \sum_{j=1}^{\infty} c_{1j} p_{ji} f_{0i} = \sum_{j=1}^{\infty} c_{0j} p_{ji}$$

Without loss of generality, assume the series functions are orthonormal. Then

$$\mathrm{E}\left[f_{1i}' f_{1i}\right] = \sum_{j=1}^{\infty} c_{1j}^2 \mathrm{E}\left[f_{0i}' f_{0i}\right] = \sum_{j=1}^{\infty} c_{0j}^2$$

Also,

$$b_\ell' b_\ell = \sum_{j=1}^{k} c_{\ell j}^2$$

Thus,

$$\mathrm{E}\left[\mathrm{var}\left(h(t)|x_i,z_i\right)\right] \gtrsim \mathrm{mineig}(b_1' b_1) \vee \mathrm{mineig}(b_0' b_0) > 0$$

□

B.4. **Conservative Inference with Discrete Covariates.** Let $\Theta(x,\alpha) = \left[\theta_0(x,\alpha), \theta_1(x,\alpha)\right]$, and to simplify notation suppress the dependence of Θ and θ_ℓ on (x,α) and let the instruments coincide with $x = [x_1\ x_2]'$, with $x_1 = 1$ and $x_2 \in \mathbb{R}^{d-1}$. Let $\Sigma = \mathrm{E}(xx')^{-1}$, $z = x + \sigma\,[0\ \eta]'$, with $\eta \sim N(0, I)$ and independent of x and θ_ℓ, $\ell = 0, 1$, where I denotes the identity matrix. Note that $\mathrm{E}(xx') = \mathrm{E}(zx')$, and define

$$B = \Sigma\mathbf{E}\left(x\Theta\right), \qquad \tilde{B} = \Sigma\mathbf{E}\left(z\Theta\right),$$

where $\mathbf{E}\left(\cdot\right)$ denotes the Aumann expectation of the random set in parenthesis, see Molchanov (2005, Chapter 2). Denote by $\widehat{\tilde{B}}$ the estimator of $\tilde{B}$ (the unique convex set corresponding to the estimated support function) and by $\mathcal{B}_{\widehat{c}_{n(1-\tau)}}$ a ball in $\mathbb{R}^d$ centered at zero and with radius

$\widehat{c}_{n(1-\tau)}$, with $\widehat{c}_{n(1-\tau)}$ the Bayesian bootstrap estimate of the $1-\tau$ quantile of $f\left(\mathbb{G}\left[h_k\left(t\right)\right]\right)$, with $f(s(t)) = \sup_{t\in T}\left\{-s\left(t\right)\right\}_+$, see Section 4.3. Following arguments in Beresteanu and Molinari (2008, Section 2.3), one can construct a (convex) confidence set CS_n such that $\sup_{\alpha\in\mathcal{A}}\left(\sigma_{\widehat{\widetilde{B}}}(q,\alpha) - \sigma_{CS_n}(q)\right) = \widehat{c}_{n(1-\tau)}$ for all $q \in \mathcal{S}^{d-1}$, where $\sigma_A\left(\cdot\right)$ denotes the support function of the set A. It then follows that

$$\lim_{n\to\infty} \mathrm{P}\left(\sup_{q,\alpha\in\mathcal{S}^{d-1}\times\mathcal{A}} \left|\sigma_{\tilde{B}}(q,\alpha) - \sigma_{CS_n}(q)\right|_+ = 0\right) = 1-\tau.$$

Lemma 7. *For a given* $\delta > 0$, *one can jitter* x *via* $z = x + \sigma_\delta\left[0\ \eta\right]'$, *so as to obtain a set* $\tilde{B}$ *such that* $\sup_{\alpha\in\mathcal{A}}\rho_H\left(\tilde{B},B\right) \leq \delta$ *and*

$$1-\tau-\gamma(\delta) \geq \lim_{n\to\infty} \mathrm{P}\left(\sup_{q,\alpha\in\mathcal{S}^{d-1}\times\mathcal{A}} \left|\sigma_B(q,\alpha) - \left(\sigma_{CS_n}(q)+\delta\right)\right|_+ = 0\right) \geq 1-\tau, \qquad \text{(B.12)}$$

where $\gamma(\delta) = \mathrm{P}\left(\sup_{t\in T}\left\{-\mathbb{G}\left[h_k\left(t\right)\right]\right\}_+ > 2\delta\right)$.

Proof. Observe that $\rho_H\left(\tilde{B},B\right) = \rho_H\left(\Sigma\mathbf{E}\left(z\Theta\right),\Sigma\mathbf{E}\left(x\Theta\right)\right)$. By the properties of the Aumann expectation (see, e.g., Molchanov (2005, Theorem 2.1.17)),

$$\rho_H\left(\Sigma\mathbf{E}\left(z\Theta\right),\Sigma\mathbf{E}\left(x\Theta\right)\right) \leq \mathrm{E}\left[\rho_H\left(\Sigma\left(z\Theta\right),\Sigma\left(x\Theta\right)\right)\right].$$

In turn,

$$\begin{aligned}
&\sup_{\alpha\in\mathcal{A}}\mathrm{E}\left[\rho_H\left(\Sigma\left(z\Theta\right),\Sigma\left(x\Theta\right)\right)\right]\\
= \;& \sup_{\alpha\in\mathcal{A}}\mathrm{E}\left[\sup_{v=\Sigma'q:\|v\|=1}\left|\sup_{\theta\in\Theta}\left(v_1+z_2v_2\right)\tilde{\theta} - \sup_{\theta\in\Theta}\left(v_1+x_2v_2\right)\theta\right|\right]\\
= \;& \sup_{\alpha\in\mathcal{A}}\mathrm{E}\left[\sup_{v=\Sigma'q:\|v\|=1}\left|\left(v_1+x_2v_2+\sigma\eta v_2\right)\left(\theta_0 1\left(v_1+x_2v_2+\sigma\eta v_2<0\right)+\theta_1 1\left(v_1+x_2v_2+\sigma\eta v_2>0\right)\right)\right.\right.\\
&\left.\left.-\left(v_1+x_2v_2\right)\left(\theta_0 1\left(v_1+x_2v_2<0\right)+\theta_1 1\left(v_1+x_2v_2>0\right)\right)\right|\right]\\
\leq \;& \sup_{\alpha\in\mathcal{A}}\mathrm{E}\left[\sup_{v=\Sigma'q:\|v\|=1}\left|\sigma\eta v_2\left(\theta_0 1\left(v_1+x_2v_2+\sigma\eta v_2<0\right)+\theta_1 1\left(v_1+x_2v_2+\sigma\eta v_2>0\right)\right)\right|\right]\\
&+\sup_{\alpha\in\mathcal{A}}\mathrm{E}\left[\sup_{v=\Sigma'q:\|v\|=1}\left|\left(v_1+x_2v_2\right)\left(\theta_1-\theta_0\right)\left(1\left(0<-\left(v_1+x_2v_2\right)<\sigma\eta v_2\right)-1\left(0<v_1+x_2v_2<-\sigma\eta v_2\right)\right)\right|\right]\\
\leq \;& \sigma\mathrm{E}\left|\eta\right|\left(\sup_{\alpha\in\mathcal{A}}\mathrm{E}\left|\theta_0(x,\alpha)\right|+\sup_{\alpha\in\mathcal{A}}\mathrm{E}\left|\theta_1(x,\alpha)\right|+\sup_{\alpha\in\mathcal{A}}\mathrm{E}\left|\theta_1(x,\alpha)-\theta_0(x,\alpha)\right|\right).
\end{aligned}$$

Hence, we can choose $\sigma_\delta = \frac{\delta}{\mathrm{E}|\eta|\left(\sup_{\alpha\in\mathcal{A}}\mathrm{E}|\theta_0(x,\alpha)|+\sup_{\alpha\in\mathcal{A}}\mathrm{E}|\theta_1(x,\alpha)|+\sup_{\alpha\in\mathcal{A}}\mathrm{E}|\theta_1(x,\alpha)-\theta_0(x,\alpha)|\right)}$.

Now observe that because $\sup_{\alpha\in\mathcal{A}} \rho_H\left(\tilde{B}, B\right) \leq \delta$, we have $B(\alpha) \subseteq \tilde{B}(\alpha) \oplus \mathcal{B}_\delta$ for all $\alpha \in \mathcal{A}$, where "$\oplus$" denotes Minkowski set summation, and therefore

$$\begin{aligned}\sup_{\alpha\in\mathcal{A}} \left(\sigma_{\tilde{B}}(q,\alpha) - \sigma_{CS_n}(q)\right) &\leq 0 \ \forall\ q \in \mathcal{S}^{d-1} \\ &\Longrightarrow \sup_{\alpha\in\mathcal{A}} \left(\sigma_B(q,\alpha) - (\sigma_{CS_n}(q) + \delta)\right) \leq 0 \ \forall\ q \in \mathcal{S}^{d-1},\end{aligned}$$

from which the second inequality in (B.12) follows. Notice also that $\tilde{B}(\alpha) \subseteq B(\alpha) \oplus \mathcal{B}_\delta$ for all $\alpha \in \mathcal{A}$, and therefore

$$\begin{aligned}\sup_{\alpha\in\mathcal{A}} \left(\sigma_B(q,\alpha) - (\sigma_{CS_n}(q) + \delta)\right) &\leq 0 \ \forall\ q \in \mathcal{S}^{d-1} \\ &\Longrightarrow \sup_{\alpha\in\mathcal{A}} \left(\sigma_{\tilde{B}}(q,\alpha) - (\sigma_{CS_n}(q) + 2\delta)\right) \leq 0 \ \forall\ q \in \mathcal{S}^{d-1},\end{aligned}$$

from which the first inequality in (B.12) follows. Because $\delta > 0$ is chosen by the researcher, inference is arbitrarily slightly conservative. Note that a similar argument applies if one uses a Kolmogorov statistic rather than a directed Kolmogorov statistic. Moreover, the Hausdorff distance among convex compact sets is larger than the L_p distance among them (see, e.g., Vitale (1985, Theorem 1)), and therefore a similar conclusion applies for Cramer-Von-Mises statistics. □

B.5. **Lemmas on Entropy Bounds.** We collect frequently used facts in the following lemma.

Lemma 8. *Let Q be any probability measure whose support concentrates on a finite set.*

(1) *Let $\mathcal{F}$ be a measurable VC class with a finite VC index k or any other class whose entropy is bounded above by that of such a VC class, then its entropy obeys*

$$\log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L^2(Q)) \lesssim 1 + k \log(1/\epsilon)$$

Examples include e.g., linear functions $\mathcal{F} = \{\alpha' w_i, \alpha \in \mathbb{R}^k, \|\alpha\| \leq C\}$ and their indicators $\mathcal{F} = \{1\{\alpha' w_i > 0\}, \alpha \in \mathbb{R}^k, \|\alpha\| \leq C\}$.

(2) *Entropies obey the following rules for sets created by addition, multiplication, and unions of measurable function sets $\mathcal{F}$ and $\mathcal{F}'$:*

$$\begin{aligned}&\log N(\epsilon \|F + F'\|_{Q,2}, \mathcal{F} + \mathcal{F}', L^2(Q)) \leq B \\ &\log N(\epsilon \|F \cdot F'\|_{Q,2}, \mathcal{F} \cdot \mathcal{F}', L^2(Q)) \leq B \\ &\log N(\epsilon \|F \vee F'\|_{Q,2}, \mathcal{F} \cup \mathcal{F}', L^2(Q)) \leq B \\ &B = \log N\left(\frac{\epsilon}{2}\|F\|_{Q,2}, \mathcal{F}, L^2(Q)\right) + \log N\left(\frac{\epsilon}{2}\|F'\|_{Q,2}, \mathcal{F}', L^2(Q)\right).\end{aligned}$$

(3) *Entropies are preserved by multiplying a measurable function class $\mathcal{F}$ with a random variable g_i:*

$$\log N(\epsilon \||g|F\|_{Q,2}, g\mathcal{F}, L^2(Q)) \lesssim \log N\left(\epsilon/2 \|F\|_{Q,2}, \mathcal{F}, L^2(Q)\right)$$

(4) *Entropies are preserved by integration or taking expectation: for* $f^*(x) := \int f(x,y)d\mu(y)$ *where* μ *is some probability measure,*

$$\log N(\epsilon\|F\|_{Q,2}, \mathcal{F}^*, L^2(Q)) \leq \log N\left(\epsilon\|F\|_{Q,2}, \mathcal{F}, L^2(Q)\right)$$

Proof. For the proof of (1)-(3) see e.g., Andrews (1994). For the proof of (4), see e.g., Ghosal, Sen, and van der Vaart (2000, Lemma A2). □

Next consider function classes and their envelops

$$\begin{aligned}
\mathcal{H}_1 &= \{q'\Sigma \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i < 0\}] J_0^{-1}(\alpha) p_i \varphi_{i0}(\alpha), t \in T\}, \quad H_1 \lesssim \|z_i\| \xi_k F_1 \\
\mathcal{H}_2 &= \{q'\Sigma \mathrm{E}[z_i p_i' 1\{q'\Sigma z_i > 0\}] J_1^{-1}(\alpha) p_i \varphi_{i1}(\alpha), t \in T\}, \quad H_2 \lesssim \|z_i\| \xi_k F_1 \\
\mathcal{H}_3 &= \{q'\Sigma x_i z_i' \Sigma \mathrm{E}\left[z_i w_{i,q'\Sigma}(\alpha)\right], t \in T\}, \quad H_3 \lesssim \|x_i\|\|z_i\| \\
\mathcal{H}_4 &= \{q'\Sigma z_i w_{i,q'\Sigma}(\alpha), t \in T\}, \quad H_4 \lesssim \|z_i\| F_2 \\
\mathcal{F}_3 &= \{\bar{\mu}' J^{-1}(\alpha) p_i \varphi_i(\alpha), \ \alpha \in \mathcal{A}\}, \quad F_3 \lesssim \xi_k F_1,
\end{aligned} \tag{B.13}$$

where $\bar{\mu}'$ is defined in equation (B.9).

Lemma 9. *1. (a) The following bounds on the empirical entropy apply*

$$\begin{aligned}
\log N(\epsilon\|H_1\|_{\mathbb{P}_n,2}, \mathcal{H}_1, L^2(\mathbb{P}_n)) &\lesssim_{\mathrm{P}} \log n + \log(1/\epsilon) \\
\log N(\epsilon\|H_2\|_{\mathbb{P}_n,2}, \mathcal{H}_2, L^2(\mathbb{P}_n)) &\lesssim_{\mathrm{P}} \log n + \log(1/\epsilon) \\
\log N(\epsilon\|F_3\|_{\mathbb{P}_n,2}, \mathcal{F}_3, L^2(\mathbb{P}_n)) &\lesssim_{\mathrm{P}} \log n + \log(1/\epsilon)
\end{aligned}$$

(b) Moreover similar bounds apply to function classes $g_i(\mathcal{H}_l^o - \mathcal{H}_l^o)$ *with the envelopes given by* $|g_i| 4H\ell$*, where* g_i *is a random variable.*

2. (a) The following bounds on the uniform entropy apply

$$\begin{aligned}
\sup_Q \log N(\epsilon\|H_1\|_{Q,2}, \mathcal{H}_1, L^2(Q)) &\lesssim k\log(1/\epsilon) \\
\sup_Q \log N(\epsilon\|H_2\|_{Q,2}, \mathcal{H}_2, L^2(Q)) &\lesssim k\log(1/\epsilon) \\
\sup_Q \log N(\epsilon\|F_3\|_{Q,2}, \mathcal{F}_3, L^2(Q)) &\lesssim k\log(1/\epsilon) \\
\sup_Q \log N(\epsilon\|H_3\|_{Q,2}, \mathcal{H}_3, L^2(Q)) &\lesssim \log(1/\epsilon) \\
\sup_Q \log N(\epsilon\|H_4\|_{Q,2}, \mathcal{H}_4, L^2(Q)) &\lesssim \log(1/\epsilon).
\end{aligned}$$

(b) Moreover similar bounds apply to function classes $g_i(\mathcal{H}_l^o - \mathcal{H}_l^o)$ *with the envelopes given by* $|g_i| 4H\ell$*, where* g_i *is a random variable.*

Proof. Part 1 (a). Case of $\mathcal{H}_1$ and $\mathcal{H}_2$. We shall detail the proof for this case, while providing shorter arguments for others, as they are simpler or similar.

Note that $\mathcal{H}_1 \subseteq \mathcal{M}_1 \cdot \mathcal{M}_2 \cdot \mathcal{F}_1$, where $\mathcal{M}_1 = \{q'\Sigma z_i, q \in \mathcal{S}^{d-1}\}$ with envelope $M_1 = \|z_i\|$ is VC with index $\dim(z_i) + \dim(x_i)$, and $\mathcal{M}_2 = \{\gamma(q) J_0^{-1}(\alpha) p_i, (q,\alpha) \in \mathcal{S}^{d-1} \times \mathcal{A}\}$ with envelope $M_2 \lesssim \|\xi_k\|$, $\mathcal{F}_1 = \{\varphi_{i0}(\alpha), \alpha \in \mathcal{A}\}$ with envelope F_1, where $\gamma(q)$ is uniformly

Holder in $q \in \mathcal{S}^{d-1}$ by Lemma 3. Elementary bounds yield

$$\begin{aligned} &\|m_2(t) - m_2(\tilde{t})\|_{\mathbb{P}_n,2} \leq L_{1n}\|\alpha - \tilde{\alpha}\| + L_{2n}\|q - \tilde{q}\|, \\ &L_{1n} \lesssim \sup_{\alpha \in \mathcal{A}} \|J^{-1}(\alpha)\| \|\xi_k\| \quad L_{2n} \lesssim \|\mathbb{E}_n[p_i p_i']\|, \\ &\log L_{1n} \lesssim_{\mathrm{P}} \log n \quad \text{and} \quad \log L_{2n} \lesssim_{\mathrm{P}} 1. \end{aligned}$$

Note that $\log \xi_k \lesssim \log n$ by assumption, $\sup_{\alpha \in \mathcal{A}} \|J^{-1}(\alpha)\| \lesssim 1$ by assumption, $\|\mathbb{E}_n[p_i p_i']\| \lesssim_{\mathrm{P}} 1$ by Lemma 10. The sets $\mathcal{S}^{d-1}$ and $\mathcal{A}$ are compact subsets of Euclidian space of fixed dimension, and so can be covered by a constant times $1/\epsilon^c$ balls of radius ϵ for some constant $c > 0$. Therefore, we can conclude

$$\log N(\epsilon \|M_2\|_{\mathbb{P}_n,2}, \mathcal{M}_2, L_2(\mathbb{P}_n)) \lesssim_{\mathrm{P}} \log n + \log(1/\epsilon).$$

Repeated application of Lemma 8 yields the conclusion, given the assumption on the function class $\mathcal{F}_1$. The case for $\mathcal{H}_2$ is very similar.

Case of $\mathcal{F}_3$. Note that $\mathcal{F}_3 \subset \mathcal{M}_2 \cdot \mathcal{F}_1$ and $\|\bar{\mu}\| = o_{\mathrm{P}}(1)$ by Step 4 in the proof of Lemma 1. Repeated application of Lemma 8 yields the conclusion, given the assumption on the function class $\mathcal{F}_1$.

Part 1 (b). Note that $\mathcal{H}^o = \mathcal{H} - \mathrm{E}[\mathcal{H}^o]$, so it is created by integration and summation. Hence repeated application of Lemma 8 yields the conclusion.

Part 2. (a) Case of $\mathcal{H}_1, \mathcal{H}_2$, and $\mathcal{F}_3$. Note that all of these classes are subsets of $\{\mu' p_i, \|\mu\| \leq C\} \cdot \mathcal{F}_1$ with envelope $\xi_k F_1$. The claim follows from repeated application of Lemma 8.

Case of $\mathcal{H}_3$. Note that $\mathcal{H}_3 \subset \{q'\Sigma x_i z_i' \mu, \|\mu\| \leq C\}$ with envelope $\|x_i\|\|z_i\|$. The claim follows from repeated application of Lemma 8.

Case of $\mathcal{H}_4$. Note that $\mathcal{H}_4$ is a subset of a function class created from taking the class $\mathcal{F}_2$ multiplying it with indicator function class $1\{q'\Sigma z_i > 0, q \in \mathcal{S}^{d-1}\}$ and with function class $\{q'\Sigma z_i, q \in \mathcal{S}^{d-1}\}$ and then adding the resulting class to itself. The claim follows from repeated application of Lemma 8.

Part 2 (b). Note that $\mathcal{H}^o = \mathcal{H} - \mathrm{E}[\mathcal{H}^o]$, so it is created by integration and summation. Hence repeated application of Lemma 8 yields the conclusion.

□

B.6. **Auxiliary Maximal and Random Matrix Inequalities.** We repeatedly use the following matrix LLN.

Lemma 10 (Matrix LLN)**.** *Let $Q_1, ..., Q_n$ be i.i.d. symmetric non-negative matrices such that $Q = \mathrm{E}Q_i$ and $\|Q_i\| \leq M$, then for $\widehat{Q} = \mathbb{E}_n Q_i$*

$$\mathrm{E}\|\widehat{Q} - Q\| \lesssim \sqrt{\frac{M(1 + \|Q\|)\log k}{n}}.$$

In particular, if $Q_i = p_i p_i'$, with $\|p_i\| \leq \xi_k$, then

$$\mathrm{E}\|\widehat{Q} - Q\| \lesssim \sqrt{\frac{\xi_k^2(1 + \|Q\|)\log k}{n}}.$$

Proof. This is a variant of a result from Rudelson (1999). By the symmetrization lemma,

$$\Delta := \mathrm{E}\left\|\widehat{Q} - Q\right\| \leq 2\mathrm{E}\mathrm{E}_\epsilon \left\|\mathbb{E}_n[\epsilon_i Q_i]\right\|$$

where ϵ_i are Rademacher random variables. The Khintchine inequality for matrices, which was shown by Rudelson (1999) to follow from the results of Lust-Piquard and Pisier (1991), states that

$$\mathrm{E}_\epsilon \left\|\mathbb{E}_n[\epsilon_i Q_i]\right\| \lesssim \sqrt{\frac{\log k}{n}} \left\|\left(\mathbb{E}_n[Q_i^2]\right)^{1/2}\right\|.$$

Since (remember that $\|\cdot\|$ is the operator norm)

$$\mathrm{E}\left\|\left(\mathbb{E}_n[Q_i^2]\right)^{1/2}\right\| = \mathrm{E}\left\|\left(\mathbb{E}_n[Q_i^2]\right)\right\|^{1/2} \leq \left[M\mathrm{E}\left\|\mathbb{E}_n Q_i\right\|\right]^{1/2},$$

and

$$\left\|\mathbb{E}_n Q_i\right\| \leq \Delta + \|Q\|,$$

one has

$$\Delta \leq 2\sqrt{\frac{M \log k}{n}} \left[\Delta + \|Q\|\right]^{1/2}.$$

Solving for Δ gives

$$\Delta \leq \sqrt{\frac{4M\|Q\|\log k}{n} + \left(\frac{M\log k}{n}\right)^2} + \frac{M\log k}{n},$$

which implies the result stated in the lemma if $\frac{M\log k}{n} < 1$. □

We also use the following maximal inequality.

Lemma 11. *Consider a separable empirical process* $\mathbb{G}_n(f) = n^{-1/2}\sum_{i=1}^n \{f(Z_i) - \mathrm{E}[f(Z_i)]\}$, *where* $Z_1, \ldots, Z_n$ *is an underlying independent data sequence on the space* $(\Omega, \mathcal{G}, \mathrm{P})$, *defined over the function class* $\mathcal{F}$, *with an envelope function* $F \geq 1$ *such that* $\log[\max_{i\leq n}\|F\|] \lesssim_{\mathrm{P}} \log n$ *and*

$$\log N\left(\varepsilon\|F\|_{\mathbb{P}_n,2}, \mathcal{F}, L_2(\mathbb{P}_n)\right) \leq \upsilon m \log(\kappa/\epsilon),\ \ 0 < \epsilon < 1,$$

with some constants $0 < \log\kappa \lesssim \log n$, m *potentially depending on* n, *and* $1 < \upsilon \lesssim 1$. *For any* $\delta \in (0,1)$, *there is a large enough constant* K_δ, *such that for* n *sufficiently large, then*

$$\mathrm{P}\left\{\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)| \leq K_\delta\sqrt{m\log n}\max\left\{\sup_{i\leq n, f\in\mathcal{F}}\|f(Z_i)\|_{\mathrm{P},2},\ \sup_{f\in\mathcal{F}}\|f\|_{\mathbb{P}_n,2}\right\}\right\} \geq 1-\delta.$$

Proof. TO BE ADDED. This is a restatement of Lemma 19 from Belloni and Chernozhukov (2009b). □

References

Andrews, D. W., and P. Jia (2008): "Inference for Parameters Defined by Moment Inequalities: A Recommended Moment Selection Procedure," *SSRN eLibrary*.

Andrews, D. W. K. (1994): *Empirical Process Methods in Econometrics*vol. IV of *Handbook of Econometrics*, pp. 2248–2294.

Andrews, D. W. K., and X. Shi (2009): "Inference Based on Conditional Moment Inequalities," mimeo.

Andrews, D. W. K., and G. Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," *Econometrica*, 78, 119–157.

Artstein, Z., and R. A. Vitale (1975): "A Strong Law of Large Numbers for Random Compact Sets," *The Annals of Probability*, 3(5), 879–882.

Belloni, A., and V. Chernozhukov (2009a): "On the computational complexity of MCMC-based estimators in large samples," *The Annals of Statistics*, 37(4), 2011–2055.

——— (2009b): "Post-L1-Penalized Estimators in High-Dimensional Linear Regression Models," .

Belloni, A., V. Chernozhukov, and I. Fernandez-Val (2011): "Conditional Quantile Processes based on Series or Many Regressors," arXiv:1105.6154v1.

Beresteanu, A., and F. Molinari (2008): "Asymptotic Properties for a Class of Partially Identified Models," *Econometrica*, 76(4), 763–814.

Blau, F. D., and L. M. Kahn (1997): "Swimming Upstream: Trends in the Gender Wage Differential in the 1980s," *Journal of Labor Economics*, 15, 1–42.

Blundell, R., A. Gosling, H. Ichimura, and C. Meghir (2007): "Changes in the Distribution of Male and Female Wages Accounting for Employment Composition UsingBounds," *Econometrica*, 75(2), 323–363.

Bontemps, C., T. Magnac, and E. Maurin (2010): "Set Identified Linear Models," Discussion paper, CeMMAP Working Paper CWP 13/11.

Bugni, F. A. (2010): "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set," *Econometrica*, 78, 735–753.

Canay, I. A. (2010): "EL Inference for Partially Identified Models: Large Deviations Optimality and Bootstrap Validity," *Journal of Econometrics*, 156, 408–425.

Card, D. (1999): *The Causal Effect of Education on Earnings*chap. 30, pp. 1801–1863. Elsevier.

Card, D., and J. E. DiNardo (2002): "Skill Biased Technological Change and Rising Wage Inequality: Some Problems and Puzzles," *Journal of Labor Economics*, 20, 733–783.

Chen, X. (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics*, 6.

Chernozhukov, V., H. Hong, and E. Tamer (2007): "Estimation and Confidence Regions for Parameter Sets in Econometric Models," *Econometrica*, 75(5), 1243–1284.

Chernozhukov, V., S. Lee, and A. Rosen (2009): "Intersection Bounds: Estimation and Inference," CeMMAP Working Paper CWP19/09.

Davydov, Y. A., M. A. Lifshits, and N. V. Smorodina (1998): *Local Properties of Distribuions of Stochastic Functionals*. American Mathematical Society.

Einav, L., A. Finkelstein, S. Ryan, P. Schrimpf, and M. Cullen (2011): "Selection on Moral Hazard in Health Insurance," NBER Working Paper Nr 16969.

Foresi, S., and F. Peracchi (1995): "The Conditional Distribution of Excess Returns: An Empirical Analysis," *Journal of the American Statistical Association*, 90, 451–466.

Galichon, A., and M. Henry (2009): "A test of non-identifying restrictions and confidence regions for partially identified parameters," *Journal of Econometrics*, 152(2), 186–196.

Ghosal, S., A. Sen, and A. W. van der Vaart (2000): "Testing Monotonicity of Regression," *Annals of Statistics*, 28, 1054–1082.

Han, A., and J. A. Hausman (1990): "Flexible Parametric Estimation of Duration and Competing Risk Models," *Journal of Applied Econometrics*, 5, 1–28.

He, X., and Q.-M. Shao (2000): "On Parameters of Increasing Dimensions," *Journal of Multivariate Analysis*, 73(1), 120–135.

Hirano, K., G. Imbens, and G. Ridder (2003): "Efficient estimation of average treatment effects using the estimated propensity score," *Econometrica*, 71(4), 1161–1189.

Imbens, G. W., and C. F. Manski (2004): "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845–1857.

Imbens, G. W., and J. M. Wooldridge (2009): "Recent developments in the econometrics of program evaluation," *Journal of Economic Literature*, 47, 5–86.

Jasso, G., and M. R. Rosenzweig (2008): "Selection Criteria and the Skill Composition of Immigrants: A Comparative Analysis of Australian and U.S. Employment Immigration," DP 3564, IZA.

Juster, F. T., and R. Suzman (1995): "An Overview of the Health and Retirement Study," *Journal of Human Resources*, 30 (Supplement), S7–S56.

Kaido, H. (2010): "A Dual Approach to Inference for Partially Identified Econometric Models," mimeo.

Kline, P., and A. Santos (2010): "Sensitivity to Missing Data Assumptions: Theory and An Evaluation of the U.S. Wage Structure," mimeo.

Lorentz, G. G. (1986): *Approximation of Functions*. Chelsea.

Lust-Piquard, F., and G. Pisier (1991): "Non commutative Khintchine and Paley inequalities," *Arkiv för Matematik*, 29(1), 241–260.

Magnac, T., and E. Maurin (2008): "Partial Identification in Monotone Binary Models: Discrete Regressors and Interval Data," *Review of Economic Studies*, 75(3), 835–864.

Manski, C. F. (1989): "Anatomy of the Selection Problem," *The Journal of Human Resources*, 24(3), 343–360.

——— (2003): *Partial Identification of Probability Distributions*. Springer Verlag, New York.

——— (2007): *Identification for Prediction and Decision*. Harvard University Press, Cambridge, MA.

Manski, C. F., and E. Tamer (2002): "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70(2), 519–546, ArticleType: primary_article / Full publication date: Mar., 2002 / Copyright 2002 The Econometric Society.

Molchanov, I. (2005): *Theory of Random Sets*. Springer Verlag, London.

Mulligan, C. B., and Y. Rubinstein (2008): "Selection, Investment, and Women's Relative Wages Over Time," *Quarterly Journal of Economics*, 123(3), 1061–1110.

Newey, W. K. (1997): "Convergence rates and asymptotic normality for series estimators," *Journal of Econometrics*, 79(1), 147–168.

Olley, G. S., and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64(6), 1263–1297, ArticleType: primary_article / Full publication date: Nov., 1996 / Copyright 1996 The Econometric Society.

Picketty, T. (2005): "Top Income Shares in the Long Run: An Overview," *Journal of the European Economic Association*, 3, 382–392.

Pollard, D. (2002): *A User's Guide to Measure Theoretic Probability*. Cambridge.

Ponomareva, M., and E. Tamer (2010): "Misspecification in Moment Inequality Models: Back to Moment Equalities?," *Econometrics Journal*, forthcoming.

Rockafellar, R. (1970): *Convex Analysis*. Princeton University Press.

Romano, J. P., and A. M. Shaikh (2008): "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

——— (2010): "Inference for the Identified Set in Partially Identified Econometric Models," *Econometrica*, 78, 169–211.

Rosen, A. M. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146(1), 107–117.

Rudelson, M. (1999): "Random vectors in the isotropic position," *Journal of Functional Analysis*, 164(1), 60–72.

Schneider, R. (1993): *Convex Bodies: The Brunn-Minkowski Theory*, Encyclopedia of Mathematics and Its Applications. Cambridge University Press, Cambridge, UK.

Stoye, J. (2007): "Bounds on Generalized Linear Predictors with Partially Identified Outcomes," *Reliable Computing*, 13, 293–302.

——— (2009): "More on Confidence Intervals for Partially Identified Parameters," *Econometrica*, 77, 1299–1315.

Townsend, R. M., and S. S. Urzua (2009): "Measuring the Impact of Financial Intermediation: Linking Contract Theory to Econometric Policy Evaluation," *Macroeconomic Dynamics*, 13(Supplement S2), 268–316.

van der Vaart, A. W. (2000): *Asymptotic statistics*. Cambridge University Press.

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

Vitale, R. A. (1985): "L_p Metric for Compact, Convex Sets," *Journal of Approximation Theory*, 45, 280–287.

Department of Economics, M. I. T.

Department of Economics, M. I. T.

Department of Economics, Cornell University

Department of Economics, University of British Columbia