

The sorted effects method: discovering heterogeneous effects beyond their averages

Victor Chernozhukov
Ivan Fernandez-Val
Ye Luo

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP74/15

THE SORTED EFFECTS METHOD: DISCOVERING HETEROGENEOUS EFFECTS BEYOND THEIR AVERAGES

VICTOR CHERNOZHUKOV, IVAN FERNANDEZ-VAL, YE LUO

ABSTRACT. The partial (*ceteris paribus*) effects of interest in nonlinear and interactive linear models are heterogeneous as they can vary dramatically with the underlying observed or unobserved covariates. Despite the apparent importance of heterogeneity, a common practice in modern empirical work is to largely *ignore it* by reporting average partial effects (or, at best, average effects for some groups, see e.g. Angrist and Pischke (2008)). While average effects provide very convenient scalar summaries of typical effects, by definition they fail to reflect the entire variety of the heterogeneous effects. In order to discover these effects much more fully, we propose to estimate and report *sorted effects* – a collection of estimated partial effects sorted in increasing order and indexed by percentiles. By construction the sorted effect curves *completely represent* and help visualize all of the heterogeneous effects in one plot. They are as convenient and easy to report in practice as the conventional average partial effects. We also provide a quantification of uncertainty (standard errors and confidence bands) for the estimated sorted effects. We apply the *sorted effects method* to demonstrate several striking patterns of gender-based discrimination in wages, and of race-based discrimination in mortgage lending.

Using differential geometry and functional delta methods, we establish that the estimated sorted effects are consistent for the true sorted effects, and derive asymptotic normality and bootstrap approximation results, enabling construction of pointwise confidence bands (pointwise with respect to percentile indices). We also derive functional central limit theorems and bootstrap approximation results, enabling construction of simultaneous confidence bands (simultaneous with respect to percentile indices). The derived statistical results in turn rely on establishing Hadamard differentiability of the multivariate sorting operator, a result of independent mathematical interest.

Keywords: Sorting, Partial Effect, Marginal Effect, Sorted Percentiles, Nonlinear Model, Functional Analysis, Differential Geometry

Date: December 17, 2015.

MIT, BU and University of Florida. We thank Isaiah Andrews, Joshua Angrist, Denis Chetverikov, Jerry Hausman, Anna Mikusheva, Whitney Newey and participants at MIT econometrics seminar for helpful discussions.

1. INTRODUCTION

In nonlinear and interactive linear models the partial (*ceteris paribus*) effects of interest often vary with respect to the underlying covariates. For example, in a binary response model with conditional choice probability $P(Y = 1 | X) = F(X^\top \beta)$, the predictive effect (PE) of a marginal change in a continuous covariate X_j on the conditional choice probability is

$$\Delta(X) = f(X^\top \beta) \beta_j, \quad f = \partial F,$$

which generally varies with the covariate vector X , as X varies according to some distribution, say μ . A common empirical practice is to report the average partial effect (APE),

$$E\Delta(X) = \int \Delta(x) d\mu(x),$$

as a single summary measure of the PE (e.g., Wooldridge (2010, Chap. 2)). However, this measure completely disregards the heterogeneity of the PE and may give a very incomplete picture of the impact of the covariates.

In this paper we propose to complement the APE by reporting the entire set of PEs sorted in increasing order and indexed by a ranking with respect to the distribution of the covariates in the part of the population of interest. These sorted effects correspond to percentiles of the PE,

$$\Delta_\mu^*(u) = u - \text{quantile of } \Delta(X), \quad X \sim \mu, \quad (1.1)$$

and provide a more complete representation of the heterogeneity of $\Delta(X)$. We note that the definition (1.1) applies to any type of PE, and that the covariates X can contain both observed and unobserved components.¹ We shall call these effects as sorted predictive effects (SPE) by default, as most models are predictive.²

Figure 1 illustrates the SPE of being black on the probability of mortgage denial, conditional on race and other applicant characteristics relevant for the bank decision. The SPE varies strongly from 0% to 15%, and does not coincide with the average effect of 5%. Single, black applicants with bad credit histories have the highest PEs. This fact can be deduced by classification – determining the features X of the subpopulation where the PE is large, i.e. all X 's such that $\Delta(X) > .14$. We refer the reader to Section 4 for a detailed discussion of this example.

¹We defer the discussion of unobserved components to Section 2.

²When the underlying model has a structural or causal interpretation, we may use the name sorted structural effects or sorted treatment effects.

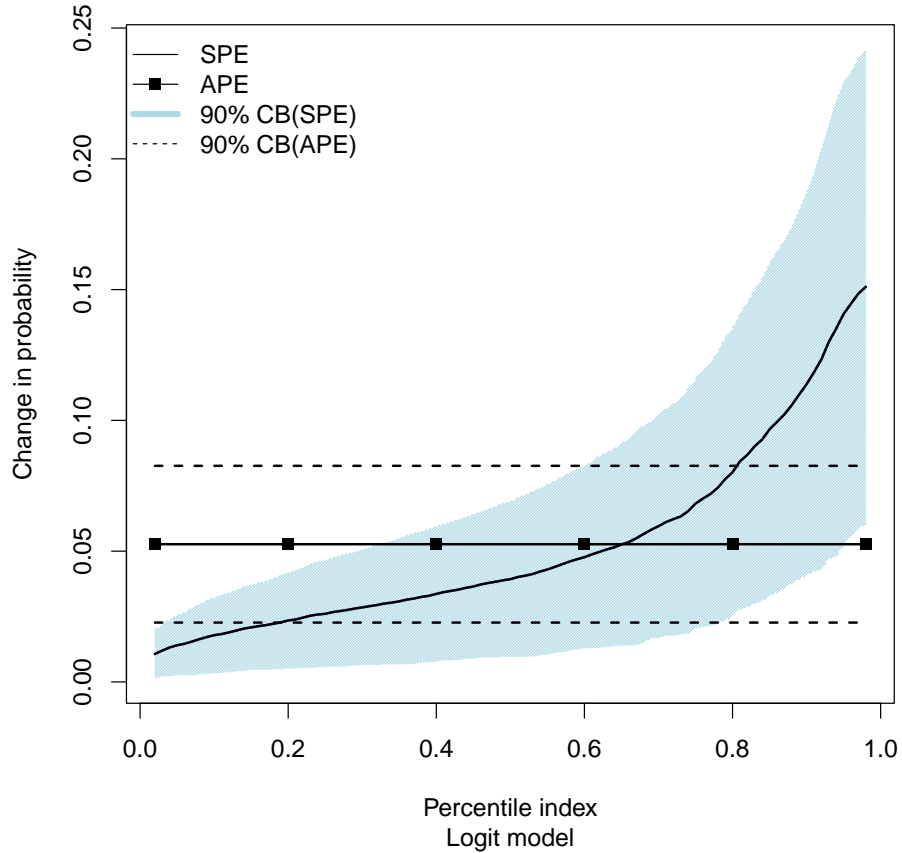


FIGURE 1. APE and SPE (introduced in this paper) of being black on the probability of mortgage denial. Estimates and 90% bootstrap uniform confidence bands (derived in this paper) based on a logit model are shown.

Heterogeneous effects also arise in the most basic interactive linear models. Consider a conditional mean model for the Mincer earnings function:

$$Y = g(X) + \epsilon = P(T, W)^\top \beta + \epsilon, \quad E[\epsilon | T, W] = 0, \quad X = (T, W),$$

where Y is log-wage, T is an indicator of gender (or race, treatment, program participation), and W is a vector of labor market relevant characteristics. The vector $P(T, W)$ is a collection of transformations of T and W , involving some interaction between T and W . For example, the most basic specification is $P(T, W) = (TW, (1 - T)W)$. Then, the PE of changing $T = 0$ to $T = 1$ is

$$\Delta(X) = P(1, W)^\top \beta - P(0, W)^\top \beta,$$

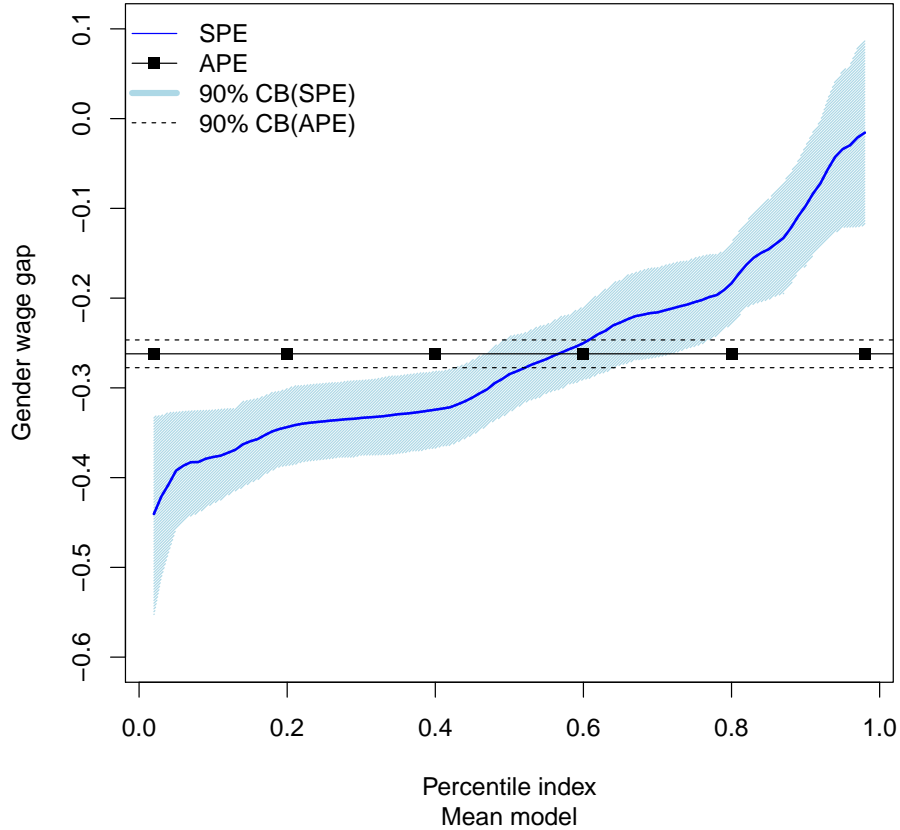


FIGURE 2. APE and SPE (introduced in this paper) of the gender wage gap for women. Estimates and 90% bootstrap uniform confidence bands (derived in this paper) based on a linear model with interactions for the conditional expectation function are shown.

which is a measure of the gender wage gap conditional on worker characteristics. The SPE function $u \mapsto \Delta_{\mu}^*(u)$ provides again a complete summary of the entire variety of PEs. Figure 2 illustrates the SPE of the conditional gender wage gap. The SPE varies strongly from -45% to 0% , and does not coincide with the average PE of -27% . The PE is especially (negatively) strong for married women, low educated or with high potential experience – this follows from looking at the average characteristics of the subpopulation with covariate values X such that $\Delta(X) \leq -0.4$. We refer the reader to Section 4 for a detailed discussion of this example.

The general settings that we deal with in this paper as well as the specific results we obtain are as follows: Let X denote a covariate vector, $\Delta(X)$ denote a generic PE of interest, μ denote the distribution of X , and \mathcal{X} denote the support of X . The SPE is obtained by sorting the multivariate function $x \mapsto \Delta(x)$ in increasing order with respect to μ . Using tools from differential geometry, we show that this multivariate sorting operator is Hadamard differentiable with respect to the PE function Δ and the distribution μ at the regular values of $x \mapsto \Delta(x)$ on \mathcal{X} . This result allows us to derive the large sample properties of the empirical SPE, which replace Δ and μ by sample analogs, obtained from parametric or semi-parametric estimators, using the functional delta method. In particular, we derive a functional central limit theorem and a bootstrap functional central limit theorem for the empirical SPE. The main requirement of these theorems is that the empirical Δ and μ also satisfy functional central limit theorems, which hold for many estimators used in empirical economics under general sampling conditions. We use the properties of the empirical SPE to construct confidence sets for the SPE that hold uniformly over quantile indexes.

Related literature: Previously, Chernozhukov, Fernández-Val, and Galichon (2010) derived the properties of the rearrangement (sorting) operator in the univariate case with known μ (standard uniform distribution). Those results were motivated by solving a different problem – the crossing problem in conditional quantile estimation where monotonicity with respect to the quantile index needs to be restored– instead of the problem of summarizing heterogeneous effects by the sorted predictive effects. These prior results are inapplicable to our case, as soon as the dimension of X is greater than one, which is the case in all modern applications where heterogeneous effects are of interest. Moreover the previous results are not applicable even in the univariate case, since the measure μ is not known in all envisioned applications. The properties of the sorting operator are different in the multivariate case and require tools from differential geometry: computation of functional (Hadamard) derivatives of the sorting operator with respect to perturbations of Δ require us to work with integration on $(d_x - 1)$ -dimensional manifolds of the type $\{\Delta(x) = \delta\}$, where $d_x = \dim X$. Moreover, we also need to compute functional derivatives with respect to suitable perturbations of the measure μ .

In econometrics or statistics, Sasaki (2015) also used differential geometry to characterize the structural properties of derivatives of conditional quantiles in nonseparable models; and Kim and Pollard (1990) used tools from differential geometry to derive the large sample properties of maximum score and other cube root consistent estimators. Our uses of differential geometry and our results are different. Moreover, our results on the functional differentiability of the sorting operator in the multivariate case constitute a new mathematical result, which is of interest in its own right.

Organization of the paper: In Section 2 we discuss the quantities of interest in nonlinear and interactive linear models with examples, and introduce the SPE. In Section 3 we introduce the empirical SPE and outline the main inferential results. Section 4 provides two empirical examples. In Section 5 we characterize the analytical properties of the multivariate sorting operator. In Section 6 we derive the properties of the empirical SPE in large samples and show how to use these properties to make inference on the SPE uniformly over quantile indexes. We discuss how to incorporate discrete covariates in the PE and gather the proofs of the main results in the Appendix.

Notation: For a random variable X , \mathcal{X} denotes the interior of the support of X , μ denotes the distribution of X over \mathcal{X} , and $\hat{\mu}$ denotes an estimator of μ . We denote the expectation with respect to the distribution $\tilde{\mu}$ by $\mathbb{E}_{\tilde{\mu}}$. We denote the PE as $\Delta(x)$, the empirical PE as $\hat{\Delta}(x)$, and $\partial\Delta(x) := \partial\Delta(x)/\partial x$, the gradient of $x \mapsto \Delta(x)$. We also use $a \wedge b$ to denote the minimum of a and b . For a vector $v = (v_1, \dots, v_{d_v}) \in \mathbb{R}^{d_v}$, $\|v\|$ denotes the Euclidian norm of v , that is $\|v\| = \sqrt{v^\top v}$, where the superscript \top denotes transpose. For a non-negative integer r and an open set \mathcal{K} , the class \mathcal{C}^r on \mathcal{K} includes the set of r times continuously differentiable real valued functions on \mathcal{K} . The symbol \rightsquigarrow denotes weak convergence (convergence in distribution), and $\rightarrow_{\mathbb{P}}$ denotes convergence in (outer) probability.

2. SORTED EFFECTS IN NONLINEAR AND INTERACTIVE LINEAR MODELS

We discuss the objects of interest in nonlinear and interactive linear models and introduce the sorted effects.

2.1. Effects of Interest. We consider a general model characterized by a predictive function $g(X)$, where X is a d_x -vector of covariates that may contain unobserved components, as in quantile regression models. The function g usually arises from a model for a response variable Y , which can be discrete or continuous. We call the function g predictive because the underlying model can be either predictive or causal under additional assumptions, but we do not insist on estimands having a causal interpretation. For example, in a mean regression model, $g(X) = \mathbb{E}[Y | X]$ corresponds to the expectation function of Y conditional on X ; in a binary response model, $g(X) = \mathbb{P}[Y = 1 | X]$ corresponds to the choice probability of $Y = 1$ conditional on X ; in a quantile regression model, $g(X) = \mathbb{Q}_Y[\epsilon | Z]$, where the covariate $X = (\epsilon, Z)$ consist of the unobservable rank variable ϵ with a uniform distribution, $\epsilon | Z \sim U(0, 1)$, and the observed covariate vector Z , and where $\mathbb{Q}_Y[\tau | Z]$ denotes the conditional τ -quantile of Y given Z .

Let $X = (T, W)$, where T is the key covariate or treatment of interest, and W is a vector of control variables. We are interested in the effects of changes in the variable T on the function g holding W constant. These effects are usually called partial effects, marginal effects, or treatment effects. We call them predictive effects (PE) throughout the paper, as such name most accurately described the meaning of the estimand (especially when a causal interpretation is not available). If T is discrete, the PE is

$$\Delta(x) = \Delta(t, w) = g(t_1, w) - g(t_0, w) \quad (2.2)$$

where t_1 and t_0 are two values of T that might depend on t (e.g., $t_0 = t$ and $t_1 = t + 1$). This PE measures the effect of changing T from t_0 to t_1 holding W constant at w . If T is continuous and $t \mapsto \Delta(t, w)$ is differentiable, the PE is

$$\Delta(x) = \Delta(t, w) = \partial_t g(t, w), \quad (2.3)$$

where ∂_t denotes $\partial/\partial t$, the partial derivative with respect to t . This PE measures the effect of a marginal change of T from the level t holding W constant at w .

We consider the following examples in the empirical applications of Section 4.

Example 1 (Binary response model). Let Y be a binary response variable such as an indicator of mortgage denial, and X be a vector of covariates related to Y . The predictive function of the probit or logit model takes the form:

$$g(X) = \text{P}(Y = 1 \mid X) = F(P(X)^\top \beta),$$

where $P(X)$ is a vector of known transformations of X , β is a parameter vector, and F is the known distribution function (the standard normal distribution function in the probit model or logistic distribution function in the logit model). If T is a binary variable such as an indicator of black applicant and W is a vector of applicant characteristics relevant for the bank decision, the PE,

$$\Delta(x) = \Phi(P(1, w)^\top \beta) - \Phi(P(0, w)^\top \beta),$$

describes the difference in predicted probability of mortgage denial for a black applicant and a white applicant, conditional on a specific value w of the observable characteristics W . \square

Example 2 (Interactive linear model with additive error). Let Y be the logarithm of the wage. Suppose $X = (T, W)$, where T is the indicator of female worker and W are other worker characteristics. Suppose we model the conditional expectation function of log-wages using the linear interactive model:

$$Y = g(X) + \epsilon = P(T, W)^\top \beta + \epsilon, \quad \text{E}[\epsilon \mid T, W] = 0, \quad X = (T, W),$$

where $P(T, W)$ is a collection of transformations of T and W , involving some interaction between T and W . For example, $P(T, W) = (TW, (1 - T)W)$. Then the PE

$$\Delta(x) = P(1, w)^\top \beta - P(0, w)^\top \beta$$

is the difference between the expected log-wage of a woman and a man, conditional on a specific value w of the characteristics W . \square

Example 3 (Linear model with non-additive error, or QR model). Let Y be log-wage, T be an the indicator of female worker, and W be a vector of worker characteristics as in the previous example. Suppose we model the conditional quantile function of log-wages using the linear interactive model:

$$Y = g(X) = P(T, W)^\top \beta(\epsilon), \quad \epsilon \mid T, W \sim U(0, 1), \quad X = (T, W, \epsilon),$$

where $P(T, W)^\top \beta(\tau)$ is the conditional τ -quantile of Y given T and W . Thus the covariate vector $X = (T, W, \epsilon)$ includes the observed covariates (T, W) as well as the rank variable ϵ , which is an unobserved factor (e.g. “ability”). Here $P(T, W)$ is a collection of transformations of T and W , e.g., $P(T, W) = (TW, (1 - T)W)$. Then the PE

$$\Delta(x) = P(1, w)^\top \beta(\tau) - P(0, w)^\top \beta(\tau), \quad x = (t, w, \tau),$$

is the difference between the conditional τ -quantile of log-wage for a woman and a man, conditional on a specific value w of the characteristics W .

Note that in this case,

$$X \sim \mu, \quad \mu = F_{T,W} \times F_\epsilon,$$

where F_ϵ is the distribution function of the standard uniform random variable, and $F_{T,W}$ is the distribution of (T, W) . For estimation purposes, we will have to exclude the tail quantile indices, and so μ would be redefined to have support on a set of the form $\mathcal{X} \times [\ell, 1 - \ell]$, where $0 < \ell < 0.5$ is a small positive number. \square

2.2. The Sorted Effects. In Examples 1–3, the PE $\Delta(x)$ is a function of x and therefore can be different for each observational unit. To summarize this effect in a single measure, a common practice in empirical economics is to average the PE

$$\mathbb{E}_\mu[\Delta(X)] = \int \Delta(x) d\mu(x),$$

where μ is the distribution of X on the part of the population of interest. For example, when μ is the distribution on the entire population we obtain the average predictive effect (APE); whereas if μ is the distribution on a group characterized by X taking values in some specified

set, we obtain a conditional average predictive effect. Averaging, however, masks most the heterogeneity in the PE allowed by nonlinear or interactive linear models.

We propose reporting the entire set of values of the PE sorted in increasing order and indexed by a ranking $u \in [0, 1]$ with respect to the population of interest. These sorted effects provide a more complete representation of the heterogeneity in the PE than the average effects.

Definition 2.1 (*u-SPE*). The *u-sorted predictive effect* with respect to μ is

$$\Delta_{\mu}^*(u) := u - \text{quantile of } \Delta(X), \quad X \sim \mu.$$

The *u-SPE* is the u^{th} -quantile of $\Delta(X)$ when X is distributed according to μ . As for the average effect, μ can be chosen to select a target subpopulation of the entire population. For example, if T is a treatment indicator:

- the *u-SPE* corresponds to the u -quantile of the PE in the entire population when μ is set to the distribution of X in the entire population; we call this *population u-SPE*;
- the *u-SPE* corresponds to the u -quantile of the PE in the subpopulation of “treated” when μ is set to the distribution of X conditional on $T = 1$; we call this *u-SPE on the treated*;

By considering $\Delta_{\mu}^*(u)$ at multiple indexes u , we obtain a one-dimensional representation of the heterogeneity of the PE. Accordingly, our object of interest is the SPE-function

$$\{u \mapsto \Delta_{\mu}^*(u) : u \in \mathcal{U}\}, \quad \mathcal{U} \subseteq [0, 1],$$

where \mathcal{U} is the set of indices of interest. For example, as we previewed in the introduction, we shall find substantial heterogeneity in the SPE-function of gender on wages in Section 4, which was completely missed by traditional empirical analyses that only report average effects.

3. ESTIMATION AND INFERENCE METHODS

3.1. Empirical SPE. In practice, we replace the PE Δ and the distribution μ by sample analogs to construct plug-in estimators of the SPE. Let $\widehat{\Delta}(x)$ and $\widehat{\mu}(x)$ be estimators of $\Delta(x)$ and $\mu(x)$ obtained from a sample of size n . The estimator of Δ_{μ}^* is

$$\widehat{\Delta}_{\mu}^*(u) := \widehat{\Delta}_{\widehat{\mu}}^*(u) = \inf_{\delta \in \mathbb{R}} \{F_{\widehat{\Delta}, \widehat{\mu}}(\delta) \geq u\},$$

where $F_{\widehat{\Delta}, \widehat{\mu}}(\delta) = \mathbb{E}_{\widehat{\mu}}[1\{\widehat{\Delta}(X) \leq \delta\}] =: \widehat{F}_{\widehat{\Delta}, \widehat{\mu}}(\delta)$. Then the empirical SPE-function is

$$\{u \mapsto \widehat{\Delta}_{\mu}^*(u) : u \in \mathcal{U}\}, \quad \mathcal{U} \subseteq [0, 1],$$

where \mathcal{U} is the set of indices of interest that typically excludes tail indexes and satisfies other technical conditions stated in Section 6.

Example 1 (Binary response model, cont.) Given $\{(Y_i, X_i) : 1 \leq i \leq n\}$, an i.i.d. sample of (Y, X) , the estimator of the PE is

$$\widehat{\Delta}(x) = F\left(P(1, w)^\top \widehat{\beta}\right) - F\left(P(0, w)^\top \widehat{\beta}\right), \quad (3.4)$$

where $\widehat{\beta}$ is the maximum likelihood estimator (MLE) of β ,

$$\widehat{\beta} \in \arg \max_{b \in \mathbb{R}^{d_p}} \sum_{i=1}^n [Y_i \log F(P(X_i)^\top b) + (1 - Y_i) \log \{1 - F(P(X_i)^\top b)\}],$$

for $d_p = \dim P(X)$.

Example 2 (Interactive linear model with additive error, cont.) Given $\{(Y_i, T_i, W_i) : 1 \leq i \leq n\}$, an i.i.d. sample of (Y, T, W) , the estimator of the PE is

$$\widehat{\Delta}(x) = P(1, w)^\top \widehat{\beta} - P(0, w)^\top \widehat{\beta}, \quad (3.5)$$

where $\widehat{\beta}$ is the ordinary least squares (OLS) estimator of β ,

$$\widehat{\beta} \in \arg \min_{b \in \mathbb{R}^{d_p}} \sum_{i=1}^n [Y_i - P(T_i, W_i)^\top b]^2,$$

for $d_p = \dim P(T, W)$. □

Example 3 (Linear model with non-additive error, cont.) Given $\{(Y_i, T_i, W_i) : 1 \leq i \leq n\}$, an i.i.d. sample of (Y, T, W) , the estimator of the PE is

$$\widehat{\Delta}(x) = P(1, w)^\top \widehat{\beta}(\tau) - P(0, w)^\top \widehat{\beta}(\tau), \quad (3.6)$$

where $\widehat{\beta}(\tau)$ is the Koenker and Basset (1978) quantile regression (QR) estimator of $\beta(\tau)$,

$$\widehat{\beta}(\tau) \in \arg \min_{b \in \mathbb{R}^{d_p}} \sum_{i=1}^n \rho_\tau(Y_i - P(T_i, W_i)^\top b),$$

for $d_p = \dim P(T, W)$, and $\rho_\tau(v) = (\tau - 1\{v < 0\})v$. □

Remark 3.1 (Estimation of μ). The distribution μ can be the distribution of X in the entire population or some subpopulation of interest. In either case we can estimate μ by the corresponding empirical distribution. Let S denote the indicator of an observation unit belonging to the subpopulation of interest. For example, if $S = T$, then $S = 1$ indicates the unit is in the subpopulation of the treated; $S = 0$ indicates the unit is in the subpopulation of the untreated. The indicator S can also incorporate other restrictions, for example $S = 1\{X \in \mathcal{X}\}$ restricts

the support of covariate X to the region \mathcal{X} . Finally, if S is always 1, then this means that we work with the entire population. Estimation of μ can be done using the empirical distribution:

$$\widehat{\mu}(x) = \sum_{i=1}^n S_i 1\{X_i \leq x\} / \sum_{i=1}^n S_i.$$

An alternative would be to use the smoothed empirical distribution.

If μ can be decomposed into known and unknown parts, then we only need to estimate the unknown parts. Thus, $\mu = F_{T,W} \times F_\epsilon$ in Example 3, where F_ϵ is the uniform distribution and $F_{T,W}$ can be estimated by the empirical distribution of (T, W) in the part of the population of interest. \square

The set of examples listed above are the most basic, leading cases, arising mostly in predictive analysis and program evaluation. Our theoretical results are rather general and are not limited to these cases. For example, our theoretical results allow for both Δ and μ to arise from structural models, as well as to be estimated by structural methods.

3.2. Inference for SPE. The main inferential result established in this paper can be previewed as follows:

Proposition 1 (Limit Theory and Bootstrap for Empirical SPE). *Under the regularity conditions specified in Theorem 3, stated in Section 6, the empirical SPE-process converges in distribution, namely*

$$\sqrt{n}(\widehat{\Delta}_\mu^*(u) - \Delta_\mu^*(u)) \rightsquigarrow Z_\infty(u), \quad (3.7)$$

as a stochastic process indexed by $u \in \mathcal{U}$ in $\ell^\infty(\mathcal{U})$, the metric space of bounded functions. The limit $Z_\infty(u)$ is a centered Gaussian process, defined in equation (6.22) of Theorem 2, and the set \mathcal{U} is a compact subset of $(0, 1)$ that obeys the conditions stated in Lemma 2. Moreover, the weighted bootstrap algorithm specified in Algorithm 1 estimates consistently the law of $Z_\infty(u)$.

We can construct asymptotically valid confidence intervals for the SPE using this result. The next corollary to Proposition 1 provides uniform bands that cover the SPE-function simultaneously over a region of values of u with prespecified probability in large samples. It does cover pointwise confidence bands for the SPE-function at a specific quantile index u as a special case by simply taking \mathcal{U} to be the singleton set $\{u\}$.

Corollary 1 (Inference on SPE-function using Limit Theory and Bootstrap). *Under the assumptions of Theorem 2, for any $0 < \alpha < 1$,*

$$\mathbb{P} \left\{ \Delta_\mu^*(u) \in \left[\widehat{\Delta}_\mu^*(u) - \widehat{t}_{1-\alpha}(\mathcal{U}) \widehat{\Sigma}(u)^{1/2} / \sqrt{n}, \widehat{\Delta}_\mu^*(u) + \widehat{t}_{1-\alpha}(\mathcal{U}) \widehat{\Sigma}(u)^{1/2} / \sqrt{n} \right] : u \in \mathcal{U} \right\} \rightarrow 1 - \alpha,$$

where $\hat{t}_{1-\alpha}(\mathcal{U})$ is any consistent estimator of $t_{1-\alpha}(\mathcal{U})$, the $(1 - \alpha)$ -quantile of

$$t(\mathcal{U}) := \sup_{u \in \mathcal{U}} |Z_\infty(u)| \Sigma(u)^{-1/2}$$

and $u \mapsto \hat{\Sigma}(u)$ is a uniformly consistent estimator of $u \mapsto \Sigma(u)$, the variance function of $u \mapsto Z_\infty(u)$. We provide consistent estimators of $t_{1-\alpha}(\mathcal{U})$ and $u \mapsto \Sigma(u)$ in Algorithm 1.

We now describe a practical bootstrap algorithm to obtain the quantiles of $t(\mathcal{U})$. Let $(\omega_1, \dots, \omega_n)$ denote the bootstrap weights, which are nonnegative random variables independent from the data, and which obey the conditions stated in van der Vaart and Wellner (1996). For example, $(\omega_1, \dots, \omega_n)$ is a multinomial vector with dimension n and probabilities $(1/n, \dots, 1/n)$ in the empirical bootstrap. In what follows B is the number of bootstrap draws, such that $B \rightarrow \infty$. In our experience, setting $B \geq 500$ suffices for good accuracy.

Algorithm 1 (Bootstrap law of $t(\mathcal{U})$ and its quantiles). **1)** Draw a realization of the bootstrap weights $(\omega_1, \dots, \omega_n)$. **2)** For each $u \in \mathcal{U}$, compute $\tilde{\Delta}_\mu^*(u) = \tilde{\Delta}_\mu^*(u)$, a bootstrap draw of $\hat{\Delta}_\mu^*(u) = \hat{\Delta}_\mu^*(u)$, where $\tilde{\Delta}$ and $\tilde{\mu}$ are the bootstrap versions of $\hat{\Delta}$ and $\hat{\mu}$ that use $(\omega_1, \dots, \omega_n)$ as sampling weights in the computation of the estimators. Construct a bootstrap draw of $Z_\infty(u)$ as $\tilde{Z}_\infty(u) = \sqrt{n}(\tilde{\Delta}_\mu^*(u) - \hat{\Delta}_\mu^*(u))$. **3)** Repeat steps (1)-(2) B times with $B \rightarrow \infty$. **4)** For each $u \in \mathcal{U}$, compute a bootstrap estimator of $\Sigma(u)^{1/2}$ such as the bootstrap interquartile range rescaled with the normal distribution $\hat{\Sigma}(u)^{1/2} = (q_{0.75}(u) - q_{0.25}(u))/(z_{0.75} - z_{0.25})$, where $q_p(u)$ is the p th sample quantile of $\tilde{Z}_\infty(u)$ in the B draws and z_p is the p th quantile of $N(0, 1)$. **5)** Use the empirical distribution of $\tilde{t}(\mathcal{U}) = \sup_{u \in \mathcal{U}} |\tilde{Z}_\infty(u)| \hat{\Sigma}(u)^{-1/2}$ across the B draws to approximate the distribution of $t(\mathcal{U}) = \sup_{u \in \mathcal{U}} |Z_\infty(u)| \Sigma(u)^{-1/2}$. In particular, construct $\hat{t}_{1-\alpha}(\mathcal{U})$, an estimator $t_{1-\alpha}(\mathcal{U})$, as the $(1 - \alpha)$ -quantile of the B draws of $\tilde{t}(\mathcal{U})$.

4. SORTED EFFECTS IN EMPIRICAL ACTION

4.1. Effect of Race on Mortgage Denials. We use data of mortgage applications in Boston for the year 1990 collected by the Federal Reserve Bank of Boston in relation to the Home Mortgage Disclosure Act (HMDA) to study racial discrimination in the bank decisions of mortgage denials, see Munnell, Tootell, Browne, and McEneaney (1996). The HMDA was passed to monitor minority access to the mortgage market. Providing better access to credit markets can arguably help the disadvantaged groups escape poverty traps. Following Stock and Watson (2011, Chap 11), we focus on white and black applicants for single-family residences. The sample comprises 2,380 observations corresponding to 2,041 white applicants and 339 black applicants.

We estimate a binary response model where the outcome variable Y is an indicator of mortgage denial, the key covariate T is an indicator for the applicant being black, and the controls W contain financial and other characteristics of the applicant that the banks take into account in the mortgage decisions. These include the monthly debt to income ratio; monthly housing expenses to income ratio; a categorical variable for “bad” consumer credit score with 6 categories: 1 if no slow payments or delinquencies, 2 if one or two slow payments or delinquencies, 3 if more than two slow payments or delinquencies, 4 if insufficient credit history for determination, 5 if delinquent credit history with payments 60 days overdue, and 6 if delinquent credit history with payments 90 days overdue; a categorical variable for “bad” mortgage credit score with 4 categories: 1 if no late mortgage payments, 2 if no mortgage payment history, 3 if one or two late mortgage payments, and 4 if more than two late mortgage payments; an indicator for public record of credit problems including bankruptcy, charge-offs, and collective actions; an indicator for denial of application for mortgage insurance; two indicators for medium and high loan to property value ratio, where medium is between .80 and .95 and high is above .95; an three indicators for self-employed, single, and high school graduated.

TABLE 1. Descriptive Statistics of Mortgage Applicants

	All	Black	White
deny	0.12	0.28	0.09
black	0.14	1.00	0.00
payment-to-income ratio	0.33	0.35	0.33
expenses-to-income ratio	0.26	0.27	0.25
bad consumer credit	2.12	3.02	1.97
bad mortgage credit	1.72	1.88	1.69
credit problems	0.07	0.18	0.06
denied mortgage insurance	0.02	0.05	0.02
medium loan-to-value ratio	0.37	0.56	0.34
high loan-to-value ratio	0.03	0.07	0.03
self-employed	0.12	0.07	0.12
single	0.39	0.52	0.37
high school graduated	0.98	0.97	0.99
number of observations	2,380	339	2,041

Table 1 reports the sample means of the variables used in the analysis. The probability of having the mortgage denied is 19% higher for black applicants than for white applicants.

However, black applicants are more likely to have socio-economic characteristics linked to a denial of the mortgage, as Table 1 shows.

Figure 1 of Section 1 plots estimates and 90% confidence sets of the population APE and SPE-function of being black. The PEs are obtained as described in Example 1 using a logit model with $P(X) = X = (T, W)$ and $\hat{\mu}$ equal to the empirical distribution of X in the whole sample. The confidence bands are constructed using Algorithm 1 with multinomial weights (empirical bootstrap) and $B = 500$, and are uniform for the SPE-function over the grid $\mathcal{U} = \{.02, .03, \dots, .98\}$. We monotone the bands using the rearrangement method of Chernozhukov, Fernández-Val, and Galichon (2009). After controlling for applicant characteristics, black applicants are still on average 5.3% more likely to have the mortgage denied than white applicants. Moreover, the SPE-function shows significant heterogeneity, with the PE ranging between 0 and 15%. Thus, there exists a subgroup of applicants that is 15% more likely to be denied a mortgage if they were black, and there is a subgroup of applicants that is not affected by racial discrimination. Table 2 shows the results of the classification analysis, answering the question “who gets affected the most and who the least?” The table shows that the 5% of the applicants *most affected* by racial discrimination are *more likely* to have either of the following characteristics relative to the 5% of the *least affected* applicants: self employed, single, black, high debt to income, high expense to income, high loan to value ratios, medium or high loan-to-income ratio, bad consumer or credit scores, and credit problems.

4.2. Gender Wage Gap in 2012. We next consider the gender wage gap using data from the U.S. March Supplement of the Current Population Survey (CPS) in 2012. We select white non-hispanic individuals, aged 25 to 64 years, and working more than 35 hours per week during at least 50 weeks of the year. We exclude self-employed workers; individuals living in group quarters; individuals in the military, agricultural or private household sectors; individuals with inconsistent reports on earnings and employment status; and individuals with allocated or missing information in any of the variables used in the analysis. The resulting sample consists of 29,217 workers including 16,690 men and 12,527 of women.

We estimate interactive linear models with additive and non-additive errors, using mean and quantile regressions, respectively. The outcome variable Y is the logarithm of the hourly wage rate constructed as the ratio of the annual earnings to the total number of hours worked, which is constructed in turn as the product of number of weeks worked and the usual number of hours worked per week. The key covariate T is an indicator for female worker, and the control variables W include 5 marital status indicators for widowed, divorced, separated, never married, and married; 6 educational attainment indicators for 0-8 years of schooling completed,

TABLE 2. **Who gets affected the most and who the least?** Classification Analysis. Averages of Characteristics of the Mortgage Applicants Least and Most Affected by Racial Discrimination

Characteristics of the Group	5% Most Affected	5% Least Affected
	PE > .14	PE < .014
deny	0.54	0.15
black	0.41	0.09
debt-to-income ratio	0.40	0.24
expense-to-income ratio	0.29	0.20
consumer credit score	4.85	1.49
mortgage credit score	1.99	1.33
credit problem	0.64	0.10
denied mortgage insurance	0.00	0.10
medium loan-to-house ratio	0.60	0.08
high loan-to- house value	0.10	0.03
self employed	0.18	0.08
single	0.56	0.13
high school graduate	0.92	0.99

high school dropouts, high school graduates, some college, college graduate, and advanced degree; 4 region indicators for midwest, south, west, and northeast; and a quartic in potential experience constructed as the maximum of age minus years of schooling minus 7 and zero, i.e., $experience = \max(age - education - 7, 0)$, interacted with the educational attainment indicators.³ All the calculations use the CPS sampling weights to account for nonrandom sampling in the March CPS.

Table 3 reports sample means for the variables used in the analysis. Working women are more highly educated than working men, have slightly less potential experience, and are less likely to be married and more likely to be divorced or widowed. The unconditional gender wage gap is 25%.

Figure 2 of Section 1 and Figure 3 plot estimates and 90% confidence sets of the APE and SPE-function on the treated of the conditional gender wage gap using additive and non-additive error models, respectively. The PEs are obtained as described in Examples 2 and 3 with $P(T, W) = (TW, (1 - T)W)$. The distribution $F_{T,W}$ is estimated by the empirical

³The sample selection criteria and the variable construction follow Mulligan and Rubinstein (2008).

TABLE 3. Descriptive Statistics of Workers

	All	Men	Women
log wage	2.79	2.90	2.65
female	0.43	0.00	1.00
married	0.66	0.69	0.63
widowed	0.01	0.00	0.02
divorced	0.12	0.10	0.15
separated	0.02	0.02	0.02
never married	0.19	0.19	0.18
0-8 years completed	0.00	0.01	0.00
high school dropout	0.02	0.03	0.02
high school graduated	0.25	0.27	0.23
some college	0.28	0.27	0.30
college graduated	0.28	0.28	0.29
advanced degree	0.15	0.14	0.17
northeast	0.20	0.20	0.19
midwest	0.27	0.27	0.28
south	0.35	0.35	0.35
west	0.18	0.19	0.18
potential experience	18.96	19.01	18.90
number of observations	29,217	16,690	12,527

Source: March Supplement CPS 2012

distribution of (T, W) for women, and F_ϵ is approximated by a uniform distribution over the grid $\{.02, .03, \dots, .98\}$. The confidence bands are constructed using Algorithm 1 with multinomial weights (empirical bootstrap) and $B = 500$, and are uniform for the SPE-function over the grid $\mathcal{U} = \{.01, .02, \dots, .98\}$. We monotone the bands using the rearrangement method of Chernozhukov, Fernández-Val, and Galichon (2009). After controlling for worker characteristics, the gender wage gap for women remains on average around 26% in both models. More importantly, we clearly reject the hypothesis that the SPE-function is constant and uncover a striking amount of heterogeneity, with the PE ranging between 0 and 43% in the additive error model and between -5 and 51% in the non-additive error model.

Table 4 shows the results of a classification analysis, exhibiting characteristics of women that are most and least affected by gender discrimination. We focus here on the non-additive

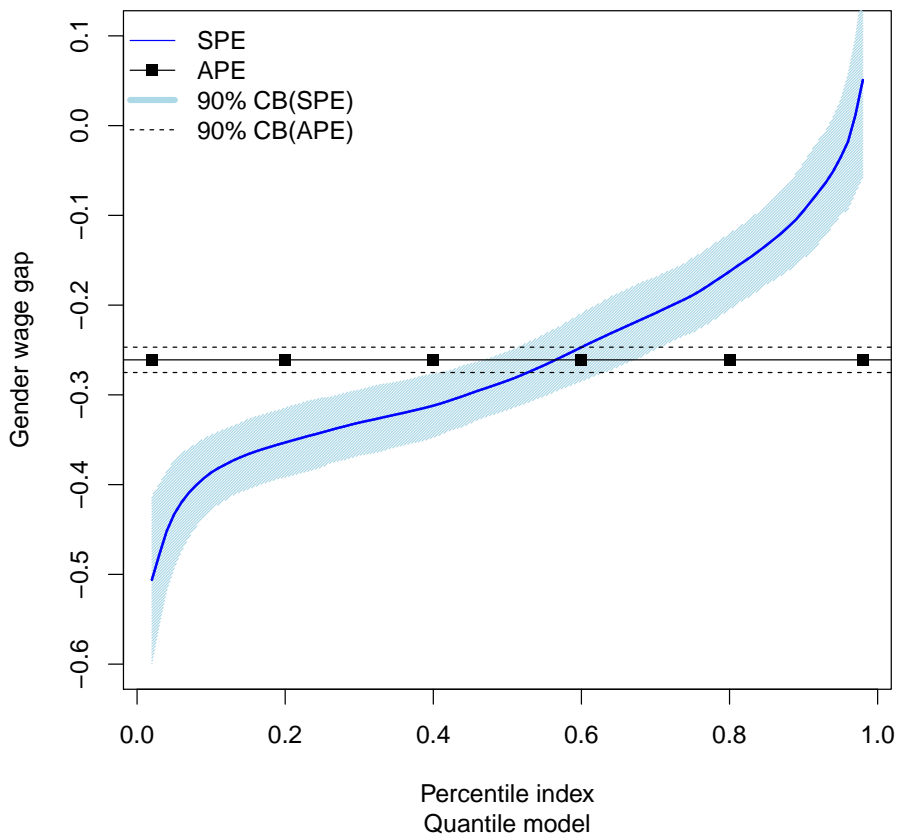


FIGURE 3. APE and SPE of the gender wage gap for women. Estimates and 90% bootstrap uniform confidence bands based on a linear model with interactions for the conditional quantile function are shown.

model, but the results from the additive model are similar. According to this model the 5% of the women *most affected* by gender discrimination earn higher wages, are much more likely to be married, have either very low or very high education, and possess much more potential experience than the 5% least affected women.

We explore further these findings by analyzing the APE and SPE on the treated *conditionally* on marital status and potential experience. Here we show estimates and 90% confidence bands of the APE and SPE-function of the gender wage gap based on the non-additive model for 3 subpopulations defined by marital status: never married, married and divorced women; and 3 subpopulations defined by experience: low, medium and high experienced women, where the experience cutoffs are 11 and 26, the first and third sample quartiles of potential experience for

women. The confidence bands are constructed as in Fig. 3. We find significant heterogeneity in the gender gap within each subpopulation, and also between subpopulations defined by marital status and experience. The SPE-function is much more heterogeneous for women with low experience and never married. Married and high experienced women suffer from the highest gender wage gaps. This pattern is consistent with preferences where single young women are more career-oriented.⁴

TABLE 4. **Who gets affected the most and the least?** Classification Analysis – Averages of Characteristics of the Women Least and Most Affected by Gender Discrimination

Characteristics of the Group	5% Most Affected	5% Least Affected
	PE < $-.43$	PE > $-.05$
log wage	2.73	2.66
female	1.00	1.00
married	0.86	0.11
widowed	0.06	0.03
divorced	0.04	0.08
separated	0.01	0.01
never married	0.03	0.77
0-8 years completed	0.02	0.01
high school dropout	0.11	0.02
high school graduated	0.19	0.03
some college	0.08	0.17
college graduated	0.11	0.41
advanced degree	0.48	0.35
northeast	0.22	0.27
midwest	0.30	0.26
south	0.29	0.29
west	0.19	0.17
potential experience	24.35	7.85

PE estimated from a linear conditional quantile model with interactions

⁴We find similar results using the additive error model. We do not report these results for the sake of brevity.

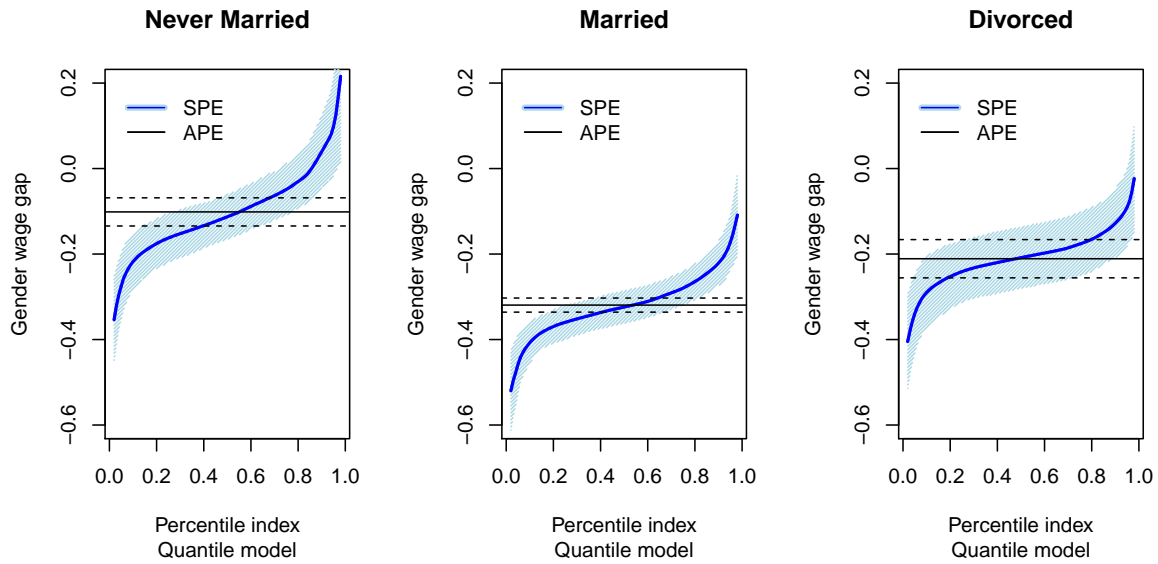


FIGURE 4. APE and SPE of the gender wage gap for women by marital status. Estimates and 90% bootstrap uniform confidence bands based on a linear model with interactions for the conditional quantile function are shown.

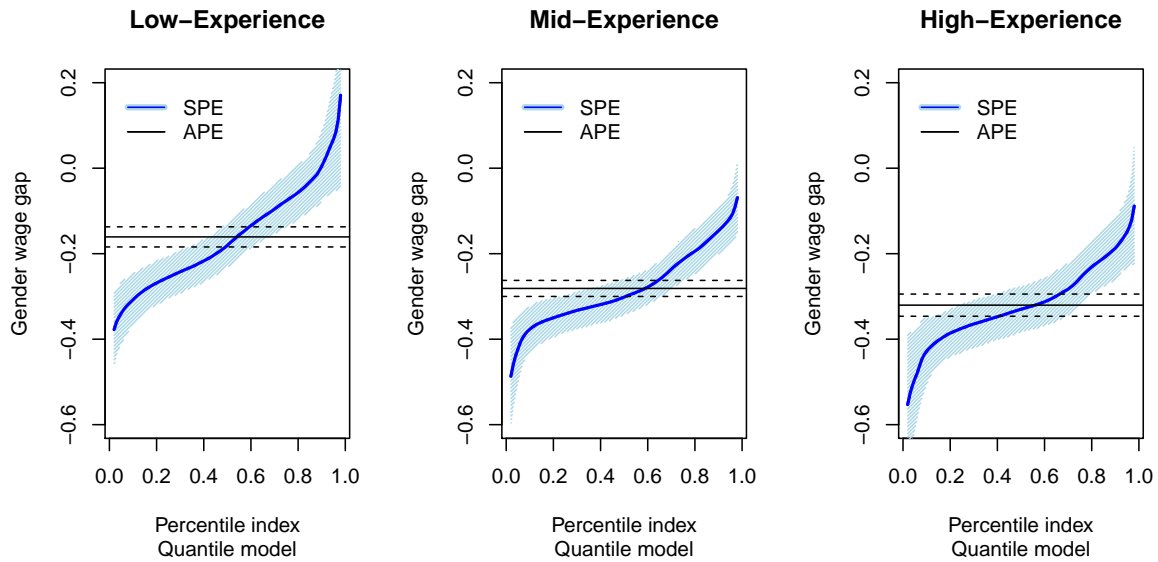


FIGURE 5. APE and SPE of the gender wage gap for women by experience level. Estimates and 90% bootstrap uniform confidence bands based on a linear model with interactions for the conditional quantile function are shown.

5. THEORETICAL ANALYSIS I: BASIC ANALYTICAL PROPERTIES OF THE SORTING
OPERATOR ACTING ON MULTIVARIATE FUNCTIONS

To analyze the analytical properties of the SPE-function, it is convenient to treat the PE as a multivariate real-valued function

$$\Delta : B(\mathcal{X}) \rightarrow \mathbb{R},$$

where $B(\mathcal{X}) \subseteq \mathbb{R}^{d_x}$ contains the set \mathcal{X} . Let μ be a distribution function over \mathcal{X} . The distribution of Δ with respect to μ is the function $F_{\Delta, \mu} : \mathbb{R} \rightarrow [0, 1]$ with

$$F_{\Delta, \mu}(\delta) = \int_{\mathcal{X}} 1\{\Delta(x) \leq \delta\} d\mu(x). \quad (5.8)$$

The SPE-function is the map

$$\Delta_{\mu}^* : \mathcal{U} \subseteq [0, 1] \rightarrow \mathbb{R},$$

defined at each point as the left-inverse function of $F_{\Delta, \mu}$, i.e.,

$$\Delta_{\mu}^*(u) := F_{\Delta, \mu}^{\leftarrow}(u) := \inf_{\delta \in \mathbb{R}} \{F_{\Delta, \mu}(\delta) \geq u\}. \quad (5.9)$$

From this functional perspective, the map $u \mapsto \Delta_{\mu}^*(u)$ is the result of applying a sorting operator to the map $x \mapsto \Delta(x)$, where the operator sorts values in increasing order weighted by μ . In this section we do the following:

- 1) characterize some analytical properties of the distribution function $\delta \mapsto F_{\Delta, \mu}(\delta)$ and the sorted function $u \mapsto \Delta_{\mu}^*(u)$, and
- 2) derive the functional derivatives of $F_{\Delta, \mu}$ and Δ_{μ}^* with respect to Δ and μ .

5.1. Background on Differential Geometry. We recall some definitions from differential geometry that are used in the analysis. For a continuously differentiable function $\Delta : B(\mathcal{X}) \rightarrow \mathbb{R}$ defined on an open set $B(\mathcal{X}) \subseteq \mathbb{R}^{d_x}$ containing the set \mathcal{X} , $x \in \mathcal{X}$ is a *critical point* of Δ on \mathcal{X} , if

$$\partial\Delta(x) = 0, \quad (5.10)$$

where $\partial\Delta(x)$ is the gradient of $\Delta(x)$; otherwise x is a *regular point* of Δ on \mathcal{X} . A value δ is a *critical value* of Δ on \mathcal{X} , if the set $\{x \in \mathcal{X} : \Delta(x) = \delta\}$ contains at least one critical point. Otherwise δ is a *regular value* of Δ on \mathcal{X} .

In the multi-dimensional space, $d_x > 1$, a function Δ can have continuums of critical points. For example, the function $\Delta(x_1, x_2) = \cos(x_1^2 + x_2^2)$ has continuums of critical points on the circles $x_1^2 + x_2^2 = k\pi$ for each positive integer k .

We recall now several core concepts related to manifolds from Spivak (1965) and Munkres (1991).

Definition 5.1 (Manifold). Let d_k , d_x and r be positive integers such that $d_x \geq d_k$. Suppose that \mathcal{M} is a subspace of \mathbb{R}^{d_x} that satisfies the following property: for each point $m \in \mathcal{M}$, there is a set \mathcal{V} containing m that is open in \mathcal{M} , a set \mathcal{K} that is open in \mathbb{R}^{d_k} , and a continuous map $\alpha_m : \mathcal{K} \rightarrow \mathcal{V}$ carrying \mathcal{K} onto \mathcal{V} in a one-to-one fashion, such that: (1) α_m is of class \mathcal{C}^r on \mathcal{K} , (2) $\alpha_m^{-1} : \mathcal{V} \rightarrow \mathcal{K}$ is continuous, (3) the Jacobian matrix of α_m , $D\alpha_m(k)$, has rank d_k for each $k \in \mathcal{K}$. Then \mathcal{M} is called a d_k -manifold without boundary in \mathbb{R}^{d_x} of class \mathcal{C}^r . The map α_m is called a *coordinate patch* on \mathcal{M} about m . A set of coordinate patches that covers \mathcal{M} is called an *atlas*.

Definition 5.2 (Connected Branch). For any subset \mathcal{M} of a topological space, if any two points m_1 and m_2 can not be connected via path in \mathcal{M} , then we say that m_1 and m_2 are not connected. Otherwise, we say that m_1 and m_2 are connected. We say that $\mathcal{V} \subseteq \mathcal{M}$ is a *connected branch* of \mathcal{M} if all points of \mathcal{V} are connected to each other and do not connect to any points in $\mathcal{M} \setminus \mathcal{V}$.

Definition 5.3 (Volume). For a $d_x \times d_k$ matrix $A = (x_1, x_2, \dots, x_{d_k})$ with $x_i \in \mathbb{R}^{d_x}$, $1 \leq i \leq d_k \leq d_x$, let $\text{Vol}(A) = \sqrt{\det(A^T A)}$, which is the *volume* of the parallelepiped $P(A)$ with edges given by the columns of A , $P(A) = \{c_1 x_1 + \dots + c_{d_k} x_{d_k} : 0 \leq c_i \leq 1, i = 1, \dots, d_k\}$.

The volume measures the amount of mass in \mathbb{R}^{d_k} of a d_k -dimensional parallelepiped in \mathbb{R}^{d_x} , $d_k \leq d_x$. This concept is essential for integration on manifolds, which we recall here. First we recall the integration over parameterized manifold:

Definition 5.4 (Integration on a parametrized manifold). Let \mathcal{K} be open in \mathbb{R}^{d_k} , and let $\alpha : \mathcal{K} \rightarrow \mathbb{R}^{d_x}$ be of class \mathcal{C}^r on \mathcal{K} , $r \geq 1$. The set $\mathcal{M} = \alpha(\mathcal{K})$ together with the map α constitute a *parametrized d_k -manifold* in \mathbb{R}^{d_x} of class \mathcal{C}^r . Let g be a real-valued continuous function defined at each point of \mathcal{M} . The *integral of g over \mathcal{M} with respect to volume* is defined by

$$\int_{\mathcal{M}} g(m) d\text{Vol} := \int_{\mathcal{K}} (g \circ \alpha)(k) \text{Vol}(D\alpha(k)) dk, \quad (5.11)$$

provided that the right side integral exists. Here $D\alpha(k)$ is the Jacobian matrix of the mapping $k \mapsto \alpha(k)$, and $\text{Vol}(D\alpha(k))$ is the volume of matrix $D\alpha(k)$ as defined in Definition 5.3.

The above definition coincides with the usual interpretation of integration. The integral can be extended to manifolds that do not admit a global parametrization α using the notion of partition of unity. This partition is a set of smooth local functions defined in a neighborhood

of the manifold. The following Lemma shows the existence of the partition of unity and is proven in Lemma 25.2 in Munkres (1991).

Lemma 1 (Partition of Unity on \mathcal{M} of class \mathcal{C}^∞). *Let \mathcal{M} be a d_k -manifold without boundary in \mathbb{R}^{d_x} of class \mathcal{C}^r , $r \geq 1$, and let ϑ be an open cover of \mathcal{M} . Then, there is a collection $\mathcal{P} = \{p_i \in \mathcal{C}^\infty : i \in \mathcal{I}\}$, where p_i is defined on an open set containing \mathcal{M} for all $i \in \mathcal{I}$, with the following properties: (1) For each $m \in \mathcal{M}$ and $i \in \mathcal{I}$, $0 \leq p_i(m) \leq 1$. (2) For each $m \in \mathcal{M}$ there is an open set $\mathcal{V} \in \vartheta$ containing m such that all but finitely many $p_i \in \mathcal{P}$ are 0 on \mathcal{V} . (3) For each $m \in \mathcal{M}$, $\sum_{p_i \in \mathcal{P}} p_i(m) = 1$. (4) For each $p_i \in \mathcal{P}$ there is an open set $\mathcal{U} \in \vartheta$, such that $\text{supp}(p_i) \subseteq \mathcal{U}$.*

Now we are ready to recall the definition of integration on a manifold.

Definition 5.5 (Integration on a manifold with partition of unity). Let $\vartheta := \{\vartheta_j : j \in \mathcal{J}\}$ be an open cover of a d_k -manifold without boundary \mathcal{M} in \mathbb{R}^{d_x} of class \mathcal{C}^r , $r \geq 1$. Suppose there is an coordinate patch $\alpha_j : \mathcal{V}_j \subseteq \mathbb{R}^{d_k} \rightarrow \vartheta_j$, one-to-one and of class \mathcal{C}^r on \mathcal{V}_j for each $j \in \mathcal{J}$. Denote $\mathcal{K}_j = \alpha_j^{-1}(\mathcal{M} \cap \vartheta_j)$. Then for a real-valued continuous function g defined on an open set that contains \mathcal{M} , the *integral of g over \mathcal{M} with respect to volume* is defined by:

$$\int_{\mathcal{M}} g(m) d\text{Vol} := \sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}} \int_{\mathcal{K}_j} [(p_i g) \circ \alpha_j](k) \text{Vol}(D\alpha_j(k)) dk, \quad (5.12)$$

where $\{p_i \in \mathcal{C}^\infty : i \in \mathcal{I}\}$ is a partition of unity on \mathcal{M} of class \mathcal{C}^∞ that satisfies the conditions of Lemma 1. Munkres (1991, p. 212) shows that the integral is well defined and does not depend on the choice of cover and partition of unity.

5.2. Basic Analytical Properties of Sorted Functions. Recall that the main functions in the analysis are the PE function $x \mapsto \Delta(x)$ and the distribution function $x \mapsto \mu(x)$. We make the following technical assumptions about these functions:

S.1. The part of the domain of $x \mapsto \Delta(x)$ of interest, \mathcal{X} , is open and its closure $\overline{\mathcal{X}}$ is compact. The distribution μ is absolutely continuous with respect to the Lebesgue measure with density μ' . There exists an open set $B(\mathcal{X})$ containing $\overline{\mathcal{X}}$ such that $x \mapsto \Delta(x)$ is \mathcal{C}^1 on $B(\mathcal{X})$, and $x \mapsto \mu'(x)$ is continuous on $B(\mathcal{X})$ and is zero outside the domain, i.e. $\mu'(x) = 0$ for any $x \in B(\mathcal{X}) \setminus \mathcal{X}$.

S.2. Let $\mathcal{M}_\Delta(\delta) := \{x \in \mathcal{X} : \Delta(x) = \delta\}$. For any regular value δ of Δ on $\overline{\mathcal{X}}$, we assume that the closure of $\mathcal{M}_\Delta(\delta)$, $\overline{\mathcal{M}_\Delta(\delta)}$, has a finite number of connected branches.

Remark 5.1 (Continuous X). Condition S.1 requires that all the components of the covariate X are continuous random variables. We defer the treatment of the case where X has both continuous and discrete components to Appendix A. \square

Remark 5.2 (Properties of $\mathcal{M}_\Delta(\delta)$). Lemma 8 in Appendix B shows that S.1 and S.2 imply that, for any δ that is a regular value of $x \mapsto \Delta(x)$ on $\overline{\mathcal{X}}$, $\mathcal{M}_\Delta(\delta)$ is a $(d_x - 1)$ -manifold without boundary in \mathbb{R}^{d_x} of class \mathcal{C}^1 . \square

The following lemma establishes the properties of the distribution function $\delta \mapsto F_{\Delta,\mu}(\delta)$ and the SPE-function $u \mapsto \Delta_\mu^*(u)$.

Define \mathcal{D} as a compact set consisting of regular values of $x \mapsto \Delta(x)$ on $\overline{\mathcal{X}}$.

Lemma 2 (Basic Properties of $F_{\Delta,\mu}$ and Δ_μ^*). *Under conditions S.1 and S.2:*

1. For any $\delta \in \mathcal{D}$, the derivative of $F_{\Delta,\mu}(\delta)$ with respect to δ is:

$$f_{\Delta,\mu}(\delta) := \partial_\delta F_{\Delta,\mu}(\delta) = \int_{\mathcal{M}_\Delta(\delta)} \frac{\mu'(x)}{\|\partial\Delta(x)\|} d\text{Vol}. \quad (5.13)$$

This integral is well-defined because the gradient $x \mapsto \partial\Delta(x)$ is finite, continuous and bounded away from 0 on $\mathcal{M}_\Delta(\delta) \subseteq \overline{\mathcal{X}}$. The map $\delta \mapsto f_{\Delta,\mu}(\delta)$ is uniformly continuous on \mathcal{D} .

2. Fix $\varepsilon > 0$, then for any

$$u \in \mathcal{U} := \{\tilde{u} \in [0, 1] : \Delta_\mu^*(\tilde{u}) \in \mathcal{D}, f_{\Delta,\mu}(\Delta_\mu^*(\tilde{u})) > \varepsilon\},$$

the derivative of $\Delta_\mu^*(u)$ respect to u is:

$$\partial_u \Delta_\mu^*(u) = \frac{1}{f_{\Delta,\mu}(\Delta_\mu^*(u))}. \quad (5.14)$$

Moreover, the derivative map $u \mapsto \partial_u \Delta_\mu^*(u)$ is uniformly continuous on \mathcal{U} .

Remark 5.3 (Properties of μ' at the boundary of $\overline{\mathcal{X}}$). S.1 imposes that the probability density function $x \mapsto \mu'(x)$ is continuous and vanishes at the boundary of $\overline{\mathcal{X}}$. To understand the importance of this condition, we consider the following example with dimension $d_x = 2$:

$$\Delta(x) = \sin(x_1^2 + x_2^2)$$

on $\mathcal{X} = \{(x_1, x_2) : x_1^2 + x_2^2 < \pi/6\}$, and $x \mapsto \mu'(x)$ is uniform on \mathcal{X} . It is easy to see that $\delta = 1/2$ is a regular value of Δ on $\overline{\mathcal{X}}$. However, $F_{\Delta,\mu}(\delta) = \int_{\mathcal{X}} 1\{\Delta(x) \leq \delta\} \mu'(x) dx$ is not differentiable at $\delta = 1/2$. The right derivative $\lim_{\eta \rightarrow 0^+} [F_{\Delta,\mu}(\delta + \eta) - F_{\Delta,\mu}(\delta)]/\eta = 0$, whereas the left derivative $\lim_{\delta \rightarrow 0^-} [F_{\Delta,\mu}(\delta + \eta) - F_{\Delta,\mu}(\delta)]/\eta = \sqrt{2\pi^3/3}$. Lemma 2 does not apply because $x \mapsto \mu'(x)$ is not continuous on any open set $B(\mathcal{X}) \supset \overline{\mathcal{X}}$. \square

Remark 5.4 (Relaxed Condition S.1). As a matter of generalization, our theoretical analysis allows us to replace that $x \mapsto \mu'(x)$ vanishes on $\partial\mathcal{X}$, as stated in assumption S.1 by the weaker condition

$$\int_{\mathcal{M}_{\Delta}(\delta)} 1\{x \in \partial\mathcal{X}\} \mu'(x) d\text{Vol} = 0, \quad (5.15)$$

where $\partial\mathcal{X}$ denotes the boundary of $\overline{\mathcal{X}}$. In the numerical examples of Section 6.4, we consider two designs where only this relaxed condition holds. \square

Remark 5.5 (No critical points in leading cases.). Lemma 2 states that $\delta \mapsto F_{\Delta, \mu}(\delta)$ ($u \mapsto \Delta_{\mu}^*(u)$) is \mathcal{C}^1 on any compact set \mathcal{D} (the Δ_{μ}^* pre-image of \mathcal{D}). It turns out we can set $\mathcal{D} = \Delta(\overline{\mathcal{X}}) := \{\Delta(x) : x \in \overline{\mathcal{X}}\}$ when the map $x \mapsto \Delta(x)$ does not have critical points on $\overline{\mathcal{X}}$. This case is nice, because it allows us not to worry about critical points when performing inference. Moreover, this case is practically important, as it occurs very naturally in many applications. For instance, it arises whenever $\Delta(x)$ is strictly monotonic in one of the components of x , say the first component x_1 when $x = (x_1, x_{-1})$. In this case, the derivative of $\delta \mapsto F_{\Delta, \mu}(\delta)$ can be expressed globally as the Riemann integral

$$f_{\Delta, \mu}(\delta) = \int_{\mathcal{X}_{-1}} f_{\Delta(X)|X_{-1}}(\delta | x_{-1}) \mu'_{-1}(x_{-1}) dx_{-1},$$

where μ'_{-1} is the density of X_{-1} , \mathcal{X}_{-1} is the interior of the support of μ'_{-1} ,

$$f_{\Delta(X)|X_{-1}}(\delta | x_{-1}) = \mu'_{1|-1}(\Delta^{-1}(\delta, x_{-1}) | x_{-1}),$$

$\mu'_{1|-1}$ is the density of X_1 conditional on X_{-1} , and $\delta \mapsto \Delta^{-1}(\delta, x_{-1})$ is the inverse function of $x_1 \mapsto \Delta(x_1, x_{-1})$. Chernozhukov, Fernández-Val, Hoderlein, Holzmann, and Newey (2015) use a similar monotonicity condition to identify quantile derivatives in nonseparable panel models. \square

5.3. Functional Derivatives of Sorting-Related Operators. We consider the properties of the distribution function and the SPE-function as functional operators $(\Delta, \mu) \mapsto F_{\Delta, \mu}$ and $(\Delta, \mu) \mapsto \Delta_{\mu}^*$. We show that these operators are Hadamard differentiable with respect to (Δ, μ) . These results are critical ingredients to deriving the large sample distributions of the empirical versions of $F_{\Delta, \mu}$ and Δ_{μ}^* in Section 6.

We now recall the definition of uniform Hadamard differentiability from van der Vaart and Wellner (1996).

Definition 5.6 (Hadamard Derivative Uniformly in an Index). Suppose the linear spaces \mathbb{D} and \mathbb{E} are equipped with the norms $\|\cdot\|_{\mathbb{D}}$ and $\|\cdot\|_{\mathbb{E}}$, and Θ is a compact subset of a metric space. A map $\phi_{\theta} : \mathbb{D}_{\phi} \subseteq \mathbb{D} \rightarrow \mathbb{E}$ is called Hadamard-differentiable uniformly in $\theta \in \Theta$ at

$f \in \mathbb{D}_\phi$ tangentially to a subspace $\mathbb{D}_0 \subseteq \mathbb{D}$ if there is a continuous linear map $\partial_f \phi_\theta : \mathbb{D}_0 \rightarrow \mathbb{E}$ such that uniformly in $\theta \in \Theta$:

$$\frac{\phi_\theta(f + t_n h_n) - \phi_\theta(f)}{t_n} - \partial_f \phi_\theta[h] \rightarrow 0, \quad n \rightarrow \infty, \quad (5.16)$$

for all converging real sequences $t_n \rightarrow 0$ and $\|h_n - h\|_{\mathbb{D}} \rightarrow 0$ such that $f + t_n h_n \in \mathbb{D}_\phi$ for every n , and $h \in \mathbb{D}_0$; moreover, the map $(\theta, G) \mapsto \partial_f \phi_\theta[G]$ is continuous on $\Theta \times \mathbb{D}_0$.

5.3.1. Hadamard differentiability of $F_{\Delta, \mu}$ and Δ_μ^* with respect to Δ . We first show differentiability of the sorting operator with respect to the PE map $x \mapsto \Delta(x)$.

In what follows, we let \mathbb{F} denote the space of continuous functions on $B(\mathcal{X})$ equipped with the sup-norm, and \mathbb{F}_0 denote a subset of \mathbb{F} that contains uniformly continuous functions.

Lemma 3 (Hadamard differentiability of $\Delta \mapsto F_{\Delta, \mu}$ and $\Delta \mapsto \Delta_\mu^*$). *Suppose that S.1-S.2 hold. Then:*

(a) *The map $F_{\Delta, \mu}(\delta) : \mathbb{F} \rightarrow \mathbb{R}$ is Hadamard-differentiable uniformly in $\delta \in \mathcal{D}$ at Δ tangentially to \mathbb{F}_0 , with the derivative map $\partial_\Delta F_{\Delta, \mu}(\delta) : \mathbb{F}_0 \rightarrow \mathbb{R}$ defined by*

$$G \mapsto \partial_\Delta F_{\Delta, \mu}(\delta)[G] := - \int_{\mathcal{M}_\Delta(\delta)} \frac{G(x) \mu'(x)}{\|\partial \Delta(x)\|} d\text{Vol}.$$

(b) *The map $\Delta_\mu^*(u) : \mathbb{F} \rightarrow \mathbb{R}$ is Hadamard-differentiable uniformly in $u \in \mathcal{U}$ at Δ tangentially to \mathbb{F}_0 , with the derivative map $\partial_\Delta \Delta_\mu^*(u) : \mathbb{F}_0 \rightarrow \mathbb{R}$ defined by:*

$$G \mapsto \partial_\Delta \Delta_\mu^*(u)[G] := - \frac{\partial_\Delta F_{\Delta, \mu}(\Delta_\mu^*(u))[G]}{f_{\Delta, \mu}(\Delta_\mu^*(u))}.$$

5.3.2. Hadamard differentiability of $F_{\Delta, \mu}$ and Δ_μ^* with respect to μ . To show differentiability with respect to the distribution μ , it is convenient to identify μ with an operator:

$$g \mapsto \mu(g) = \int_{\mathcal{X}} g d\mu(x),$$

mapping from the set $\mathcal{G} := \{x \mapsto 1(f(x) \leq \delta) : f \in \mathcal{F}, \delta \in \mathcal{V}\}$ to \mathbb{R} , where \mathcal{F} is a fixed subset of continuous functions on $B(\mathcal{X})$, containing Δ , and \mathcal{V} is any compact set of \mathbb{R} . We require \mathcal{G} to be totally bounded under the $L^2(\mu)$ norm. Define \mathbb{H} as the set of all bounded linear operators H on \mathcal{G} , of the form

$$g \mapsto H(g).$$

We define the boundedness of these operators with respect to the norm:

$$\|H\|_{\mathcal{G}} = \sup_{g \in \mathcal{G}} |H(g)|,$$

and define the corresponding distance between two operators H and \tilde{H} in \mathbb{H}_0 as $\|H - \tilde{H}\|_{\mathcal{G}} = \sup_{g \in \mathcal{G}} |H(g) - \tilde{H}(g)|$. Clearly, $\mu \in \mathbb{H}$. We consider \mathbb{H}_0 as a subspace of operators where the map $g \mapsto H(g)$ is uniformly continuous on $g \in \mathcal{G}$ under the $L^2(\mu)$ norm.

Lemma 4 (Hadamard Differentiability of $\mu \mapsto F_{\Delta, \mu}$ and $\mu \mapsto \Delta_{\mu}^*(u)$). *Suppose that S.1-S.2 hold. Then,*

(a) *The map $\mu \mapsto F_{\Delta, \mu}(\delta)$, mapping $\mathbb{H} \rightarrow \mathbb{R}$, is Hadamard differentiable uniformly in $\delta \in \mathcal{D}$ at μ tangentially to \mathbb{H}_0 with the derivative map $F_{\Delta, \mu}(\delta) : \mathbb{H}_0 \rightarrow \mathbb{R}$ defined by*

$$\partial_{\mu} F_{\Delta, \mu}(\delta)[H] := H(g_{\Delta, \delta}), \quad g_{\Delta, \delta}(x) := 1\{\Delta(x) \leq \delta\}. \quad (5.17)$$

(b) *The map $\mu \mapsto \Delta_{\mu}^*(u)$, mapping $\mathbb{H} \rightarrow \mathbb{R}$, is Hadamard differentiable uniformly in $u \in \mathcal{U}$ at μ tangentially to \mathbb{H}_0 with the derivative map $\Delta_{\mu}^*(u) : \mathbb{H}_0 \rightarrow \mathbb{R}$ defined by*

$$H \mapsto \partial_{\mu} \Delta_{\mu}^*(u)[H] := -\frac{\partial_{\mu} F_{\Delta, \mu}(\Delta_{\mu}^*(u))[H]}{f_{\Delta, \mu}(\Delta_{\mu}^*(u))}. \quad (5.18)$$

5.3.3. *Hadamard differentiability of $F_{\Delta, \mu}$ and Δ_{μ}^* with respect to (Δ, μ) .* The following Lemma combines the results of the previous two subsections.

Lemma 5 (Hadamard differentiability of $(\Delta, \mu) \mapsto F_{\Delta, \mu}$ and $(\Delta, \mu) \mapsto \Delta_{\mu}^*$). *Let $\mathbb{D} := \mathbb{F} \times \mathbb{H}_0$ and $\mathbb{D}_0 := \mathbb{F}_0 \times \mathbb{H}_0$. Assume that S.1-S.2 hold. Then,*

(a) *The map $(\Delta, \mu) \mapsto F_{\Delta, \mu}(\delta)$, mapping $\mathbb{D} \rightarrow \mathbb{R}$, is Hadamard differentiable uniformly in $\delta \in \mathcal{D}$ at (Δ, μ) tangentially to \mathbb{D}_0 with the derivative map $\partial_{\Delta, \mu} F_{\Delta, \mu}(\delta) : \mathbb{D}_0 \rightarrow \mathbb{R}$ defined by*

$$(G, H) \mapsto \partial_{\Delta, \mu} F_{\Delta, \mu}(\delta)[G, H] := -\int_{\mathcal{M}_{\Delta}(\delta)} \frac{G(x)\mu'(x)}{\|\partial\Delta(x)\|} d\text{Vol} + H(g_{\Delta, \delta}).$$

(b) *The map $(\Delta, \mu) \mapsto \Delta_{\mu}^*(u)$, mapping $\mathbb{D} \rightarrow \mathbb{R}$ is Hadamard differentiable uniformly in $u \in \mathcal{U}$ at (Δ, μ) tangentially to \mathbb{D}_0 with the derivative map, $\partial_{\Delta, \mu} \Delta_{\mu}^*(u) : \mathbb{D}_0 \rightarrow \mathbb{R}$, defined by*

$$(G, H) \mapsto \partial_{\Delta, \mu} \Delta_{\mu}^*(u)[G, H] := -\frac{\partial_{\Delta, \mu} F_{\Delta, \mu}(\Delta_{\mu}^*(u))[G, H]}{f_{\Delta, \mu}(\Delta_{\mu}^*(u))}.$$

6. THEORETICAL ANALYSIS II: LARGE SAMPLE PROPERTIES OF EMPIRICAL SPE

We use the Hadamard differentiability of the sorting-related operators and the functional delta method to derive functional central limit theorems for $\delta \mapsto \widehat{F}_{\Delta, \mu}(\delta)$ and $u \mapsto \widehat{\Delta}_{\mu}^*(u)$ over regions that exclude the critical values of $x \mapsto \Delta(x)$ on $B(\mathcal{X})$. To describe this results, let $\ell^{\infty}(\mathcal{T})$ denote the set of bounded and measurable functions $g : \mathcal{T} \mapsto \mathbb{R}$. We consider two different cases depending on whether the distribution μ is treated as known or unknown.

6.1. **Case 1: Δ unknown, μ known.** We first discuss the properties of $F_{\widehat{\Delta},\mu}$ and $\widehat{\Delta}_\mu^*$, the estimators of $F_{\Delta,\mu}$ and Δ_μ^* when μ is treated as known. We make the following assumptions about the estimator of the PE:

S.3. $\widehat{\Delta}$, the estimator of Δ , belongs to \mathcal{F} and obeys a functional central limit theorem, namely,

$$a_n(\widehat{\Delta} - \Delta) \rightsquigarrow G_\infty \text{ in } \ell^\infty(B(\mathcal{X})),$$

where a_n is a sequence such that $a_n \rightarrow \infty$ as $n \rightarrow \infty$, and $x \mapsto G_\infty(x)$ is a tight process that has almost surely uniformly continuous sample paths on $B(\mathcal{X})$.

The following result is a corollary of Lemma 3:

Theorem 1 (FCLT for $F_{\widehat{\Delta},\mu}$ and $\widehat{\Delta}_\mu^*$). *Under S.1-S.3, as $n \rightarrow \infty$:*

(a) *The estimator of the distribution of the PE obeys a functional central limit theorem, namely, in $\ell^\infty(\mathcal{D})$,*

$$a_n(F_{\widehat{\Delta},\mu}(\delta) - F_{\Delta,\mu}(\delta)) \rightsquigarrow \partial_\Delta F_{\Delta,\mu}(\delta)[G_\infty] = - \int_{\mathcal{M}_\Delta(\delta)} \frac{G_\infty(x)\mu'(x)}{\|\partial\Delta(x)\|} d\text{Vol} =: T_\infty(\delta),$$

as a stochastic process indexed by $\delta \in \mathcal{D}$.

(b) *The empirical SPE-process obeys a functional central limit theorem, namely in $\ell^\infty(\mathcal{U})$,*

$$a_n(\widehat{\Delta}_\mu^*(u) - \Delta_\mu^*(u)) \rightsquigarrow \partial_\Delta \Delta_\mu^*(u)[G_\infty] = - \frac{T_\infty(\Delta_\mu^*(u))}{f_{\Delta,\mu}(\Delta_\mu^*(u))}, \quad (6.19)$$

as a stochastic process indexed by $u \in \mathcal{U}$.

Remark 6.1 (Form of the Limit Process for SPE). Replacing the expressions of $T_\infty(\delta)$ and $f_{\Delta,\mu}(\delta)$ in (6.19),

$$\partial_\Delta \Delta_\mu^*(u)[G_\infty] = \frac{\int_{\mathcal{M}_\Delta(\Delta_\mu^*(u))} \frac{G_\infty(x)\mu'(x)}{\|\partial\Delta(x)\|} d\text{Vol}}{\int_{\mathcal{M}_\Delta(\Delta_\mu^*(u))} \frac{\mu'(x)}{\|\partial\Delta(x)\|} d\text{Vol}}.$$

The limit process is therefore the average of the process $G_\infty(x)$ on $\mathcal{M}_\Delta(\Delta_\mu^*(u))$ with respect to the density

$$\frac{\frac{\mu'(x)}{\|\partial\Delta(x)\|}}{\int_{\mathcal{M}_\Delta(\Delta_\mu^*(u))} \frac{\mu'(x)}{\|\partial\Delta(x)\|} d\text{Vol}}.$$

6.2. Case 2: Both Δ and μ unknown. We consider now the most empirically relevant case where both Δ and μ are estimated. We make the following assumption about $\hat{\mu}$, the estimator of the distribution μ :

S.4. The function $x \mapsto \hat{\mu}(x)$ is a distribution over $B(\mathcal{X})$ obeying in \mathbb{H} ,

$$b_n(\hat{\mu} - \mu) \rightsquigarrow H_\infty, \quad (6.20)$$

where $g \mapsto H_\infty(g)$ is a.s. an element of \mathbb{H}_0 , i.e. it has almost surely uniformly continuous sample paths on \mathcal{G} with respect to the $L^2(\mu)$ metric; and where b_n is a sequence such that $b_n \rightarrow \infty$ as $n \rightarrow \infty$.

Remark 6.2 (Empirical distribution). When $\hat{\mu}$ is the empirical distribution based on a random sample from the population with distribution μ , then $b_n = \sqrt{n}$ and $H_\infty := B_\mu$, where B_μ is a μ -Brownian Bridge, i.e. a Gaussian process with zero mean and covariance function $(g_1, g_2) \mapsto \mu(g_1 g_2) - \mu(g_1)\mu(g_2)$. \square

Remark 6.3 (Donsker condition). When $\hat{\mu}$ is the empirical distribution, condition S.4 imposes that the function class

$$\mathcal{G} = \{1(f \leq \delta) : f \in \mathcal{F}, \delta \in \mathcal{V}\}$$

is μ -Donsker. Note that \mathcal{F} is the parameter space that contains $\Delta(x)$ as well as $\hat{\Delta}(x)$ in S.3. In parametric models for the PE where $\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$, f is known, $\theta \subseteq \mathbb{R}^{d_\theta}$ with $d_\theta < \infty$, and $x \mapsto f(x, \theta)$ is \mathcal{C}^1 on \mathcal{X} for all $\theta \in \Theta$, the class \mathcal{G} is μ -Donsker under mild conditions specified for example in van der Vaart (1998, Chap. 19). Examples 1 and 2 specify the PE parametrically. Lemma 6 gives other sufficient conditions for the Donsker property. \square

Lemma 6 (Sufficient conditions for \mathcal{G} being μ -Donsker). *Suppose S.1-S.2 hold, and \mathcal{V} is the union of a finite number of compact intervals. Suppose that \mathcal{F} satisfies:*

$$\sup_{\tilde{\Delta} \in \mathcal{F}} \sup_{x \in B(\mathcal{X})} \|\partial \tilde{\Delta}(x) - \partial \Delta(x)\| + \sup_{\tilde{\Delta} \in \mathcal{F}} \sup_{x \in B(\mathcal{X})} |\tilde{\Delta}(x) - \Delta(x)| < c_0.$$

Let $N(\epsilon, \mathcal{F}, \|\cdot\|_\infty)$ be the ϵ -covering number of the class \mathcal{F} under L_∞ norm. Suppose that $\int_0^1 \sqrt{\log N(\epsilon^2, \mathcal{F}, \|\cdot\|_\infty)} d\epsilon < \infty$. If c_0 is small enough, then, \mathcal{G} is μ -Donsker.

Let $r_n := a_n \wedge b_n$, the slowest of the rates of convergence of $\hat{\Delta}$ and $\hat{\mu}$. Assume $r_n/a_n \rightarrow s_\Delta \in [0, 1]$ and $r_n/b_n \rightarrow s_\mu \in [0, 1]$, where $s_\Delta = 0$ when $b_n = o(a_n)$ and $s_\mu = 0$ when $a_n = o(b_n)$.

The following result is a corollary of Lemma 5.

Theorem 2 (FCLT for $F_{\hat{\Delta}, \hat{\mu}}$ and $\hat{\Delta}_\mu^*$). *Suppose that S.1-S.4 hold, and the convergence in S.3 and S.4 holds jointly. Then, as $n \rightarrow \infty$,*

(a) *The estimator of the distribution of PE obeys a functional central limit theorem, namely, in $\ell^\infty(\mathcal{D})$,*

$$r_n(F_{\widehat{\Delta}_\mu}(\delta) - F_{\Delta,\mu}(\delta)) \rightsquigarrow \partial_{\Delta,\mu} F_{\Delta,\mu}(\delta)[s_\Delta G_\infty, s_\mu H_\infty] = s_\Delta T_\infty(\delta) + s_\mu H_\infty(g_{\Delta,\delta}), \quad (6.21)$$

as a stochastic process indexed by $\delta \in \mathcal{D}$.

(b) *The empirical SPE-process obeys a functional central limit theorem, namely in $\ell^\infty(\mathcal{U})$,*

$$\begin{aligned} r_n(\widehat{\Delta}_\mu^*(u) - \Delta_\mu^*(u)) \rightsquigarrow \partial_{\Delta,\mu} \Delta_\mu^*(u)[s_\Delta G_\infty, s_\mu H_\infty] &= -\frac{s_\Delta T_\infty(\Delta_\mu^*(u)) + s_\mu H_\infty(g_{\Delta,\Delta_\mu^*(u)})}{f_{\Delta,\mu}(\Delta_\mu^*(u))} \\ &=: Z_\infty(u), \end{aligned} \quad (6.22)$$

as a stochastic process indexed by $u \in \mathcal{U}$.

Theorem 1 can be seen a special case of Theorem 2 with $r_n = a_n$, $s_\Delta = 1$, and $s_\mu = 0$.

6.3. Bootstrap Inference. Corollary 1 of Section 3 uses critical values of a statistic related to the limit process Z_∞ to construct confidence bands. These critical values can be hard to obtain in practice. In principle one can use simulation, but it might be difficult to numerically locate and parametrize the manifold $\mathcal{M}_\Delta(\delta)$, and to evaluate the integrals on $\mathcal{M}_\Delta(\delta)$ needed to compute $Z_\infty(u)$. This creates a real challenge to implement our inference methods. To deal with this challenge we employ (exchangeable) bootstrap to compute critical values (Præstgaard and Wellner, 1993; van der Vaart and Wellner, 1996), instead of simulation. We show that the bootstrap law is consistent to approximate the distribution of the limit process Z_∞ of Theorem 2.

To state the bootstrap validity result formally, we follow the notation and definitions in van der Vaart and Wellner (1996). Let D_n denote the data vector and let $B_n = (\omega_1, \dots, \omega_n)$ be the vector of bootstrap weights. Consider a random element $\tilde{Z}_n = Z_n(D_n, B_n)$ in a normed space \mathbb{D} . We say that the bootstrap law of \tilde{Z}_n consistently estimates the law of some tight random element Z_∞ and write $\tilde{Z}_n \rightsquigarrow_{\mathbb{P}} Z_\infty$ if

$$\sup_{h \in \text{BL}_1(\mathbb{D})} |\mathbb{E}_{B_n} h(\tilde{Z}_n) - \mathbb{E}_{\mathbb{P}} h(Z_\infty)| \rightarrow_{\mathbb{P}} 0,$$

where $\text{BL}_1(\mathbb{D})$ denotes the space of functions with Lipschitz norm at most 1; \mathbb{E}_{B_n} denotes the conditional expectation with respect to B_n given the data D_n ; $\mathbb{E}_{\mathbb{P}}$ denotes the conditional expectation with respect to \mathbb{P} , the distribution of the data D_n ; and $\rightarrow_{\mathbb{P}}$ denotes convergence in (outer) probability.

The next result is a consequence of the functional delta method for the exchangeable bootstrap.

Theorem 3 (Bootstrap FCLT for $\widehat{\Delta}_\mu^*$). *Suppose that the assumptions of Theorem 2 hold, and that the bootstrap is consistent for the law of the estimator of the PE, namely $a_n(\tilde{\Delta} - \widehat{\Delta}) \rightsquigarrow_{\mathbb{P}} G_\infty$ in $\ell^\infty(B(\mathcal{X}))$, and for the law of the estimated measure, namely $b_n(\tilde{\mu} - \widehat{\mu}) \rightsquigarrow_{\mathbb{P}} H_\infty$ in \mathbb{H} . Then, the bootstrap is consistent for the law of the empirical SPE-process, namely*

$$r_n(\widehat{\Delta}_\mu^* - \widehat{\Delta}_\mu^*) \rightsquigarrow_{\mathbb{P}} Z_\infty \text{ in } \ell^\infty(\mathcal{U}).$$

Remark 6.4 (Bootstrap FCLT for estimators of Δ and μ). See van der Vaart and Wellner (1996) and Chernozhukov, Fernández-Val, and Melly (2013) for bootstrap FCLT for parametric and semi parametric estimators of Δ including least squares, quantile regression, and distribution regression; and nonparametric estimators of μ including the empirical distribution function. \square

Remark 6.5 (Exchangeable Bootstrap). Theorem 3 employs the high-level condition that the bootstrap can approximate consistently the laws of $\widehat{\Delta}$ and $\widehat{\mu}$, after suitable rescaling. In Examples 1-3 when $\widehat{\mu}$ is the empirical measure based on the random sample of size n , the exchangeable bootstrap method entails to randomly reweighing the sample using the weights $(\omega_1, \dots, \omega_n)$. In this case the high level condition holds if the weights satisfy:

B. $B_n = (\omega_1, \dots, \omega_n)$ is an exchangeable, nonnegative random vector, which is independent of the data D_n , such that for some $\epsilon > 0$,

$$\sup_n \mathbb{E}[\omega_i^{2+\epsilon}] < \infty, \quad n^{-1} \sum_{i=1}^n (\omega_i - \bar{\omega})^2 \rightarrow_{\mathbb{P}} 1, \quad \bar{\omega} \rightarrow_{\mathbb{P}} 1,$$

where $\bar{\omega} = n^{-1} \sum_{i=1}^n \omega_i$.⁵ \square

6.4. Some Numerical Illustrations. We evaluate the accuracy of the asymptotic approximations to the distribution of the empirical SPE in small samples using numerical simulations. In particular, we compare pointwise 95% confidence intervals for the SPE based on the asymptotic and exact distributions of the empirical SPE. The exact distribution is approximated numerically by simulation. The asymptotic distribution is obtained analytically from the FCLT of Theorem 2, and approximated by bootstrap using Theorem 3. We consider two simulation

⁵A sequence of random variables $\omega_1, \omega_2, \dots, \omega_n$ is exchangeable if for any finite permutation σ of the indices $1, 2, \dots, n$ the joint distribution of the permuted sequence $\omega_{\sigma(1)}, \omega_{\sigma(2)}, \dots, \omega_{\sigma(n)}$ is the same as the joint distribution of the original sequence.

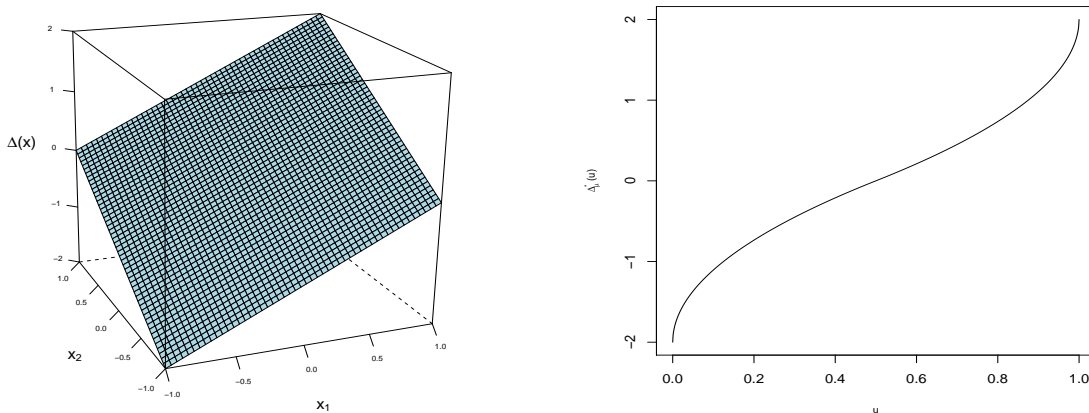


FIGURE 6. PE-function and SPE-function in Design 1. Left: PE function $x \mapsto \Delta(x)$. Right: SPE function $u \mapsto \Delta_\mu^*(u)$.

designs where the limit process in Theorem 2 has a convenient closed-form analytical expression. The designs differ on whether the PE-function $x \mapsto \Delta(x)$ has critical points or not. We hold fix the values of the covariate vector X in all the calculations, and accordingly we treat the distribution μ as known. For the bootstrap inference, we use empirical bootstrap with $B = 3,000$ repetitions. All the results are based on $S = 3,000$ simulations.

Design 1 (No critical points). We consider the PE-function

$$\Delta(x) = x_1 + x_2, \quad x = (x_1, x_2),$$

with the covariate vector X uniformly distributed in $\mathcal{X} = (-1, 1) \times (-1, 1)$. The corresponding SPE is

$$\Delta_\mu^*(u) = 2(\sqrt{2u} - 1)1(u \leq 1/2) + 2(1 - \sqrt{2(1-u)})1(u > 1/2),$$

where $1(\cdot)$ denotes the indicator function, and we use that $\Delta(X)$ has a triangular distribution with parameters $(-2, 0, 2)$. The sample size is $n = 441$ and the values of X are hold fixed in the grid $\{-1, -0.9, \dots, 1\} \times \{-1, -0.9, \dots, 1\}$. Figure 6 plots $x \mapsto \Delta(x)$ on \mathcal{X} , and $u \mapsto \Delta_\mu^*(u)$ on $(0, 1)$. Here we see that $x \mapsto \Delta(x)$ does not have critical values, and that $u \mapsto \Delta_\mu^*(u)$ is a smooth function.

To obtain an analytical expression of the limit $Z_\infty(u)$ of Theorem 2, we make the following assumption on the estimator of the PE:

$$\sqrt{n}(\widehat{\Delta}(x) - \Delta(x)) = \exp[\Delta(x)] \sum_{i=1}^n Z_i / \sqrt{n},$$

where Z_1, \dots, Z_n is an i.i.d. sequence of standard normal random variables. Hence

$$Z_\infty(u) \sim N(0, \exp[2\Delta_\mu^*(u)]),$$

so that $\widehat{\Delta}_\mu^*(u) \stackrel{a}{\sim} N(\Delta_\mu^*(u), \exp[2\Delta_\mu^*(u)]/n)$, where $\stackrel{a}{\sim}$ denotes asymptotic approximation to the distribution.

Table 5 reports biases and compares the standard deviations of the empirical SPE with the asymptotic standard deviations, $\exp[\Delta_\mu^*(u)]/\sqrt{n}$, at the quantile indexes $u \in \{0.1, 0.2, \dots, 0.9\}$. The biases are small relative to dispersions and the asymptotic approximations are very close to the exact standard deviations. We also find that 95% confidence intervals constructed using the asymptotic approximations, $\widehat{\Delta}_\mu^*(u) \pm 1.96 \exp[\Delta_\mu^*(u)]/\sqrt{n}$, have coverage probabilities close to their nominal levels at all quantiles. These asymptotic confidence intervals are not feasible in general, because $\Delta_\mu^*(u)$ are unknown or more generally because it is not possible to characterize analytically the distribution of $Z_\infty(u)$. In practice we propose to approximate this distribution by bootstrap. In this case the empirical bootstrap version of the empirical SPE is constructed from the bootstrap PE

$$\tilde{\Delta}(x) = \Delta(x) + \exp[\Delta(x)] \sum_{i=1}^n \omega_i Z_i / n,$$

where $(\omega_1, \dots, \omega_n)$ is a multinomial vector with dimension n and probabilities $(1/n, \dots, 1/n)$ independent of Z_1, \dots, Z_n . The last column of the table shows that the empirical coverages of bootstrap 95% confidence intervals are close to their nominal levels at all quantiles indexes.

Design 2 (Critical points). We consider the PE-function

$$\Delta(x) = x^3 - 3x,$$

with covariate X uniformly distributed on $\mathcal{X} = (-3, 3)$. Figure 7 plots $x \mapsto \Delta(x)$ on \mathcal{X} , and $u \mapsto \Delta_\mu^*(u)$ on $(0, 1)$.⁶ Here we see that $x \mapsto \Delta(x)$ has two critical points at $x = -1$ and $x = 1$ with corresponding critical values at $\delta = 2$ and $\delta = -2$. The SPE-function $u \mapsto \Delta_\mu^*(u)$ has two kinks at $u = 1/6$ and $u = 5/6$, the Δ_μ^* pre-images of the critical values.

⁶We obtain $u \mapsto \Delta_\mu^*(u)$ analytically using the characterization of Chernozhukov, Fernández-Val, and Galichon (2010) for the univariate case.

TABLE 5. Properties of Empirical SPE in Design 1

u	Bias	Std. Dev.		Pointwise Coverage (%)	
	($\times 100$)	Exact	Asymptotic	Asymptotic	Bootstrap [†]
0.1	0.016	0.014	0.014	95.10	95.03
0.2	0.024	0.021	0.021	95.10	95.03
0.3	0.032	0.029	0.029	95.10	95.03
0.4	0.044	0.039	0.039	95.10	95.03
0.5	0.053	0.047	0.048	95.10	95.03
0.6	0.065	0.058	0.058	95.10	95.03
0.7	0.088	0.078	0.079	95.10	95.03
0.8	0.119	0.105	0.106	95.10	95.03
0.9	0.177	0.157	0.158	95.10	95.03

Notes: 3,000 simulations with sample size $n = 441$.

[†]3,000 bootstrap repetitions. Nominal level is 95%.

To obtain an analytical expression of the limit $Z_\infty(u)$ of Theorem 2, we make the following assumption on the estimator of the PE:

$$\sqrt{n}(\widehat{\Delta}(x) - \Delta(x)) = (x/2)^2 \sum_{i=1}^n Z_i / \sqrt{n},$$

where Z_1, \dots, Z_n is an i.i.d. sequence of standard normal variables. This assumption is analytically convenient because after some calculations we find that for $u \notin \{1/5, 5/6\}$,

$$Z_\infty(u) \sim N(0, S(\Delta_\mu^*(u))^2 / (4n)),$$

where

$$S(\delta) = 1(\delta < -2)\check{\Delta}_1(\delta)^2 + 1(-2 < \delta < 2) \sum_{k=1}^3 \frac{\check{\Delta}_k(\delta)^2 |\check{\Delta}_k(\delta)^2 - 1|^{-1}}{\sum_{j=1}^3 |\check{\Delta}_j(\delta)^2 - 1|^{-1}} + 1(\delta > 2)\check{\Delta}_3(\delta)^2,$$

and $\check{\Delta}_1(\delta)$, $\check{\Delta}_2(\delta)$ and $\check{\Delta}_3(\delta)$ are the real roots of $\Delta(x) - \delta = 0$ sorted in increasing order.⁷ Hence, $\widehat{\Delta}_\mu^*(u) \stackrel{a}{\sim} N(\Delta_\mu^*(u), S(\Delta_\mu^*(u))^2 / (4n))$.

Table 6 reports biases and compares the standard deviations of the empirical SPE in samples of size $n = 601$ with the asymptotic standard deviations at the quantile indexes $u \in \{1/12, 2/12, \dots, 11/12\}$, where the values of X are hold fixed in the grid $\{-3, -2.99, \dots, 3\}$.

⁷The equation $\Delta(x) - \delta = x^3 - 3x - \delta = 0$ has three real roots when $\delta \in (-2, 2)$, and one real root when $\delta < -2$ or $\delta > 2$.

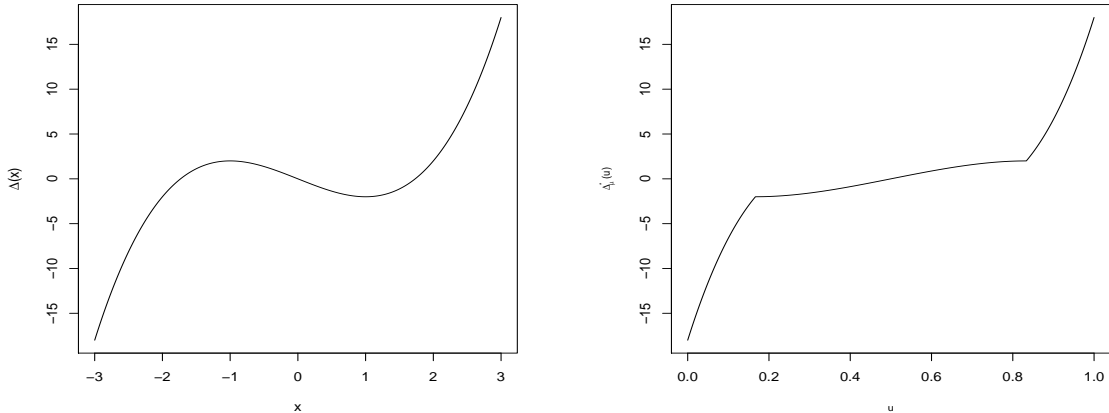


FIGURE 7. PE-function and SPE-function in Design 2. Left: PE function $x \mapsto \Delta(x)$. Right: SPE function $u \mapsto \Delta_\mu^*(u)$.

The biases are small relative to dispersion except at the kinks $u = 1/5$ and $u = 5/6$. The asymptotic approximation is close to the exact standard deviation, except for the quantiles at the kinks where the asymptotic standard deviations are not well-defined because $\check{\Delta}_k(\delta)^2 - 1 = 0$. We also find that pointwise 95% confidence intervals constructed using the asymptotic distribution and empirical bootstrap have coverage probabilities close to their nominal levels. Interestingly, the bootstrap still provides coverages close to the nominal levels even at the kinks.

APPENDIX A. EXTENSION OF THEORETICAL ANALYSIS TO DISCRETE VARIABLES

We consider the case where the covariate X includes discrete components. Without loss of generality we assume that the first component of X is discrete and the rest are continuous. Accordingly, we do the partition $X = (D, C)$. Let $\mathcal{X}_{c|d}$ denote the interior of the support of C conditional on $D = d$, \mathcal{X}_d denote the support of D , $\mu_{c|d}$ the distribution of C conditional on $D = d$, μ_d is the distribution of D , and $\pi_d(d) = P(D = d)$. The distribution μ is then defined

TABLE 6. Properties of Empirical SPE in Design 2

u	Bias	Std. Dev.		Pointwise Coverage (%)	
	($\times 100$)	Exact	Asymptotic	Asymptotic	Bootstrap [†]
1/12	0.068	0.126	0.127	95.67	95.80
1/6	-2.393	0.054	–	–	95.67
1/4	-0.005	0.025	0.025	95.83	95.77
1/3	-0.016	0.028	0.028	95.80	95.90
5/12	0.045	0.030	0.030	95.63	95.47
1/2	0.023	0.030	0.031	92.73	97.53
7/12	-0.020	0.030	0.030	95.20	95.80
2/3	0.049	0.028	0.028	95.53	95.67
3/4	0.039	0.025	0.025	95.53	95.73
5/6	2.447	0.053	–	–	95.73
11/12	0.068	0.126	0.127	95.67	95.80

Notes: 3,000 simulations with sample size $n = 601$.

[†]3,000 bootstrap repetitions. Nominal level is 95%.

by $\mu(c, d) := \mu_d(d)\mu_{c|d}(c)$. We continue denoting by $d_x = \dim(X)$, and by \mathcal{D} a compact set consisting of regular values of Δ on $\bar{\mathcal{X}} := \cup_{d \in \mathcal{X}_d} \{d\} \times \bar{\mathcal{X}}_{c|d}$, where $\bar{\mathcal{X}}_{c|d}$ is the closure of $\mathcal{X}_{c|d}$.

We adjust S.1-S.4 to hold conditionally at each value of the discrete covariate.

S.1'. The set \mathcal{X}_d is finite. For any $d \in \mathcal{X}_d$: the set $\mathcal{X}_{c|d}$ is open and its closure $\bar{\mathcal{X}}_{c|d}$ is compact; the distribution $\mu_{c|d}$ is absolutely continuous with respect to the Lebesgue measure with density $\mu'_{c|d}$; and there exists an open set $B(\mathcal{X}_{c|d})$ containing $\bar{\mathcal{X}}_{c|d}$ such that $c \mapsto \Delta(d, c)$ is \mathcal{C}^1 on $B(\mathcal{X}_{c|d})$, and $c \mapsto \mu'_{c|d}(c)$ is continuous on $B(\mathcal{X}_{c|d})$ and is zero outside $\mathcal{X}_{c|d}$, i.e. $\mu'(x) = 0$ for any $x \in B(\mathcal{X}_{c|d}) \setminus \mathcal{X}_{c|d}$.

S.2'. For any $d \in \mathcal{X}_d$ and any regular value δ of Δ on $\bar{\mathcal{X}}_{c|d}$, $\mathcal{M}_{\Delta|d}(\delta) := \{c \in \bar{\mathcal{X}}_{c|d} : \Delta(d, c) = \delta\}$ is either a $(d_x - 2)$ -manifold without boundary on $\mathbb{R}^{d_x - 1}$ of class \mathcal{C}^1 with finite number of connected branches, or an empty set.

S.3'. $\hat{\Delta}$, the estimator of Δ , obeys a functional central limit theorem, namely,

$$a_n(\hat{\Delta} - \Delta) \rightsquigarrow G_\infty \text{ in } \ell^\infty(B(\mathcal{X})),$$

where a_n is a sequence such that $a_n \rightarrow \infty$ as $n \rightarrow \infty$, and $c \mapsto G_\infty(d, c)$ is a tight process that has almost surely uniformly continuous sample paths on $B(\mathcal{X}_{c|d})$ for all $d \in \mathcal{X}_d$.

Let $B(\mathcal{X}) := \cup_{d \in \mathcal{X}_d} \{d\} \times B(\mathcal{X}_{c|d})$; \mathcal{F} denote a set of continuous functions on $B(\mathcal{X})$ equipped with the sup-norm; \mathcal{V} be any compact subset of \mathbb{R} ; \mathbb{H} be the set of all bounded operators $H : g \mapsto H(g)$ on $\mathcal{G} = \{1(f \leq \delta) : f \in \mathcal{F}, \delta \in \mathcal{V}\}$ that are represented as:

$$H(g) = \sum_{d \in \mathcal{X}_d} H_d(d) \int_{\mathcal{X}_{c|d}} g(c, d) d\mu_{c|d}(c) + \sum_{d \in \mathcal{X}_d} \pi_d(d) H_{c|d}(g(\cdot, d)),$$

where $d \mapsto H_d(d)$ is a function that takes on finitely many values and $g \mapsto H_{c|d}(g)$ is a bounded linear operator on \mathcal{G} . Equip the space \mathbb{H} with the sup norm $\|\cdot\|_{\mathcal{G}}$: $\|H\|_{\mathcal{G}} = \sup_{g \in \mathcal{G}} |H(g)|$. Let $\mu(x) = \mu_d(d) \mu_{c|d}(c)$ and $\widehat{\mu}(x) = \widehat{\mu}_d(d) \widehat{\mu}_{c|d}(c)$. Let \mathbb{H}_0 be the subset of \mathbb{H} , which contains operators that are uniformly continuous with respect to the $L^2(\mu)$ norm on \mathcal{G} .

S.4'. The function $x \mapsto \widehat{\mu}(x)$ is a distribution over $B(\mathcal{X})$ obeying in \mathbb{H} ,

$$b_n(\widehat{\mu} - \mu) \rightsquigarrow H_\infty,$$

where $H_\infty \in \mathbb{H}_0$ a.s., and b_n is a sequence such that $b_n \rightarrow \infty$ as $n \rightarrow \infty$, and H_∞ can be represented as:

$$H_\infty(g) = \sum_{d \in \mathcal{X}_d} H_{d,\infty}(d) \int_{\mathcal{X}_{c|d}} g(c, d) d\mu_{c|d}(c) + \sum_{d \in \mathcal{X}_d} \pi_d(d) H_{c|d,\infty}(g(\cdot, d)).$$

We generalize Lemmas 2 and 5 to the case where X includes discrete components.

Define $\mathbb{D} := \mathbb{F} \times \mathbb{H}$ and $\mathbb{D}_0 := \mathbb{F}_0 \times \mathbb{H}_0$, where \mathbb{F} is the set of continuous functions on $B(\mathcal{X})$ and \mathbb{F}_0 is the subset of uniformly continuous functions.

Lemma 7 (Properties of $F_{\Delta,\mu}$ and Δ_μ^* with discrete X). *Suppose that S.1' and S.2' hold. Then, $\delta \mapsto F_{\Delta,\mu}(\delta)$ is differentiable at any $\delta \in \mathcal{D}$, with derivative function $f_{\Delta,\mu}(\delta)$ defined as:*

$$f_{\Delta,\mu}(\delta) := \partial_\delta F_{\Delta,\mu}(\delta) = \sum_{d \in \mathcal{X}_d} \pi_d(d) \int_{\mathcal{M}_{\Delta|d}(\delta)} \frac{\mu'_{c|d}(c)}{\|\partial_c \Delta(d, c)\|} d\text{Vol}.$$

(1) *The map $F_{\Delta,\mu}(\delta) : \mathbb{D} \rightarrow \mathbb{R}$ is Hadamard differentiable uniformly in $d \in \mathcal{D}$ at (Δ, μ) tangentially to \mathbb{D}_0 , with derivative map $\partial_{\Delta,\mu} F_{\Delta,\mu}(\delta) : \mathbb{D}_0 \rightarrow \mathbb{R}$ defined by:*

$$\begin{aligned} (G, H) \mapsto \partial_{\Delta,\mu} F_{\Delta,\mu}(\delta)[G, H] := & - \sum_{d \in \mathcal{X}_d} \pi_d(d) \int_{\mathcal{M}_{\Delta|d}(\delta)} \frac{G(d, c) \mu'_{c|d}(c)}{\|\partial_c \Delta(d, c)\|} d\text{Vol}(c) \\ & + \sum_{d \in \mathcal{X}_d} H_d(d) \int_{\mathcal{X}_{c|d}} 1\{\Delta(d, c) \leq \delta\} \mu'_{c|d}(c) dc \\ & + \sum_{d \in \mathcal{X}_d} \pi_d(d) H_{c|d}(1\{\Delta(\cdot, d) \leq \delta\}). \end{aligned}$$

(2) The map $\Delta_\mu^*(u) : \mathbb{D} \rightarrow \mathbb{R}$ is Hadamard differentiable uniformly in $u \in \mathcal{U}$ at (Δ, μ) tangentially to \mathbb{D}_0 , with derivative map $\partial_{\Delta, \mu} \Delta_\mu^*(u) : \mathbb{D}_0 \rightarrow \mathbb{R}$ defined by:

$$(G, H) \mapsto \partial_{\Delta, \mu} \Delta_\mu^*(u)[G, H] := -\frac{\partial F_{\Delta, \mu}(\Delta_\mu^*(u))[G, H]}{f_{\Delta, \mu}(\Delta_\mu^*(u))},$$

where $\mathcal{U} = \{\tilde{u} \in [0, 1] : \Delta_\mu^*(\tilde{u}) \in \mathcal{D}, f_{\Delta, \mu}(\Delta_\mu^*(\tilde{u})) > \varepsilon\}$ for fixed $\varepsilon > 0$.

We are now ready to derive a functional central limit theorem for the empirical SPE-function. As in Theorem 2, let $r_n := a_n \wedge b_n$, the slowest of the rates of convergence of $\widehat{\Delta}$ and $\widehat{\mu}$, where $r_n/a_n \rightarrow s_\Delta \in [0, 1]$ and $r_n/b_n \rightarrow s_\mu \in [0, 1]$.

Theorem 4 (FCLT for $\widehat{\Delta}_\mu^*(u)$ with discrete X). *Suppose that S.1'-S.4' hold, the convergence in S.3' and S.4' holds jointly, and $\widehat{\Delta} \in \mathcal{F}$ with probability approaching 1. Then, the empirical SPE-process obeys a functional central limit theorem, namely in $\ell^\infty(\mathcal{U})$,*

$$r_n(\widehat{\Delta}_\mu^*(u) - \Delta_\mu^*(u)) \rightsquigarrow \partial_{\Delta, \mu} \Delta_\mu^*(u)[s_\Delta G_\infty, s_\mu H_\infty], \quad (\text{A.23})$$

as a stochastic process indexed by $u \in \mathcal{U}$, where \mathcal{U} is defined in Lemma 7.

Remark A.1 (Bootstrap FCLT for $\widehat{\Delta}_\mu^*(u)$ with discrete X). The exchangeable bootstrap is consistent to approximate the distribution of the limit process in (A.23) under the same conditions as in Theorem 3, replacing S.1-S.4 by S.1'-S.4'. Accordingly, we do not repeat the statement here. \square

APPENDIX B. PROOFS OF SECTION 5

B.1. A lemma regarding Remark 3.2.

Lemma 8. *Let $\Delta : B(\mathcal{X}) \rightarrow \mathbb{R}$ be a \mathcal{C}^1 function on an open set $B(\mathcal{X})$ that contains $\overline{\mathcal{X}}$. If δ is a regular value of Δ on $\overline{\mathcal{X}}$, then the set $\mathcal{M}_\Delta(\delta) = \{x \in \mathcal{X} : \Delta(x) = \delta\}$ is a $(d_x - 1)$ -manifold in \mathbb{R}^{d_x} of class \mathcal{C}^1 .*

Proof. This follows from Theorem 5-1 in Spivak (1965, p. 111). \square

B.2. Proof of Lemma 2. We use the following results in the proof for Lemma 2.

Lemma 9. *If $\Delta : B(\mathcal{X}) \rightarrow \mathbb{R}$ is \mathcal{C}^1 on an open set $B(\mathcal{X}) \subseteq \mathbb{R}^{d_x}$, then for any compact subset $\overline{\mathcal{X}}$ of $B(\mathcal{X})$, the sets of critical points and critical values of $x \mapsto \Delta(x)$ on $\overline{\mathcal{X}}$ are closed.*

Proof. (1) Critical points: since $x \mapsto \partial\Delta(x)$ is continuous on $\overline{\mathcal{X}}$ and $\overline{\mathcal{X}}$ is compact, the set of points $x \in \overline{\mathcal{X}}$ such that $\partial\Delta(x) = 0$ is closed.

(2) Critical values: since $x \mapsto \Delta(x)$ is continuous and $\overline{\mathcal{X}}$ is compact, the image set $\Delta(\overline{\mathcal{X}})$ is a compact set in \mathbb{R} . For any sequence of critical values $\{\delta_i\}_{i \geq 1}$ in $\Delta(\overline{\mathcal{X}})$, there is a corresponding sequence $\{x_i\}_{i \geq 1}$ in $\overline{\mathcal{X}}$ such that $\Delta(x_i) = \delta_i$. Suppose $\{\delta_i\}_{i \geq 1}$ converges to $\delta_0 \in \Delta(\overline{\mathcal{X}})$. By compactness of $\overline{\mathcal{X}}$, we can find a converging subsequence of $\{x_i\}_{i \geq 1}$ with limit $x_0 \in \overline{\mathcal{X}}$ such that $\Delta(x_i) = \delta_i$. Then by continuity of $x \mapsto \partial\Delta(x)$, $\partial\Delta(x_0) = 0$. By continuity of $x \mapsto \Delta(x)$, $\Delta(x_0) = \delta_0$, and therefore $\delta_0 = \Delta(x_0)$ is a critical value of $\Delta(x)$. Hence the set of critical values is closed. \square

Lemma 10. *For a compact set \mathcal{V} in a metric space \mathbb{D} , suppose there is an open cover $\{\theta_i : i \in I\}$ of \mathcal{V} . Then there exists a finite open sub-cover of \mathcal{V} and $\eta > 0$, such that for every point $x \in \mathcal{V}$, the η -ball around x is contained in the finite sub-cover.*

Proof of Lemma 10. Since \mathcal{V} is a compact set in the metric space \mathbb{D} (with metric $\|\cdot\|_{\mathbb{D}}$), then any open cover $\{\theta_i : i \in I\}$ of \mathcal{V} has a finite open subcover $\{\tilde{\theta}_i : i = 1, 2, \dots, m\}$ which covers \mathcal{V} .

Let $\Theta = \cup_{i=1}^m \tilde{\theta}_i$. We prove the statement of the lemma by contradiction. Suppose for any $i > 0$, there exists some point $x_i \in \mathbb{D}$ such that $d(x_i, \mathcal{V}) := \inf_{v \in \mathcal{V}} \|x_i - v\|_{\mathbb{D}} < i^{-1}$ and $x_i \notin \Theta$. Then, by compactness of \mathcal{V} there exists $v_i \in \mathcal{V}$ such that $d(x_i, \mathcal{V}) = d(x_i, v_i) < i^{-1}$. Let v_0 be the limit of $\{v_i : i \geq 1\}$. By compactness of \mathcal{V} , $v_0 \in \mathcal{V}$. Since $d(x_i, v_0) \rightarrow 0$ as $i \rightarrow \infty$ and Θ is an open cover of \mathcal{V} , there must be a open ball $B(v_0)$ around v_0 such that $B(v_0) \subseteq \Theta$, which contradicts with $x_i \notin \Theta$, for i large enough. Therefore there must be an η such that the η -ball around any $x \in \mathcal{V}$ is covered by Θ . \square

Proof of Lemma 2. The proof of statement (2) follows directly from the inverse function theorem.

The proof of statement (1) is divided in two steps. Step 1 constructs a finite set of open rectangles that covers the set $\mathcal{M}_{\Delta}(\delta)$ and has certain properties that allow us to apply a change of variable to the derivative of $\delta \mapsto F_{\Delta, \mu}(\delta)$. Step 2 expresses the derivative as an integral on a manifold.

For a subset $\mathcal{S} \subseteq \mathbb{R}^{d_x}$ and $\eta > 0$, define $B_{\eta}(\mathcal{S}) := \{x \in \mathbb{R}^{d_x} : d(x, \mathcal{S}) = \inf_{s \in \mathcal{S}} \|x - s\| < \eta\}$. Similarly, for any $\delta \in \mathbb{R}$ and $\eta > 0$, define $B_{\eta}(\delta) := (\delta - \eta, \delta + \eta)$. Without loss of generality, we assume that $\mathcal{M}_{\Delta}(\delta)$ only has one connected branch. We will discuss the case where $\mathcal{M}_{\Delta}(\delta)$ has multiple connected branches at the end of the proof of this lemma.

Step 1. For any regular value $\delta \in \mathcal{D}$, the set $\mathcal{M}_\Delta(\delta)$ is a $(d_x - 1)$ -manifold in \mathbb{R}^{d_x} of class \mathcal{C}^1 by Lemma 8. Denote $\widetilde{\mathcal{M}}_\Delta(\delta) := \{x \in B(\mathcal{X}) : \Delta(x) = \delta\}$ and $\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta)) := \cup_{\delta' \in B_\eta(\delta)} \widetilde{\mathcal{M}}_\Delta(\delta')$ for $\eta > 0$. These enlargements of the set $\mathcal{M}_\Delta(\delta)$ are used to apply a change of variable technique to integrals on $\mathcal{M}_\Delta(\delta)$.

By Lemma 8 and assumptions S.1-S.2, there exists $\eta_1 > 0$ small enough and $C > c > 0$ such that:

(1) $\overline{B_{\eta_1}(\delta)} := [\delta - \eta_1, \delta + \eta_1] \subseteq \Delta(\overline{\mathcal{X}}) := \{\Delta(x) : x \in \overline{\mathcal{X}}\}$ and contains no critical values of Δ on $\overline{\mathcal{X}}$, and $B_{\eta_1}(\overline{\mathcal{X}}) \subseteq B(\mathcal{X})$.

(2) $\inf_{x \in \overline{\widetilde{\mathcal{M}}_\Delta(B_{\eta_1}(\delta)) \cap B_{\eta_1}(\overline{\mathcal{X}})}} \|\partial\Delta(x)\| > c$.

(3) $\sup_{x \in \overline{\widetilde{\mathcal{M}}_\Delta(B_{\eta_1}(\delta)) \cap B_{\eta_1}(\overline{\mathcal{X}})}} \|\partial\Delta(x)\| < C$.

(4) For any $\eta < \eta_1$, $\widetilde{\mathcal{M}}_\Delta(\delta) \cap B_\eta(\overline{\mathcal{X}})$ is a $(d_x - 1)$ -manifold in \mathbb{R}^{d_x} of class \mathcal{C}^1 .

Indeed, by Lemma 10, the set of regular values is open. Therefore, there exists a small neighborhood $B_\eta(\delta)$ with $\eta > 0$ such that there exists no critical value of Δ on $\overline{\mathcal{X}}$ in $B_\eta(\delta)$. Then any $\eta_1 < \eta$ satisfies statement (1). Statements (2) and (3) follow by the compactness of $\overline{\mathcal{X}}$, the continuity of mapping $x \mapsto \partial\Delta(x)$, and assumptions S.1 and S.2. Statement (4) is implied by Lemma 8.

Next, we establish a finite cover of $\widetilde{\mathcal{M}}_\Delta(B_{\eta_2}(\delta)) \cap B_{\eta_2}(\overline{\mathcal{X}})$ with certain good properties, for some $\eta_2 < \eta_1$.

For any $\eta_3 < \eta_1$, $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta)) \cap B_{\eta_3}(\overline{\mathcal{X}})$ satisfies the properties (2)–(4) stated above. Consider the rectangles $\theta(x) := X_1(x) \times \dots \times X_{d_x}(x)$ centered at $x = (x_1, \dots, x_{d_x})$ where $X_k(x) := (x_k - a_k(x), x_k + a_k(x))$, with $a_k(x) > 0$, $k = 1, 2, \dots, d_x$. Let $A(x) := \sup_{1 \leq k \leq d_x} a_k(x)$ be such that:

$$\overline{\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta)) \cap B_{\eta_3}(\overline{\mathcal{X}})} \subseteq \cup_{x \in \widetilde{\mathcal{M}}_\Delta(\delta) \cap B_{\eta_3}(\overline{\mathcal{X}})} \theta(x) \subseteq \widetilde{\mathcal{M}}_\Delta(B_{\eta_1}(\delta)) \cap B_{\eta_1}(\overline{\mathcal{X}}),$$

which can be fulfilled by using small enough η_3 .

By construction, for small enough $A(x)$ and any $x' \in \theta(x)$, there always exists an index $i(x) \in \{1, 2, \dots, d_x\}$ such that $|\partial_{x_{i(x)}}\Delta(x')| \geq \frac{c}{2\sqrt{d_x}}$ since $\|\partial\Delta(x')\| \geq c$ for all $x' \in \theta(x)$ by the property (2) above, where $\partial_x := \partial/\partial x$. Also we can find a finite set of $\theta(x)$'s, denoted as $\Theta := \{\theta(x^i)\}_{i=1}^m$, such that Θ forms a finite open cover of $\overline{\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta)) \cap B_{\eta_3}(\overline{\mathcal{X}})}$. We rename these open rectangles as $\theta_i := \theta(x^i)$, $i \in \{1, 2, \dots, m\}$, where $\theta_i = X_{i_1} \times \dots \times X_{i_{d_x}}$ and $X_{ik} := X_k(x^i)$, $k \in \{1, \dots, d_x\}$.

For a given $i \in \{1, 2, \dots, m\}$, consider the center of θ_i , denoted as x^i . Without loss of generality, we can assume that $i(x^i) = d_x$. Then, for all $x' \in \theta(x^i)$, $|\partial_{x_{d_x}} \Delta(x')| \geq c/2\sqrt{d_x}$. This means that $\Delta(x)$ is partially monotonic in x_{d_x} on $\theta(x^i)$. By the implicit function theorem, there exists g such that $g(x'_1, x'_2, \dots, x'_{d_x-1}, \delta') = x'_{d_x}$, for any $x' = (x'_1, x'_2, \dots, x'_{d_x}) \in \widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta)) \cap \theta(x^i)$ and $\delta' = \Delta(x')$. Also by implicit function theorem, $\|\partial g(x'_1, \dots, x'_{d_x-1}, \delta')\| \leq \frac{\|\partial \Delta(x')\|}{|\partial_{x_{d_x}} \Delta(x')|} \leq \frac{2C\sqrt{d_x}}{c} := \Lambda$ because $|\partial_{x_{d_x}} \Delta(x')| \geq c/2\sqrt{d_x}$ and $\|\partial \Delta(x')\| \leq C$. Therefore,

$$\begin{aligned} |g(x'_1, x'_2, \dots, x'_{d_x-1}, \delta') - x^i_{d_x}| &= |g(x'_1, x'_2, \dots, x'_{d_x-1}, \delta') - g(x^i_1, x^i_2, \dots, x^i_{d_x-1}, \delta)| \\ &\leq \Lambda C(\sqrt{a_1^2(x^i) + \dots + a_{d_x-1}^2(x^i)} + \eta_3). \end{aligned}$$

We can choose $a_1(x^i) = a_2(x^i) = \dots = a_{d_x-1}(x^i) = \eta_4$ and $a_{d_x}(x^i) = 2(1+\eta_3)C\Lambda(\sqrt{d_x-1}\eta_4 + \eta_3)$, using η_4 small enough. With η_4 small enough, the tube $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta)) \cap \theta_i \subseteq X_{i1} \times \dots \times X_{i,d_x-1} \times (x^i_{d_x} - a_{d_x}(x^i)/2, x^i_{d_x} + a_{d_x}(x^i)/2)$, or geometrically, $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta))$ does not intersect θ_i 's faces except at the ones which are *parallel* to the vector $(0, \dots, 0, 1) \in \mathbb{R}^{d_x}$. In such a case, we say that $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta))$ intersects θ_i at the axis x_{d_x} . More generally, for all $i \in \{1, 2, \dots, m\}$, $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta))$ intersects θ_i at axis $i(x^i)$, where x^i is the center of θ_i . This property implies that g is a well-defined injection from $X_{i1} \times \dots \times X_{i,d_x-1} \times B_{\eta_3}(\delta)$ to $X_{i1} \times \dots \times X_{i,d_x}$, for $i \in \{1, \dots, m\}$, which will allow us to perform a change of variable in the equation (B.25). Such a property holds for any $\eta_2 < \eta_3$.

Step 2. Let η_2 be such that $0 < \eta_2 < \eta_3$. We first apply partition of unity to the open cover $\Theta = \{\theta_i\}_{i=1}^m$ of $\widetilde{\mathcal{M}}_\Delta(B_{\eta_2}(\delta)) \cap B_{\eta_2}(\overline{\mathcal{X}})$ of Step 1.

By Lemma 1, for the finite open cover Θ of the manifold $\widetilde{\mathcal{M}}_\Delta(B_{\eta_2}(\delta)) \cap B_{\eta_2}(\overline{\mathcal{X}})$, we can find a set of C^∞ partition of unity p_j , $1 \leq j \leq J$ on Θ with the properties given in the lemma.

Our main goal is to compute

$$\partial_\delta F_{\Delta, \mu}(\delta) = \lim_{h \rightarrow 0} \frac{F_{\Delta, \mu}(\delta + h) - F_{\Delta, \mu}(\delta)}{h}.$$

Denote $B_\eta^+(\delta) = [\delta, \delta + \eta]$, for any $\delta \in \mathbb{R}$ and $\eta > 0$. Denote $\mathcal{M}_\Delta(B_\eta^+(\delta)) = \cup_{\delta' \in B_\eta^+(\delta)} \mathcal{M}_\Delta(\delta')$, and $\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) = \cup_{\delta' \in B_\eta^+(\delta)} \widetilde{\mathcal{M}}_\Delta(\delta')$.

For any $0 < \eta < \eta_2$, $\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) \subseteq \widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))$. Therefore, the properties (1) to (4) stated in Step 1 are satisfied when we replace $\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))$ by $\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta))$. Note that,

$$\begin{aligned} F_{\Delta,\mu}(\delta + \eta) - F_{\Delta,\mu}(\delta) &= \int_{x \in \mathcal{X}} 1(\delta \leq \Delta(x) \leq \delta + \eta) \mu'(x) dx \\ &= \int_{\mathcal{M}_\Delta(B_\eta^+(\delta))} \mu'(x) dx = \int_{\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta))} \mu'(x) dx = \int_{\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) \cap \Theta} \mu'(x) dx \\ &= \int_{\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) \cap (\cup_{i=1}^m \theta_i)} \mu'(x) \sum_{j=1}^J p_j(x) dx = \sum_{1 \leq i \leq m, 1 \leq j \leq J} \int_{\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) \cap \theta_i} p_j(x) \mu'(x) dx. \end{aligned} \quad (\text{B.24})$$

This third and fourth equalities hold because $\mu'(x) = 0$ for any $x \in \widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) \setminus \mathcal{M}_\Delta(B_\eta^+(\delta))$ and $x \in \widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) \setminus \Theta$, respectively.

For any $i \in \{1, 2, \dots, m\}$, without loss of generality, suppose that $\mathcal{M}_\Delta(B_\eta^+(\delta))$ intersects $\theta_i = X_{i1} \times \dots \times X_{id_x}$ at the x_{d_x} axis. Then, $|\partial_{x_{d_x}} \Delta(x)| \geq c/\sqrt{d_x}$ on θ_i , and we can apply the implicit function theorem to show existence of the \mathcal{C}^1 implicit function $g : X_{i1} \times \dots \times X_{i(d_x-1)} \times B_\eta^+(\delta) \rightarrow X_{id_x}$, such that $\Delta(x_1, \dots, x_{d_x-1}, g(x_1, \dots, x_{d_x-1}, \delta')) = \delta'$ for all $(x_1, \dots, x_{d_x-1}, \delta') \in X_{i1} \times \dots \times X_{i(d_x-1)} \times B_\eta^+(\delta)$. Define the injective mapping ψ_{d_x} as:

$$\begin{aligned} \psi_{d_x} : X_{i1} \times \dots \times X_{i(d_x-1)} \times B_\eta^+(\delta) &\rightarrow X_{i1} \times \dots \times X_{i(d_x-1)} \times X_{id_x}, \\ \psi_{d_x}(x_{-d_x}, \delta') &= (x_{-d_x}, g(x_{-d_x}, \delta')) \text{ for } x_{-d_x} := (x_1, x_2, \dots, x_{d_x-1}). \end{aligned}$$

In equation (B.24), we apply a change of variable defined by the map ψ_{d_x} to the (i, j) -th element of the sum:

$$\begin{aligned} \int_{\theta_i \cap \widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta))} p_j(x) \mu'(x) dx &= \int_{X_{i1} \times \dots \times X_{i(d_x-1)} \times B_\delta^+(\eta)} (p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x}) |\det(D\psi_{d_x})| d\delta' dx_{-d_x} \\ &= \int_{X_{i1} \times \dots \times X_{i(d_x-1)}} \int_{B_\delta^+(\eta)} \frac{(p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x})}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}|} d\delta' dx_{-d_x} \\ &= \eta \int_{X_{i1} \times \dots \times X_{i(d_x-1)}} \frac{(p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x})}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}|} dx_{-d_x} + o(\eta). \end{aligned} \quad (\text{B.25})$$

The second equality follows because

$$D\psi_{d_x}(x_{-d}, \delta) = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \partial_\delta g(x_{-d_x}, \delta) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 1/\partial_{x_{d_x}} \Delta(\tilde{x}) \end{bmatrix},$$

where $\tilde{x} = \psi_{d_x}(x_{-d}, \delta)$.

The last equality follows as $\eta \rightarrow 0$, because by the uniform continuity of

$$(x_{-d_x}, \delta') \mapsto (p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x}) / |\partial_{x_{d_x}} \Delta \circ \psi_{d_x}| \Big|_{(x_{-d_x}, \delta')}$$

over $(x_{-d_x}, \delta') \in X_{i_1} \times \dots \times X_{i(d_x-1)} \times B_\eta^+(\delta)$. In (B.25), the last component of ψ_{d_x} is fixed to be δ without being specified for simplicity. We will maintain this convention in the rest of the proof whenever the variable of integration is x_{-d_x} (excluding x_{d_x}).

Next, we write the first term of (B.25) as an integral on a manifold, which is

$$\eta \int_{X_{i_1} \times \dots \times X_{i(d_x-1)}} \frac{(p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x})}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}|} dx_{-d_x} = \eta \int_{\widetilde{\mathcal{M}}_\Delta(\delta) \cap \theta_i} \frac{p_j(x) \mu'(x)}{\|\partial \Delta(x)\|} d\text{Vol}. \quad (\text{B.26})$$

Summing up over i and j in (B.25) and using Definition 5.5,

$$\sum_{1 \leq i \leq m, 1 \leq j \leq J} \int_{\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) \cap \theta_i} p_j(x) \mu'(x) dx = \eta \int_{\widetilde{\mathcal{M}}_\Delta(\delta) \cap \Theta} \frac{\mu'(x)}{\|\partial \Delta(x)\|} d\text{Vol} + o(\eta). \quad (\text{B.27})$$

Let us explain (B.26). Equation (B.26) is calculated using the following fact: The mapping $\alpha : X_{i_1} \times \dots \times X_{i d_x-1} \rightarrow X_{i_1} \times \dots \times X_{i d_x}$ such that $\alpha(x_1, \dots, x_{d_x-1}) = (x_1, \dots, x_{d_x-1}, g(x_1, \dots, x_{d_x-1}, \delta))$ has Jacobian matrix

$$D\alpha^\top(x_{-d_x}) = \begin{bmatrix} 1 & 0 & \dots & 0 & \partial_{x_1} g(x_{-d_x}) \\ 0 & 1 & \dots & 0 & \partial_{x_2} g(x_{-d_x}) \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 1 & \partial_{x_{d_x-1}} g(x_{-d_x}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 & (\partial_{x_1} \Delta / \partial_{x_{d_x}} \Delta)(\tilde{x}) \\ 0 & 1 & \dots & 0 & (\partial_{x_2} \Delta / \partial_{x_{d_x}} \Delta)(\tilde{x}) \\ \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & 1 & (\partial_{x_{d_x-1}} \Delta / \partial_{x_{d_x}} \Delta)(\tilde{x}) \end{bmatrix},$$

where $\tilde{x} = (x_1, \dots, x_{d_x-1}, g(x_1, \dots, x_{d_x-1}, \delta))$. The volume of $D\alpha$ is $\text{Vol}(D\alpha) = \sqrt{\det(D\alpha^\top D\alpha)}$, where $D\alpha^\top D\alpha = I_{d_x-1} + \partial g \partial g^\top$. By the Matrix Determinant Lemma,

$$\text{Vol}(D\alpha)(x_{-d_x}) = \sqrt{1 + \partial g^\top \partial g} = \|\partial \Delta\| / |\partial_{x_{d_x}} \Delta| \Big|_{x=\tilde{x}}.$$

Hence, the left hand side of equation (B.26) is:

$$\eta \int_{X_{i_1} \times \dots \times X_{i(d_x-1)}} \frac{(p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x})}{\|\partial \Delta \circ \psi_{d_x}\|} \text{Vol}(D\alpha) dx_{-d_x},$$

and it can be further re-expressed as the right side of (B.26) using Definition 5.4.

By equations (B.24) and (B.27),

$$\frac{F_{\Delta, \mu}(\delta + \eta) - F_{\Delta, \mu}(\delta)}{\eta} = \int_{\mathcal{M}_\Delta(\delta)} \frac{\mu'(x)}{\|\partial \Delta(x)\|} d\text{Vol} + o(1), \quad (\text{B.28})$$

where we use that $\mu'(x) = 0$ for all $x \in \widetilde{\mathcal{M}}_\Delta(\delta) \setminus \mathcal{M}_\Delta(\delta)$. Similarly, we can show that

$$\frac{F_{\Delta, \mu}(\delta) - F_{\Delta, \mu}(\delta - \eta)}{\eta} = \int_{\mathcal{M}_\Delta(\delta)} \frac{\mu'(x)}{\|\partial \Delta(x)\|} d\text{Vol} + o(1).$$

Thus, we conclude that $F_{\Delta,\mu}(\delta)$ is differentiable at $\delta \in \mathcal{D}$ with derivative

$$f_{\Delta,\mu}(\delta) := \partial_\delta F_{\Delta,\mu}(\delta) = \int_{\mathcal{M}_\Delta(\delta)} \frac{\mu'(x)}{\|\partial\Delta(x)\|} d\text{Vol}.$$

Finally, if $\mathcal{M}_\Delta(\delta)$ has multiple branches but a finite number of them, we can repeat Step 1 and 2 in the proof above for each individual branch. Since the number of connected branches is finite, the remainders in equation (B.28) converge to 0 uniformly. Thus, adding up the results for all connected branches in equation (B.28), the statements of Lemma 2 hold. \square

B.3. Proofs of Lemmas 3-5.

Lemma 11 (Continuity). *Let f be a measurable function defined on $B_\eta(\mathcal{X}) \subset B(\mathcal{X})$ which vanishes outside \mathcal{X} , where $\eta > 0$ is a constant. Let δ be a regular value of Δ on $\overline{\mathcal{X}}$. Suppose f is continuous on $\widetilde{\mathcal{M}}_\Delta(B_{\eta_1}(\delta)) \cap B_{\eta_1}(\overline{\mathcal{X}})$ for some small η_1 such that $0 < \eta_1 < \eta$. Then, $\delta \mapsto \int_{\mathcal{M}_\Delta(\delta)} f d\text{Vol}$ is continuous on \mathcal{D} .*

Proof. First, we follow Step 1 in the Proof of Lemma 2. Suppose we have a set of open rectangles $\Theta = \{\theta_1, \dots, \theta_m\}$ such that $\widetilde{\mathcal{M}}_\Delta(B_{\eta_2}(\delta)) \cap B_{\eta_2}(\overline{\mathcal{X}}) \subset \cup_{i=1}^m \theta_i \subset \overline{\cup_{i=1}^m \theta_i} \subset \widetilde{\mathcal{M}}_\Delta(B_{\eta_1}(\delta)) \cap B_{\eta_1}(\overline{\mathcal{X}})$ for any $\eta_2 < \eta_3$, where η_3 is a small enough positive number, $\eta_3 < \eta_1$. Moreover, let η_3 be small enough such that all $\delta' \in B_{\eta_3}(\delta)$ are regular values. By compactness of $\overline{\cup_{i=1}^m \theta_i}$, f is bounded and uniformly continuous on $\cup_{i=1}^m \theta_i$.

By construction, θ_i , $i = 1, 2, \dots, m$, satisfies that $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3})$ intersects θ_i at axis $i(\theta_i)$, for any $\eta_2 < \eta_3$.

Then, following Step 2 in the Proof of Lemma 2, there exists a set of C^∞ partition of unity functions $x \mapsto p_j(x)$ of Θ , $j = 1, 2, \dots, J$.

Then, for any $\delta' \in B_{\eta_3}(\delta)$, by the definition of partition of unity,

$$\int_{\mathcal{M}_\Delta(\delta')} f d\text{Vol} = \sum_{1 \leq i \leq m, 1 \leq j \leq J} \int_{\widetilde{\mathcal{M}}_\Delta(\delta') \cap \theta_i} p_j(x) f(x) d\text{Vol}. \quad (\text{B.29})$$

The equation (B.29) holds since $f(x) = 0$ for all $x \notin \mathcal{X}$.

To show that $\int_{\mathcal{M}_\Delta(\delta')} f d\text{Vol}$ converges to $\int_{\mathcal{M}_\Delta(\delta)} f d\text{Vol}$ as δ' converges to δ , it suffices to show that $\int_{\widetilde{\mathcal{M}}_\Delta(\delta') \cap \theta_i} p_j(x) f(x) d\text{Vol}$ converges to $\int_{\widetilde{\mathcal{M}}_\Delta(\delta) \cap \theta_i} p_j(x) f(x) d\text{Vol}$ as δ' converges to δ , for all $i = 1, 2, \dots, m$ and $j = 1, 2, \dots, J$.

Without loss of generality, assume that $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta))$ intersects θ_i at axis $i(\theta_i) = d_x$. Then, there exists constants $c > 0$ and $C > 0$ such that $\partial_{x_{d_x}} \Delta(x) > c$ and $\|\partial \Delta(x)\| < C$ for all $x \in \theta_i$, $i = 1, 2, \dots, m$.

We can apply the implicit function theorem to establish existence of the \mathcal{C}^1 function $g : X_{i1} \times \dots \times X_{i(d_x-1)} \times B_\eta^+(\delta) \rightarrow X_{id_x}$, such that $\Delta(x_1, \dots, x_{d_x-1}, g(x_1, \dots, x_{d_x-1}, \delta')) = \delta'$ for all $(x_1, \dots, x_{d_x-1}, \delta') \in X_{i1} \times \dots \times X_{i(d_x-1)} \times B_\eta(\delta)$. Define the one-to-one mapping ψ_{d_x} as:

$$\psi_{d_x} : X_{i1} \times \dots \times X_{i(d_x-1)} \times B_\eta^+(\delta) \rightarrow X_{i1} \times \dots \times X_{i(d_x-1)} \times X_{id_x},$$

where $\psi_{d_x}(x_{-d_x}, \delta') = (x_{-d_x}, g(x_{-d_x}, \delta'))$ for $x_{-d_x} := (x_1, x_2, \dots, x_{d_x-1})$. Note that ψ_{d_x} and g are both \mathcal{C}^1 functions.

For any δ' such that $|\delta' - \delta| < \eta_3$, by the change of variables we have:

$$\int_{\widetilde{\mathcal{M}}_\Delta(\delta') \cap \theta_i} p_j(x) f(x) d\text{Vol} = \int_{X_1 \times X_2 \times \dots \times X_{d_x-1}} (p_j f) \circ \psi_{d_x}(x_{-d_x}, \delta') \frac{\|\partial \Delta \circ \psi_{d_x}(x_{-d_x}, \delta')\|}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}(x_{-d_x}, \delta')|} dx_{-d_x}. \quad (\text{B.30})$$

Since $|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}(x_{-d_x}, \delta')| = |\partial_{x_{d_x}} \Delta|_{x=\psi_{d_x}(x_1, \dots, x_{d_x-1}, \delta')} > c$ for all $\delta' \in B_{\eta_3}(\delta)$ and $x_{-d_x} \in X_1 \times X_2 \times \dots \times X_{d_x-1}$ and $p_j, f, \partial \Delta$ and $\partial_{x_{d_x}} \Delta$ are uniformly continuous functions on $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta)) \cap B_i$, conclude that the map

$$(p_j f) \circ \psi_{d_x} \frac{\|\partial \Delta \circ \psi_{d_x}\|}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}|} \text{ is uniformly continuous on } X_1 \times \dots \times X_{d_x-1} \times B_{\eta_3}(\delta).$$

Since $X_2 \times \dots \times X_{d_x-1}$ and is bounded, it immediately follows that $\delta' \mapsto \int_{\widetilde{\mathcal{M}}_\Delta(\delta') \cap \theta_i} p_j(x) f(x) d\text{Vol}$ is continuous at $\delta' = \delta$, and hence

$$\delta' \mapsto \int_{\mathcal{M}_\Delta(\delta')} f d\text{Vol} = \sum_{1 \leq i \leq m, 1 \leq j \leq J} \int_{\widetilde{\mathcal{M}}_\Delta(\delta') \cap \theta_i} p_j(x) f(x) d\text{Vol}$$

is continuous at $\delta' = \delta$.

This argument applies to every $\delta \in \mathcal{D}$, and by compactness of \mathcal{D} the continuity claim extends to the entire \mathcal{D} . \square

Proof of Lemma 3. To shows statement (a), for any $G_n \rightarrow G \in \mathbb{F}_0$ under sup-norm such that $\Delta + t_n G_n \in \mathbb{F}$, and $t_n \rightarrow 0$, we consider

$$\frac{F_{\Delta+t_n G_n, \mu}(\delta) - F_{\Delta, \mu}(\delta)}{t_n}.$$

By assumption, any function $G \in \mathbb{F}_0$ is bounded and uniformly continuous on $B(\mathcal{X})$. Hence, G_n is uniformly bounded for $n \geq N$, since $G_n \rightarrow G$ in sup-norm.

For any $\delta \in \mathcal{D}$ we consider a procedure similar to Lemma 2. We use the same notation as in Step 1 of the proof of Lemma 2. Suppose for $\eta_1 > 0$ small enough, we have a rectangle cover $\Theta = \cup_{i=1}^m \theta_i \subseteq B(\mathcal{X})$ of $\widetilde{\mathcal{M}}_\Delta(B_{\eta_1}(\delta)) \cap B_{\eta_1}(\overline{\mathcal{X}})$ such that for all $\eta < \eta_1$, $\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))$ intersects each θ_i at some axis $i(\theta_i)$, $1 \leq i \leq m$. As before, there is a partition of unity $\{p_j\}_{j=1}^J$ on the cover sets $\Theta = \{\theta_i\}_{i=1}^m$. As in the proof of Lemma 2, we can rewrite

$$\begin{aligned} & \frac{\int_{\mathcal{X}} [1\{\Delta(x) + t_n G_n(x) \leq \delta\} - 1\{\Delta(x) \leq \delta\}] \mu'(x) dx}{t_n} \\ &= \sum_{1 \leq i \leq m, 1 \leq j \leq J} \int_{\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) \cap \theta_i} p_j(x) \frac{[1\{\Delta(x) + t_n G_n(x) \leq \delta\} - 1\{\Delta(x) \leq \delta\}] \mu'(x)}{t_n} dx. \end{aligned}$$

Then, for any fixed positive number ζ , there exist N large enough such that $\sup_{x \in B(\mathcal{X}), n \geq N} |G_n - G| < \zeta$. Moreover, for any $x \in B(\mathcal{X})$, and large enough n ,

$$1\{\Delta(x) + t_n G_n(x) \leq \delta\} \leq 1\{\Delta(x) + t_n(G(x) - \zeta) \leq \delta\}.$$

As in Step 2 of the proof of Lemma 2, suppose $\theta_i = X_{i1} \times \dots \times X_{i d_x}$ intersects $\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))$ at $i(\theta_i) = x_{d_x}$. Define the parametrization

$$\psi_{d_x} : X_{i1} \times \dots \times X_{i, d_x-1} \times B_\eta(\delta) \mapsto \theta_i,$$

$$\psi_{d_x}(x_{-d_x}, \delta') = (x_{-d_x}, g(x_{-d_x}, \delta')),$$

where $g(x_{-d_x}, \delta')$ is the implicit function derived from equation $\Delta(x) = \delta'$, for any $\delta' \in B_\eta(\delta)$.

Therefore, for large enough n ,

$$\begin{aligned} & \int_{\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) \cap \theta_i} p_j(x) \frac{[1\{\Delta(x) + t_n G_n(x) \leq \delta\} - 1\{\Delta(x) \leq \delta\}] \mu'(x)}{t_n} dx \\ & \leq \frac{\int_{\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta)) \cap \theta_i} [1\{\Delta(x) + t_n(G(x) - \zeta) \leq \delta\} - 1\{\Delta(x) \leq \delta\}] \mu'(x) dx}{t_n}. \end{aligned}$$

Next, by a change of variables $\psi_{d_x}^{-1}$ from θ_i to $X_{i1} \times \dots \times X_{i, d_x-1} \times B_\eta(\delta)$,

$$\begin{aligned} & \int_{\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta)) \cap \theta_i} p_j(x) \frac{1\{\delta \leq \Delta(x) \leq \delta - t_n(G(x) - \zeta)\} \mu'(x)}{t_n} dx \\ &= \int_{X_{i1} \times \dots \times X_{i, d_x-1}} \int_{B_\eta(\delta)} \frac{(p_j \cdot \mu') \circ \psi_{d_x}(x_{-d_x}, \delta') 1\{\delta \leq \delta' \leq \delta - t_n(G \circ \psi_{d_x}(x_{-d_x}, \delta) - \zeta)\}}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}(x_{-d_x}, \delta')| t_n} d\delta' dx_{-d_x} \\ &= \int_{X_{i1} \times \dots \times X_{i, d_x-1}} \int_{B_\eta(\delta) \cap [\delta, \delta - t_n(G \circ \psi_{d_x}(x_{-d_x}, \delta) - \zeta)]} \frac{(p_j \cdot \mu') \circ \psi_{d_x}(x_{-d_x}, \delta')}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}(x_{-d_x}, \delta')| t_n} d\delta' dx_{-d_x} \\ & \leq - \int_{X_{i1} \times \dots \times X_{i, d_x-1}} \frac{(p_j \cdot \mu') \circ \psi_{d_x}(x_{-d_x}, \delta)}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}(x_{-d_x}, \delta)|} (G \circ \psi_{d_x}(x_{-d_x}, \delta) - \zeta) dx_{-d_x} + o(\eta) \end{aligned}$$

$$\begin{aligned}
&= - \int_{\theta_i \cap \widetilde{\mathcal{M}}_\Delta(\delta)} p_j(x) \mu'(x) \frac{G(x) - \zeta}{\|\partial\Delta(x)\|} d\text{Vol} + o(\eta) \\
&= - \int_{\theta_i \cap \mathcal{M}_\Delta(\delta)} p_j(x) \mu'(x) \frac{G(x) - \zeta}{\|\partial\Delta(x)\|} d\text{Vol} + o(\eta),
\end{aligned}$$

where the inequality in the above equation holds by continuity of $(x_{-d_x}, \delta') \mapsto (p_j \cdot \mu') \circ \psi_{d_x}(x_{-d_x}, \delta') / |\partial_{x_{d_x}} \Delta \circ \psi_{d_x}(x_{-d_x}, \delta')|$. More specifically, fixing $\eta > 0$ and x_{-d_x} , for $t_n \rightarrow 0$,

$$B_\eta(\delta) \cap [\delta, \delta - t_n(G \circ \psi_{d_x}(x_{-d_x}, \delta) - \zeta)] = [\delta, \delta - t_n(G \circ \psi_{d_x}(x_{-d_x}, \delta) - \zeta)]$$

and

$$\frac{(p_j \cdot \mu') \circ \psi_{d_x}(x_{-d_x}, \delta')}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}(x_{-d_x}, \delta')|} \rightarrow \frac{(p_j \cdot \mu') \circ \psi_{d_x}(x_{-d_x}, \delta)}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}(x_{-d_x}, \delta)|}$$

as $\delta' \rightarrow \delta$. The last equality above holds because $\mu'(x) = 0$ for all $x \in \widetilde{\mathcal{M}}_\Delta(\delta) \setminus \mathcal{M}_\Delta(\delta)$.

Since m and J are fixed for any $n \geq N$, and $|G \circ \psi_{d_x}(x_{-d_x}, \delta) - \zeta|$ is bounded by some absolute constant, $\sum_j p_j(x) = 1$ and $p_j(x) \geq 0$, we can let $\zeta \rightarrow 0$ to conclude that:

$$\lim_{n \rightarrow \infty} \frac{F_{\Delta+t_n G_n, \mu}(\delta) - F_{\Delta, \mu}(\delta)}{t_n} \leq \sum_{i=1}^m \sum_{j=1}^J - \int_{\theta_i \cap \mathcal{M}_\Delta(\delta)} p_j(x) \mu'(x) \frac{G(x)}{\|\partial\Delta(x)\|} d\text{Vol}.$$

The right side is given by:

$$- \int_{\mathcal{M}_\Delta(\delta)} \frac{\mu'(x) G(x)}{\|\partial\Delta(x)\|} d\text{Vol}.$$

Working from other direction using the same approaches as above with $\zeta < 0$ yields

$$\lim_{n \rightarrow \infty} \frac{F_{\Delta+t_n G_n, \mu}(\delta) - F_{\Delta, \mu}(\delta)}{t_n} \geq - \int_{\mathcal{M}_\Delta(\delta)} \frac{\mu'(x) G(x)}{\|\partial\Delta(x)\|} d\text{Vol}.$$

Combining the two inequalities, we conclude that $F_{\Delta, \mu}(\delta)$ is Hadamard-differentiable at Δ tangentially to \mathbb{F}_0 with derivative

$$\partial_\Delta F_{\Delta, \mu}(\delta)[G] = - \int_{\mathcal{M}_\Delta(\delta)} \frac{\mu'(x) G(x)}{\|\partial\Delta(x)\|} d\text{Vol}.$$

To show that the result holds uniformly in $\delta \in \mathcal{D}$, we use the equivalence between uniform convergence and continuous convergence (e.g., Resnick (1987, p.2)). Take a sequence δ_n in \mathcal{D} that converges to $\delta \in \mathcal{D}$. Then, the preceding argument applies to this sequence and $\partial_\Delta F_{\Delta, \mu}(\delta_n)[G] \rightarrow \partial_\Delta F_{\Delta, \mu}(\delta)[G]$ by uniform continuity of $\delta \mapsto \partial_\Delta F_{\Delta, \mu}(\delta)[G]$ on \mathcal{D} , which holds by Lemma 11 because G , μ' , and $\|\partial\Delta\|$ are continuous on $\overline{\mathcal{X}}$ and \mathcal{D} excludes neighborhoods of the critical values of Δ in $\overline{\mathcal{X}}$.

To show statement (b), note that by statement (a), Hadamard differentiability of the quantile map, see e.g. Lemma 3.9.20 in van der Vaart and Wellner (1996), and the chain

rule for Hadamard differentiation, the inverse map $\Delta_\mu^*(u)$ is Hadamard differentiable at Δ tangentially to \mathbb{F}_0 with the derivative map

$$\partial_\Delta \Delta_\mu^*(u)[G] = - \left. \frac{\partial_\Delta F_{\Delta,\mu}(\delta)[G]}{\partial_\delta F_{\Delta,\mu}(\delta)} \right|_{\delta=\Delta_\mu^*(u)} = \frac{\partial_\Delta F_{\Delta,\mu}(\Delta_\mu^*(u))[G]}{f_{\Delta,\mu}(\Delta_\mu^*(u))},$$

uniformly in the index $u \in \mathcal{U} = \{u \in (0, 1) : \Delta_\mu^*(u) \in \mathcal{D}, f_{\Delta,\mu}(\Delta_\mu^*(u)) \geq \varepsilon\}$. \square

Proof of Lemma 4. Statement (a):

Suppose that there are sequences of operators H_n and scalars t_n , such that $t_n \rightarrow 0$ as $n \rightarrow \infty$, $\mu + t_n H_n \in \mathbb{H}_0$, and $H_n \rightarrow H \in \mathbb{H}_0$ as $n \rightarrow \infty$, i.e.,

$$\sup_{g \in \mathcal{G}} |[H_n - H](g)| \rightarrow 0.$$

Let $\mu_n = \mu + t_n H_n$. Then, $F_{\Delta,\mu_n}(\delta) - F_{\Delta,\mu}(\delta) = [\mu_n - \mu](g_{\Delta,\delta}) = t_n H_n(g_{\Delta,\delta}) = t_n H(g_{\Delta,\delta}) + t_n [H_n - H](g_{\Delta,\delta})$. By the assumption that $H_n \rightarrow H$,

$$\frac{F_{\Delta,\mu_n}(\delta) - F_{\Delta,\mu}(\delta)}{t_n} = H(g_{\Delta,\delta}) + o(1).$$

Hence, $F_{\Delta,\mu}(\delta)$ is Hadamard differentiable at μ tangentially to \mathbb{H}_0 , with the derivative map $\partial_\mu F_{\Delta,\mu}(\delta)[H] = H(g_{\Delta,\delta})$.

To show that the result holds uniformly in $\delta \in \mathcal{D}$, we use the equivalence between uniform convergence and continuous convergence (e.g., Resnick (1987, p.2)). Take a sequence δ_n in \mathcal{D} that converges to $\delta \in \mathcal{D}$. Then, the preceding argument applies to this sequence and we conclude that

$$\frac{F_{\Delta,\mu_n}(\delta_n) - F_{\Delta,\mu}(\delta_n)}{t_n} = H(g_{\Delta,\delta_n}) + o(1).$$

Moreover, $H(g_{\Delta,\delta_n}) \rightarrow H(g_{\Delta,\delta})$ since $g_{\Delta,\delta_n}(X) = 1(\Delta(X) \leq \delta_n) \rightarrow g_{\Delta,\delta}(X) = 1(\Delta(X) \leq \delta)$ in the $L^2(\mu)$ norm, since $\Delta(X)$ has an absolutely continuous distribution, and we require the operator H to be continuous under the $L^2(\mu)$ norm.

Statement (b) follows by statement (a), the Hadamard differentiability of the quantile map uniformly in the quantile index, and the chain rule for Hadamard differentiation; see van der Vaart and Wellner (1996). \square

Proof of Lemma 5. To show Statement (a), Consider $t_n \rightarrow 0$ and $(G_n, H_n) \rightarrow (G, H) \in \mathbb{D}_0 := \mathbb{F}_0 \times \mathbb{H}_0$ as $n \rightarrow \infty$, such that $(\Delta + t_n G_n, \mu + t_n H_n) \in \mathbb{D}$. Let $\Delta_n := \Delta + t_n G_n$ and $\mu_n := \mu + t_n H_n$. Then, we can decompose

$$F_{\Delta_n,\mu_n}(\delta) - F_{\Delta,\mu}(\delta) = [F_{\Delta_n,\mu_n}(\delta) - F_{\Delta_n,\mu}(\delta)] + [F_{\Delta_n,\mu}(\delta) - F_{\Delta,\mu}(\delta)].$$

By Lemma 3,

$$\frac{F_{\Delta_n, \mu}(\delta) - F_{\Delta, \mu}(\delta)}{t_n} = - \int_{\mathcal{M}_{\Delta}(\delta)} \frac{G(x)\mu'(x)}{\|\partial\Delta(x)\|} d\text{Vol} + o(1).$$

By definition of $F_{\Delta_n, \mu_n}(\delta)$,

$$\frac{F_{\Delta_n, \mu_n}(\delta) - F_{\Delta_n, \mu}(\delta)}{t_n} = H_n(g_{\Delta_n, \delta}).$$

Note that

$$H_n(g_{\Delta_n, \delta}) - H(g_{\Delta, \delta}) = [H_n(g_{\Delta_n, \delta}) - H_n(g_{\Delta, \delta})] + [H_n - H](g_{\Delta, \delta}).$$

The second term goes to 0 by the assumption $H_n \rightarrow H$ in \mathbb{H} . For the first term, we further decompose

$$|H_n(g_{\Delta_n, \delta}) - H_n(g_{\Delta, \delta})| \leq |H_n(g_{\Delta_n, \delta}) - H(g_{\Delta_n, \delta})| + |H_n(g_{\Delta, \delta}) - H(g_{\Delta, \delta})| + |H(g_{\Delta_n, \delta}) - H(g_{\Delta, \delta})|.$$

The first two terms go to 0 by $\|H_n - H\|_{\mathcal{G}} \rightarrow 0$. Moreover, $H(g_{\Delta_n, \delta}) \rightarrow H(g_{\Delta, \delta})$ because $g_{\Delta_n, \delta}(X) = 1(\Delta_n(X) \leq \delta) \rightarrow g_{\Delta, \delta}(X) = 1(\Delta(X) \leq \delta)$ in the $L^2(\mu)$ norm, since $\Delta_n \rightarrow \Delta$ in the sup norm and $\Delta(X)$ has an absolutely continuous distribution, and since we require the operator H to be continuous under the $L^2(\mu)$ norm.

We conclude that for any $\delta \in \mathcal{D}$,

$$\frac{F_{\Delta_n, \mu_n}(\delta) - F_{\Delta, \mu}(\delta)}{t_n} \rightarrow - \int_{\mathcal{M}_{\Delta}(\delta)} \frac{G(x)\mu'(x)}{\|\partial\Delta(x)\|} d\text{Vol} + H(g_{\Delta, \delta}) = \partial_{\Delta, \mu} F_{\Delta, \mu}(\delta)[G, H].$$

By an argument similar to the proof of Lemma 3, it can be shown that the convergence is uniform in $\delta \in \mathcal{D}$.

Statement (b) follows by statement (a) and the Hadamard differentiability of the quantile map uniformly in the quantile index, see, e.g., Lemma 3.9.20 in van der Vaart and Wellner (1996). \square

APPENDIX C. PROOFS OF SECTION 6

We first recall Theorem 3.9.4 of van der Vaart and Wellner (1996).

Lemma 12 (Delta-method). *Let \mathbb{D} and \mathbb{E} be metrizable topological vector spaces, and Θ is a compact subset of a metric space. Let $\phi_\theta : \mathbb{D}_\phi \subseteq \mathbb{D} \rightarrow \mathbb{E}$ be a Hadamard differentiable mapping uniformly in $\theta \in \Theta$ at $f \in \mathbb{D}$ tangentially to $\mathbb{D}_0 \subseteq \mathbb{D}$, with derivative $\partial_f \phi_\theta$. Let $\hat{f}_n : \Omega_n \rightarrow \mathbb{D}_\phi$ be stochastic maps taking values in \mathbb{D}_ϕ such that $r_n(\hat{f}_n - f) \rightsquigarrow J_\infty$ for some sequence of constants $r_n \rightarrow \infty$, where J_∞ is separable and takes values in \mathbb{D}_0 . Then $r_n(\phi_\theta(\hat{f}_n) - \phi_\theta(f)) \rightsquigarrow \partial_f \phi_\theta[J_\infty]$, as a stochastic process indexed by $\theta \in \Theta$.*

Proof of Theorem 1. The statements follow directly from Lemma 3, and Lemma 12, by setting $\phi_\theta = F_{\Delta, \mu}(\delta)$ with $\theta = \delta$ or $\phi_\theta = \Delta_\mu^*(u)$ with $\theta = u$, $\mathbb{D}_\phi = \mathbb{D} = \mathbb{F}$, $\mathbb{E} = \mathbb{R}$, $\mathbb{D}_0 = \mathbb{F}_0$, $f = \Delta$, $\widehat{f}_n = \widehat{\Delta}$, and $J_\infty = G_\infty$. The expression of $\partial_f \phi_\theta$ for each statement is the Hadamard derivative in the corresponding statement of Lemma 3. \square

Proof of Lemma 6. Since \mathcal{V} is a union of finite number of closed intervals, for any $\zeta > 0$, we can construct a collection of closed intervals $\mathcal{I} := \{[a_i, b_i] : i = 1, 2, \dots, r\}$ such that: (1) $|b_i - a_i| < \zeta$, (2) $[a_i, b_i] \subset \mathcal{V}$, (3) $\cup_{i=1}^r [a_i, b_i] = \mathcal{V}$, (4) $a_i \leq b_i \leq a_{i+1} \leq b_{i+1}$, for all $i = 1, 2, \dots, r-1$, and (5) $r \leq \frac{C_0}{\zeta}$, where C_0 is a constant.

Using S.1 and S.2 and the assumptions of the Lemma, there exists $\eta > 0$ small enough such that the following conditions hold:

(1) There exist constants c and C such that $\|\partial \Delta(x)\| \leq C$ for all $x \in \overline{\mathcal{X}}$ and $\|\partial \Delta(x)\| \geq c$ in $\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))$ for some small $\eta > 0$ and all $\delta \in \mathcal{D}$.

(2) Uniformly in $\widetilde{\Delta} \in \mathcal{F}$,

$$\frac{c}{2} \leq \inf_{x \in \widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))} \|\partial \widetilde{\Delta}(x)\| \leq \sup_{x \in \widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))} \|\partial \widetilde{\Delta}(x)\| \leq \frac{c}{2} + C.$$

Moreover, using arguments similar to those used to show Lemma 2, we can verify that:

(3) Uniformly in $\widetilde{\Delta} \in \mathcal{F}$, uniformly in $\delta \in \mathcal{V}$,

$$f_{\widetilde{\Delta}, \mu}(\delta) = \int_{\mathcal{M}_{\widetilde{\Delta}}(\delta)} \frac{\mu'(x)}{\|\partial \widetilde{\Delta}(x)\|} d\text{Vol} < K_1,$$

for some finite constant K_1 .

Define the norm $\|g\|_{2, \mu}^2 := \int_{\mathcal{X}} g(x)^2 \mu'(x) dx$. For $\eta > 0$ small enough, for any $\delta \in \mathcal{V}$ and $\widetilde{\Delta} \in \mathcal{F}$,

$$\|1(\widetilde{\Delta} \leq \delta) - 1(\widetilde{\Delta} \leq \delta + \eta)\|_{2, \mu}^2 = \int_{x \in \mathcal{X}} 1(\delta \leq \widetilde{\Delta}(x) \leq \delta + \eta) \mu'(x) dx = \int_{\delta' \in B_\eta^+(\delta)} f_{\widetilde{\Delta}, \mu}(\delta') d\delta' \leq K_1 \eta.$$

Similarly, $\|1(\widetilde{\Delta} \leq \delta) - 1(\widetilde{\Delta} \leq \delta - \eta)\|_{2, \mu}^2 \leq K_1 \eta$.

Let $B_{\zeta, \infty}(\Delta_1), \dots, B_{\zeta, \infty}(\Delta_{q_\zeta})$ be a set of ζ -balls centered at $\Delta_1, \dots, \Delta_{q_\zeta}$ under sup norm that covers \mathcal{F} , where $q_\zeta = N(\zeta, \mathcal{F}, \|\cdot\|_\infty)$. Then, $[\Delta_j - \zeta, \Delta_j + \zeta]$ are covering brackets of \mathcal{F} , $j = 1, 2, \dots, q_\zeta$. For any $\widetilde{\Delta} \in [\Delta_j - \zeta, \Delta_j + \zeta]$ and $\delta \in [a_i, b_i]$, $i = 1, 2, \dots, r$, then the bracket $[1(\Delta_j + \zeta \leq a_i), 1(\Delta_j - \zeta \leq b_i)]$ covers $1(\widetilde{\Delta} \leq \delta)$. For ζ small enough, the size of the bracket

$[1(\Delta_j + \zeta \leq a_i), 1(\Delta_j - \zeta \leq b_i)]$ under the norm $\|\cdot\|_{2,\mu}$ is:

$$\|1(\Delta_j + \zeta \leq a_i) - 1(\Delta_j - \zeta \leq b_i)\|_{2,\mu}^2 = \|1(\Delta_j \leq b_i + \zeta) - 1(\Delta_j \leq a_i - \zeta)\|_{2,\mu}^2 \leq 3K_1\zeta,$$

since $|b_i - a_i| < \zeta$ by construction. Therefore, for ζ small enough, $\{[1(\Delta_j + \zeta \leq a_i), 1(\Delta_j - \zeta \leq b_i)] : j = 1, 2, \dots, q_\zeta, i = 1, 2, \dots, r\}$, form a set of $\sqrt{3K_1\zeta}$ -brackets under the norm $\|\cdot\|_{2,\mu}$ that covers \mathcal{G} . The total number of brackets is $rq_\zeta \leq \frac{C_0}{\zeta} N(\zeta, \mathcal{F}, \|\cdot\|_\infty)$. Or equivalently, for ζ small enough,

$$N_{[]}(\zeta, \mathcal{G}, \|\cdot\|_{2,\mu}) \leq \frac{3K_1C_0}{\zeta^2} N(\zeta^2/(3K_1), \mathcal{F}, \|\cdot\|_\infty).$$

Then by assumption,

$$\begin{aligned} \int_0^1 \sqrt{\log(N_{[]}(\zeta, \mathcal{G}, \|\cdot\|_{2,\mu}))} d\zeta &\leq \int_0^1 \sqrt{\log\left(\frac{3K_1C_0}{\zeta^2} N(\zeta^2/(3K_1), \mathcal{F}, \|\cdot\|_\infty)\right)} d\zeta \\ &\lesssim \int_0^1 \sqrt{\log\left(\frac{3K_1C_0}{\zeta^2}\right)} d\zeta + \int_0^1 \sqrt{\log(N(\zeta^2/(3K_1), \mathcal{F}, \|\cdot\|_\infty))} d\zeta < \infty. \end{aligned}$$

We conclude that \mathcal{G} is μ -Donsker by Donsker theorem (van der Vaart, 1998, Theorem 19.5). \square

Proof of Theorem 2. The statements follow directly from Lemma 5, and Lemma 12, by setting $\phi_\theta = F_{\Delta,\mu}(\delta)$ with $\theta = \delta$ or $\phi_\theta = \Delta_\mu^*(u)$ with $\theta = u$, $\mathbb{D}_\phi = \mathbb{D} = \mathbb{F} \times \mathbb{H}$, $\mathbb{E} = \mathbb{R}$, $\mathbb{D}_0 = \mathbb{F}_0 \times \mathbb{H}_0$, $f = (\Delta, \mu)$, $\hat{f}_n = (\hat{\Delta}, \hat{\mu})$, and $J_\infty = (G_\infty, H_\infty)$. The expression of $\partial_f \phi_\theta$ for each statement is the Hadamard derivative in the corresponding statement of Lemma 5. \square

To prove Theorem 3, we recall Theorem 3.9.11 of van der Vaart (1998). Here we use the notation for bootstrap convergence $\rightsquigarrow_{\mathbb{P}}$ introduced in Section 4.6.

Lemma 13 (Delta-method for bootstrap in probability). *Let \mathbb{D} and \mathbb{E} be metrizable topological vector spaces, and Θ is a compact subset of a metric space. Let $\phi_\theta : \mathbb{D}_\phi \subseteq \mathbb{D} \mapsto \mathbb{E}$ be a Hadamard-differentiable mapping uniformly in $\theta \in \Theta$ at f tangentially to \mathbb{D}_0 with derivative $\partial_f \phi_\theta$. Let \hat{f}_n be a random element such that $r_n(\hat{f}_n - f) \rightsquigarrow J_\infty$. Let \tilde{f}_n be a stochastic map in \mathbb{D} , produced by a bootstrap method, such that $r_n(\tilde{f}_n - \hat{f}_n) \rightsquigarrow_{\mathbb{P}} J_\infty$. Then, $r_n(\phi_\theta(\tilde{f}_n) - \phi_\theta(\hat{f}_n)) \rightsquigarrow_{\mathbb{P}} \partial_f \phi_\theta[J_\infty]$, as a stochastic process indexed by $\theta \in \Theta$.*

Proof of Theorem 3. The statements follow directly from Lemma 5, and Lemma 13, by setting $\phi_\theta = \Delta_\mu^*(u)$, $\theta = u$, $\mathbb{D}_\phi = \mathbb{D} = \mathbb{F} \times \mathbb{H}$, $\mathbb{E} = \mathbb{R}$, $\mathbb{D}_0 = \mathbb{F}_0 \times \mathbb{H}_0$, $f = (\Delta, \mu)$, $\hat{f}_n = (\hat{\Delta}, \hat{\mu})$, and $J_\infty = (G_\infty, H_\infty)$. The expression of $\partial_f \phi_\theta$ is the Hadamard derivative in statement (b) of Lemma 5. \square

APPENDIX D. PROOFS OF APPENDIX A

Proof of Lemma 6. Note that $F_{\Delta,\mu}(\delta) = \sum_{d \in \mathcal{X}_d} \pi_d(d) \int_{c \in \mathcal{X}_d} 1(\Delta(d, c) \leq \delta) \mu'_{c|d}(c) dc$. Given the results of Lemma 2, for each d ,

$$\partial_\delta \int_{\mathcal{X}_{c|d}} 1(\Delta(d, c) \leq \delta) \mu'_{c|d}(c) dc = \int_{\mathcal{M}_{\Delta|d}(\delta)} \frac{\mu'_{c|d}(c)}{\|\partial_c \partial(d, c)\|} d\text{Vol}.$$

Therefore, averaging over $d \in \mathcal{X}_d$,

$$f_{\Delta,\mu}(\delta) := \partial_\delta F_{\Delta,\mu}(\delta) = \sum_{d \in \mathcal{X}_d} \pi_d(d) \int_{\mathcal{M}_{\Delta|d}(\delta)} \frac{\mu'_{c|d}(c)}{\|\partial_c \Delta(d, c)\|} d\text{Vol},$$

where we use that \mathcal{X}_d is a finite set.

Next we prove the statements (1) and (2). Let $G_n \in \mathbb{F}$ and $H_n \in \mathbb{H}_0$ such that $G_n \rightarrow G \in \mathbb{F}_0$ and $H_n \rightarrow H \in \mathbb{H}_0$. Let $\Delta_n = \Delta + t_n G_n$ and $\mu_n = \mu + t_n H_n$, where $t_n \rightarrow 0$ as $n \rightarrow \infty$.

As in the proof of Lemma 5, we decompose

$$F_{\Delta_n, \mu_n}(\delta) - F_{\Delta, \mu}(\delta) = [F_{\Delta_n, \mu_n}(\delta) - F_{\Delta_n, \mu}(\delta)] + [F_{\Delta_n, \mu}(\delta) - F_{\Delta, \mu}(\delta)].$$

Applying the same argument as in the proof of Lemma 3 to each d and averaging over $d \in \mathcal{X}_d$, for any $\delta \in \mathcal{D}$

$$\frac{F_{\Delta_n, \mu}(\delta) - F_{\Delta, \mu}(\delta)}{t_n} = - \sum_{d \in \mathcal{X}_d} \mu_d(d) \int_{\mathcal{M}_{\Delta|d}(\delta)} \frac{G(d, c) \mu'_{c|d}(c)}{\|\partial_c \Delta(d, c)\|} d\text{Vol} + o(1),$$

where we use that \mathcal{X}_d is a finite set. By assumption S.4' and a similar argument to the proof of Lemma 5,

$$\frac{F_{\Delta_n, \mu_n}(\delta) - F_{\Delta_n, \mu}(\delta)}{t_n} = H(g_{\Delta, \delta}) + o(1), \quad g_{\Delta, \delta}(c, d) = 1\{\Delta(c, d) \leq \delta\}$$

We conclude that for any $\delta \in \mathcal{D}$,

$$\frac{F_{\Delta_n, \mu_n}(\delta) - F_{\Delta, \mu}(\delta)}{t_n} \rightarrow - \sum_{d \in \mathcal{X}_d} \mu_d(d) \int_{\mathcal{M}_{\Delta|d}(\delta)} \frac{G(d, c) \mu'_{c|d}(c)}{\|\partial_c \Delta(d, c)\|} d\text{Vol} + H(g_{\Delta, \delta}) = \partial_{\Delta, \mu} F_{\Delta, \mu}(\delta)[G, H].$$

By an argument similar to the proof of Lemma 3, it can be shown that the convergence is uniform in $\delta \in \mathcal{D}$. This shows statement (1). \square

Statement (2) follows from statement (1) and Theorem 3.9.20 of van der Vaart and Wellner (1996) for inverse maps, using an argument analogous to the proof of statement (b) in Lemma 3. \square

Proof of Theorem 5. The result follows from Lemma 6 and Lemma 10. \square

REFERENCES

- ANGRIST, J. D., AND J.-S. PISCHKE (2008): *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND A. GALICHON (2009): "Improving point and interval estimators of monotone functions by rearrangement," *Biometrika*, 96(3), 559–575.
- CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND A. GALICHON (2010): "Quantile and probability curves without crossing," *Econometrica*, 78(3), 1093–1125.
- CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, S. HODERLEIN, H. HOLZMANN, AND W. NEWEY (2015): "Nonparametric identification in panels using quantiles," *J. Econometrics*, 188(2), 378–392.
- CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND B. MELLY (2013): "Inference on counterfactual distributions," *Econometrica*, 81(6), 2205–2268.
- KIM, J., AND D. POLLARD (1990): "Cube Root Asymptotics," *Ann. Statist.*, 18(1), 191–219.
- KOENKER, R., AND G. BASSET (1978): "Regression Quantiles," *Econometrica*, 46(1), 33–50.
- MULLIGAN, C. B., AND Y. RUBINSTEIN (2008): "Selection, Investment, and Women's Relative Wages Over Time," *The Quarterly Journal of Economics*, 123(3), 1061–1110.
- MUNKRES, J. R. (1991): *Analysis on manifolds*. Addison-Wesley Publishing Company, Advanced Book Program, Redwood City, CA.
- MUNNELL, A. H., G. M. B. TOOTELL, L. E. BROWNE, AND J. MCEANEY (1996): "Mortgage Lending in Boston: Interpreting HMDA Data," *The American Economic Review*, 86(1), pp. 25–53.
- PRÆSTGAARD, J., AND J. A. WELLNER (1993): "Exchangeably weighted bootstraps of the general empirical process," *Ann. Probab.*, 21(4), 2053–2086.
- SASAKI, Y. (2015): "What do quantile regressions identify for general structural functions?," *Econometric Theory*, 31(5), 1102–1116.
- SPIVAK, M. (1965): *Calculus on manifolds. A modern approach to classical theorems of advanced calculus*. W. A. Benjamin, Inc., New York-Amsterdam.
- STOCK, J., AND M. WATSON (2011): *Introduction to Econometrics (3rd edition)*. Addison Wesley Longman.
- VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press.
- VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak Convergence and Empirical Processes*. Springer Series in Statistics.
- WOOLDRIDGE, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*. Cambridge, Massachusetts: The MIT Press, second edn.