SUPPLEMENT TO "THE SORTED EFFECTS METHOD: DISCOVERING HETEROGENEOUS EFFECTS BEYOND THEIR AVERAGES"

VICTOR CHERNOZHUKOV, IVAN FERNANDEZ-VAL, YE LUO

Abstract. The supplementary material contains 7 appendices with additional results and some omitted proofs. Appendix C introduces some notation. Appendix D includes a brief review of differential geometry. Appendix E gathers the proofs of the key mathematical results in Appendix A. Appendix F provides sufficient conditions for the μ-Donsker properties in Section 4. Appendix G extends the theoretical analysis to include discrete covariates. Appendices H and I report the results of 3 numerical simulations and an empirical application to the effect of race on mortgage denials, respectively.

Appendix C. Notation

For a possibly multivariate random variable X, $\mathcal{X}$ denotes the interior of the support of X in the part of the population of interest, μ denotes the distribution of X over $\mathcal{X}$, and $\widehat{\mu}$ denotes an estimator of μ. We denote the expectation with respect to the distribution $\widetilde{\mu}$ by $\mathrm{E}_{\widetilde{\mu}}$. We denote the PE as $\Delta(x)$, the empirical PE as $\widehat{\Delta}(x)$, 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 $v = (v_1, \ldots, v_{d_v}) \in \mathbb{R}^{d_v}$, $\|v\|$ denotes the Euclidian norm of v, that is $\|v\| = \sqrt{v^{\mathsf{T}}v}$, where the superscript $^{\mathsf{T}}$ 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 $\to_{\mathrm{P}}$ denotes convergence in (outer) probability.

Appendix D. Background on Differential Geometry

We recall some definitions from differential geometry that are used in the analysis. For a continuously differentiable function $\Delta : B(\mathcal{X}) \to \mathbb{R}$ 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

$$\partial\Delta(x) = 0, \tag{D.1}$$

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 $\mathcal{X}$.

Date: September 28, 2018.

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 D.1 (Manifold)**.** Let d_k, d_x and r be positive integers such that $d_x \geqslant d_k$. 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 $\alpha_m : \mathcal{K} \to \mathcal{V}$ 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) $\alpha_m^{-1} : \mathcal{V} \to \mathcal{K}$ is continuous, and (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 D.2 (Connected Branch)**.** For any subset $\mathcal{M}$ of a topological space, if any two points m_1 and m_2 cannot 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 D.3 (Volume)**.** For a $d_x \times d_k$ matrix $A = (x_1, x_2, ..., x_{d_k})$ with $x_i \in \mathbb{R}^{d_x}$, $1 \leqslant i \leqslant d_k \leqslant d_x$, let $\mathrm{Vol}(A) = \sqrt{\det(A^\intercal A)}$, which is the *volume* of the parallelepiped $P(A)$ with edges given by the columns of A, $P(A) = \{c_1 x_1 + \cdots + c_{d_k} x_{d_k} : 0 \leqslant c_i \leqslant 1, i = 1, \ldots, d_k\}$.

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

Definition D.4 (Integration on a parametrized manifold)**.** Let $\mathcal{K}$ be open in $\mathbb{R}^{d_k}$, and let $\alpha : \mathcal{K} \to \mathbb{R}^{d_x}$ be of class $\mathcal{C}^r$ on $\mathcal{K}$, $r \geqslant 1$. 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

$$\int_{\mathcal{M}} g(m) d\mathrm{Vol} := \int_{\mathcal{K}} (g \circ \alpha)(k) \mathrm{Vol}(D\alpha(k)) dk, \tag{D.2}$$

provided that the right side integral exists. Here $D\alpha(k)$ is the Jacobian matrix of the mapping $k \mapsto \alpha(k)$, and $\mathrm{Vol}(D\alpha(k))$ is the volume of matrix $D\alpha(k)$ as defined in Definition D.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 D.1 (Partition of Unity on $\mathcal{M}$ of class $\mathcal{C}^\infty$)**.** *Let $\mathcal{M}$ be a d_k-manifold without boundary in $\mathbb{R}^{d_x}$ of class $\mathcal{C}^r$, $r \geqslant 1$, 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 $i \in \mathcal{I}$, $0 \leqslant p_i(m) \leqslant 1$, (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 $m \in \mathcal{M}$, $\sum_{p_i \in \mathcal{P}} p_i(m) = 1$, and (4) for each $p_i \in \mathcal{P}$ there is an open set $\mathcal{U} \in \vartheta$, such that $supp(p_i) \subseteq \mathcal{U}$.*

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

Definition D.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 $\mathcal{C}^r$, $r \geqslant 1$. Suppose there is an coordinate patch $\alpha_j : \mathcal{V}_j \subseteq \mathbb{R}^{d_k} \to \vartheta_j$, that is 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:

$$\int_{\mathcal{M}} g(m) d\mathrm{Vol} := \sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}} \int_{\mathcal{K}_j} [(p_i g) \circ \alpha_j](k) \mathrm{Vol}(D\alpha_j(k)) dk, \tag{D.3}$$

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

Appendix E. Proofs of Appendix A

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

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

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

$$F_{\Delta,\mu}(\delta) = \int 1\{\Delta(x) \leqslant \delta\} d\mu(x). \tag{E.4}$$

The SPE-function is the map

$$\Delta^*_\mu : \mathcal{U} \subseteq [0, 1] \to \mathbb{R},$$

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

$$\Delta^*_\mu(u) := F^{\leftarrow}_{\Delta,\mu}(u) := \inf_{\delta \in \mathbb{R}} \{F_{\Delta,\mu}(\delta) \geqslant u\}. \tag{E.5}$$

From this functional perspective, the map $u \mapsto \Delta^*_\mu(u)$ is the result of applying a sorting operator to the map $x \mapsto \Delta(x)$ that sorts the values of Δ in increasing order weighted by μ. The next subsections provide the proofs of 3 results:

1) Lemma A.1, which characterizes some analytical properties of the distribution function $\delta \mapsto F_{\Delta,\mu}(\delta)$ and the sorted function $u \mapsto \Delta^*_\mu(u)$,

2) Lemma A.2, which derives the functional derivatives of $F_{\Delta,\mu}$ and Δ^*_μ with respect to Δ and μ, and
3) Lemma A.3, which derives the functional derivatives of the related classification operator $\Lambda^-_{\Delta,\mu,\delta}$ with respect to Δ, μ and δ.

E.1. **Proof of Lemma A.1.** We use the following results in the proof of Lemma A.1.

Lemma E.1. *If $\Delta : B(\mathcal{X}) \to \mathbb{R}$ 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 $\{\delta_i\}_{i\geqslant 1}^\infty$ in $\Delta(\overline{\mathcal{X}})$, there is a corresponding sequence $\{x_i\}_{i\geqslant 1}$ in $\overline{\mathcal{X}}$ such that $\Delta(x_i) = \delta_i$. Suppose $\{\delta_i\}_{i\geqslant 1}^\infty$ converges to $\delta_0 \in \Delta(\overline{\mathcal{X}})$. By compactness of $\overline{\mathcal{X}}$, we can find a converging subsequence of $\{x_i\}_{i\geqslant 1}$ 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. □

Lemma E.2. *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 E.2. 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 $\{\widetilde{\theta}_i : i = 1, 2, ..., m\}$ which covers $\mathcal{V}$.

Let $\Theta = \cup_{i=1}^m \widetilde{\theta}_i$. 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 $\{v_i : i \geqslant 1\}$. By compactness of $\mathcal{V}$, $v_0 \in \mathcal{V}$. Since $d(x_i, v_0) \to 0$ as $i \to \infty$ 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. □

Proof of Lemma A.1. 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 Theorem 5-1 in Spivak (1965, p. 111). 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 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 E.1, 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 Theorem 5-1 in Spivak (1965, p. 111).

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 $\theta(x) := X_1(x) \times ... \times X_{d_x}(x)$ centered at $x = (x_1, ..., x_{d_x})$ where $X_k(x) := (x_k - a_k(x), x_k + a_k(x))$, with $a_k(x) > 0$, $k = 1, 2, ..., d_x$. Let $A(x) := \sup_{1 \leqslant k \leqslant d_x} a_k(x)$ be such that:

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

which can be fulfilled by using small enough η_3.

By continuity of $x \mapsto \partial \Delta(x)$, for small enough $A(x)$ and any $x' \in \theta(x)$, there always exists an index $\mathrm{i}(x) \in \{1, 2, ..., d_x\}$ such that $|\partial_{x_{\mathrm{i}(x)}} \Delta(x')| \geqslant \frac{c}{2\sqrt{d_x}}$ since $\|\partial \Delta(x')\| \geqslant c$ 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 $\theta_i := \theta(x^i)$, $i \in \{1, 2, ..., m\}$, where $\theta_i = X_{i1} \times ... \times X_{id_x}$ and $X_{ik} := X_k(x^i)$, $k \in \{1, \ldots, d_x\}$.

For a given $i \in \{1, 2, ..., m\}$, consider the center of θ_i, denoted as x^i. Without loss of generality, we can assume that $\mathrm{i}(x^i) = d_x$. Then, for all $x' \in \theta(x^i)$, $|\partial_{x_{d_x}} \Delta(x')| \geqslant c/2\sqrt{d_x}$. 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 $g(x'_1, x'_2, ..., x'_{d_x-1}, \delta') = x'_{d_x}$, for any $x' = (x'_1, x'_2, ..., x'_{d_x}) \in \widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta)) \cap \theta(x^i)$ and $\delta' = \Delta(x')$. Also by the implicit function theorem,

$$\partial g(x'_1, ..., x'_{d_x-1}, \delta') = \frac{-(\partial_{x_1}\Delta(x'), \partial_{x_2}\Delta(x'), ..., \partial_{x_{d_x-1}}\Delta(x'), -1)}{\partial_{d_x}\Delta(x')}.$$

So $\|\partial g(x'_1, ..., x'_{d_x-1}, \delta')\| \leqslant \frac{\|\partial\Delta(x')\|}{|\partial_{x_{d_x}}\Delta(x')|} \leqslant \frac{2(C+1)\sqrt{d_x}}{c} := \Lambda$ because $|\partial_{x_{d_x}}\Delta(x')| \geqslant c/2\sqrt{d_x}$ and $\|\partial\Delta(x')\| \leqslant C$. Therefore,

$$\begin{aligned}
&|g(x'_1, x'_2, ..., x'_{d_x-1}, \delta') - x^i_{d_x}| = |g(x'_1, x'_2, ..., x'_{d_x-1}, \delta') - g(x^i_1, x^i_2, ..., x^i_{d_x-1}, \delta)| \\
&\quad \leqslant \sup_{x' \in \theta(x), \delta' = \Delta(x')} \|\partial g(x_1, x_2, ..., x_{d_x-1}, \delta')\| \cdot \|(x_1 - x^i_1, x_2 - x^i_2 ..., x_{d_x-1} - x^i_{d_x-1}, \delta' - \delta)\| \\
&\quad \leqslant \Lambda(\sqrt{a_1^2(x^i) + ... + a^2_{d_x-1}(x^i)} + \eta_3),
\end{aligned}$$

since $\|(x_1 - x^i_1, ..., x_{d_x-1} - x^i_{d_x-1}, \delta' - \delta)\| \leqslant \|(x_1 - x^i_1, ..., x_{d_x-1} - x^i_{d_x-1})\| + |\delta' - \delta|$, with $\|(x_1 - x^i_1, ..., x_{d_x-1} - x^i_{d_x-1})\| \leqslant \sqrt{a_1^2(x^i) + ... + a^2_{d_x-1}(x^i)}$ and $|\delta' - \delta| < \eta_3$.

We can choose $a_1(x^i) = a_2(x^i) = ... = a_{d_x-1}(x^i) = \eta_4$ and $a_{d_x}(x^i) = 2(1+\eta_3)\Lambda(\sqrt{d_x - 1}\eta_4 + \eta_3)$, using η_4 small enough in order to fulfill the following property of θ_i: with η_4 small enough,

$$\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta)) \cap \theta_i \subseteq X_{i1} \times ... \times X_{i,d_x-1} \times \left(x^i_{d_x} - \frac{a_{d_x}(x^i)}{2(1+\eta_3)}, x^i_{d_x} + \frac{a_{d_x}(x^i)}{2(1+\eta_3)} \right),$$

or geometrically, the tube $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta))$ does not intersect θ_i's faces except at the ones which are *parallel* to the vector $(0, ..., 0, 1) \in \mathbb{R}^{d_x}$. 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 $i \in \{1, 2, ..., m\}$, $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta))$ intersects θ_i at axis $\mathrm{i}(x^i)$, where x^i is the center of θ_i. This property implies that g is a well-defined injection from $X_{i1} \times ... \times X_{i,d_x-1} \times B_{\eta_3}(\delta)$ to $X_{i1} \times ... \times X_{i,d_x}$, for $i \in \{1, \ldots, m\}$, which will allow us to perform a change of variable in the equation (E.7). 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 D.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 $\mathcal{C}^\infty$ partition of unity p_j, $1 \leqslant j \leqslant J$ on Θ with the properties given in the lemma.

Our main goal is to compute

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

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,

$$\begin{aligned}
F_{\Delta,\mu}(\delta+\eta) - F_{\Delta,\mu}(\delta) &= \int_{x\in\mathcal{X}} 1(\delta \leqslant \Delta(x) \leqslant \delta+\eta)\mu'(x)dx \\
&= \int_{\mathcal{M}_\Delta(B^+_\eta(\delta))} \mu'(x)dx = \int_{\widetilde{\mathcal{M}}_\Delta(B^+_\eta(\delta))} \mu'(x)dx = \int_{\widetilde{\mathcal{M}}_\Delta(B^+_\eta(\delta))\cap\Theta} \mu'(x)dx \\
&= \int_{\widetilde{\mathcal{M}}_\Delta(B^+_\eta(\delta))\cap(\cup_{i=1}^m \theta_i)} \mu'(x)\sum_{j=1}^J p_j(x)dx = \sum_{1\leqslant i\leqslant m, 1\leqslant j\leqslant J} \int_{\widetilde{\mathcal{M}}_\Delta(B^+_\eta(\delta))\cap\theta_i} p_j(x)\mu'(x)dx. \quad \text{(E.6)}
\end{aligned}$$

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

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

In equation (E.6), we apply a change of variable defined by the map ψ_{d_x} to the (i, j)-th element of the sum:

$$\begin{aligned}
\int_{\theta_i\cap\widetilde{\mathcal{M}}_\Delta(B^+_\eta(\delta))} p_j(x)\mu'(x)dx &= \int_{X_{i1}\times...\times X_{i(d_x-1)}\times B^+_\delta(\eta)} (p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x})|\det(D\psi_{d_x})|d\delta' dx_{-d_x} \\
&= \int_{X_{i1}\times...\times X_{i(d_x-1)}} \int_{B^+_\delta(\eta)} \frac{(p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x})}{|\partial_{x_{d_x}}\Delta \circ \psi_{d_x}|} d\delta' dx_{-d_x} \\
&= \eta \int_{X_{i1}\times...\times X_{i(d_x-1)}} \frac{(p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x})}{|\partial_{x_{d_x}}\Delta \circ \psi_{d_x}|} dx_{-d_x} + o(\eta). \quad \text{(E.7)}
\end{aligned}$$

The second equality follows because

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

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

The last equality follows as $\eta \to 0$, 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 $(x_{-d_x}, \delta') \in X_{i1} \times ... \times X_{i(d_x-1)} \times B^+_\eta(\delta)$. In (E.7), 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 (E.7) as an integral on a manifold, which is

$$\eta \int_{X_{i1}\times ... \times X_{i(d_x-1)}} \frac{(p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x})}{|\partial_{x_{d_x}} \Delta \circ \psi_{d_x}|} dx_{-d_x} = \eta \int_{\widetilde{\mathcal{M}}_\Delta(\delta) \cap \theta_i} \frac{p_j(x)\mu'(x)}{\| \partial \Delta(x) \|} d\text{Vol}. \tag{E.8}$$

Summing up over i and j in (E.7) and using Definition 5.5,

$$\sum_{1\leqslant i \leqslant m, 1 \leqslant j \leqslant J} \int_{\widetilde{\mathcal{M}}_\Delta(B^+_\eta(\delta))\cap \theta_i} p_j(x)\mu'(x)dx = \eta \int_{\widetilde{\mathcal{M}}_\Delta(\delta)\cap\Theta} \frac{\mu'(x)}{\| \partial\Delta(x) \|} d\text{Vol} + o(\eta). \tag{E.9}$$

Let us explain(E.8). Equation (E.8) is calculated using the following fact: The mapping $\alpha : X_{i1} \times ... \times X_{id_x-1} \to X_{i1} \times ... \times X_{id_x}$ such that $\alpha(x_1, ..., x_{d_x-1}) = (x_1, ..., x_{d_x-1}, g(x_1, ..., x_{d_x-1}, \delta))$ has Jacobian matrix

$$D\alpha^\mathsf{T}(x_{-d_x}) = \begin{bmatrix} 1 & 0 & ... & 0 & \partial_{x_1} g(x_{-d_x}) \\ 0 & 1 & ... & 0 & \partial_{x_2} g(x_{-d_x}) \\ ... & ... & ... & ... & ... \\ 0 & ... & ... & 1 & \partial_{x_{d_x-1}} g(x_{-d_x}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & ... & 0 & (\partial_{x_1}\Delta/\partial_{x_{d_x}}\Delta)(\widetilde{x}) \\ 0 & 1 & ... & 0 & (\partial_{x_2}\Delta/\partial_{x_{d_x}}\Delta)(\widetilde{x}) \\ ... & ... & ... & ... & ... \\ 0 & ... & ... & 1 & (\partial_{x_{d_x-1}}\Delta/\partial_{x_{d_x}}\Delta)(\widetilde{x}) \end{bmatrix},$$

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

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

Hence, the left hand side of equation (E.8) is:

$$\eta \int_{X_{i1}\times ... \times X_{i(d_x-1)}} \frac{(p_j \circ \psi_{d_x}) \cdot (\mu' \circ \psi_{d_x})}{\|\partial\Delta \circ \psi_{d_x}\|} \text{Vol}(D\alpha) dx_{-d_x},$$

and it can be further re-expressed as the right side of (E.8) using Definition 5.4.

By equations (E.6) and (E.9),

$$\frac{F_{\Delta,\mu}(\delta + \eta) - F_{\Delta,\mu}(\delta)}{\eta} = \int_{\mathcal{M}_\Delta(\delta)} \frac{\mu'(x)}{\| \partial\Delta(x) \|} d\text{Vol} + o(1), \tag{E.10}$$

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 (E.10) converge to 0 uniformly. Thus, adding up the results for all connected branches in equation (E.10), the statements of Lemma A.1 hold. □

E.2. **Proof of Lemma A.2.** We use the following results in the proof of Lemma A.2.

Lemma E.3 (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 any $\delta \in \mathcal{D}$ and some small η_1 such that $0 < \eta_1 < \eta$. Then, $\delta \mapsto \int_{\mathcal{M}_\Delta(\delta)} f d\mathrm{Vol}$ is continuous on $\mathcal{D}$.*

Proof. First, we follow Step 1 in the Proof of Lemma A.1. Suppose we have a set of open rectangles $\Theta = \{\theta_1, ..., \theta_m\}$ such that $\overline{\widetilde{\mathcal{M}}_\Delta(B_{\eta_2}(\delta)) \cap B_{\eta_2}(\overline{\mathcal{X}})} \subset \cup_{i=1}^m \theta_i \subset \overline{\cup_{i=1}^m \theta_i} \subset \widetilde{\mathcal{M}}_\Delta(B_{\eta_1}(\delta)) \cap B_{\eta_1}(\overline{\mathcal{X}})$ 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, $i = 1, 2, ..., m$, satisfies that $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3})$ intersects θ_i at axis $\mathrm{i}(\theta_i)$, for any $\eta_2 < \eta_3$.

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

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

$$\int_{\mathcal{M}_\Delta(\delta')} f d\mathrm{Vol} = \sum_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant J} \int_{\widetilde{\mathcal{M}}_\Delta(\delta') \cap \theta_i} p_j(x) f(x) d\mathrm{Vol}. \tag{E.11}$$

The equation (E.11) holds since $f(x) = 0$ for all $x \notin \mathcal{X}$.

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

Without loss of generality, assume that $\widetilde{\mathcal{M}}_\Delta(B_{\eta_3}(\delta))$ intersects θ_i at axis $\mathrm{i}(\theta_i) = d_x$. 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 $x \in \theta_i$, $i = 1, 2, ..., m$.

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

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

where $\psi_{d_x}(x_{-d_x}, \delta') = (x_{-d_x}, g(x_{-d_x}, \delta'))$ for $x_{-d_x} := (x_1, x_2, ..., x_{d_x - 1})$. 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:

$$\int_{\widetilde{\mathcal{M}}_\Delta(\delta')\cap\theta_i} p_j(x) f(x) d\mathrm{Vol} = \int_{X_1\times X_2\times\ldots\times X_{d_x-1}} (p_j f)\circ\psi_{d_x}(x_{-d_x},\delta') \frac{\|\partial\Delta\circ\psi_{d_x}(x_{-d_x},\delta')\|}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta')|} dx_{-d_x}. \tag{E.12}$$

Since $|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x}, \delta')| = |\partial_{x_{d_x}}\Delta|_{x=\psi_{d_x}(x_1,\ldots,x_{d_x-1},\delta')} > c$ for all $\delta' \in B_{\eta_3}(\delta)$ and $x_{-d_x} \in X_1 \times X_2 \times \ldots \times X_{d_x-1}$ 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

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

Since $X_1 \times \ldots \times X_{d_x-1}$ and is bounded, it immediately follows that $\delta' \mapsto \int_{\widetilde{\mathcal{M}}_\Delta(\delta')\cap\theta_i} p_j(x) f(x) d\mathrm{Vol}$ is continuous at $\delta' = \delta$, and hence

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

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 $\mathcal{D}$. □

Lemma E.4 (Hadamard differentiability of $\Delta \mapsto F_{\Delta,\mu}$ and $\Delta \mapsto \Delta^*_\mu$)**.** *Suppose that* S.1-S.2 *hold. Then:*

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

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

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

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

Proof of Lemma E.4. To shows statement (a), for any $G_n \to G \in \mathbb{F}_0$ under sup-norm such that $\Delta + t_n G_n \in \mathbb{F}$, and $t_n \to 0$, we consider

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

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 $n \geqslant N$, since $G_n \to G$ in sup-norm.

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

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

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

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

As in Step 2 of the proof of Lemma A.1, suppose $\theta_i = X_{i1} \times ... \times X_{id_x}$ intersects $\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))$ at $\mathrm{i}(\theta_i) = x_{d_x}$. Define the parametrization

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

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

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,

$$\int_{\widetilde{\mathcal{M}}_\Delta(B_\eta^+(\delta)) \cap \theta_i} p_j(x) \frac{\left[1\{\Delta(x) + t_n G_n(x) \leqslant \delta\} - 1\{\Delta(x) \leqslant \delta\}\right] \mu'(x)}{t_n} dx$$
$$\leqslant \quad \frac{\int_{\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta)) \cap \theta_i} \left[1\{\Delta(x) + t_n(G(x) - \zeta) \leqslant \delta\} - 1\{\Delta(x) \leqslant \delta\}\right] \mu'(x) dx}{t_n}.$$

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

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

$$= -\int_{\theta_i \cap \widetilde{M}_\Delta(\delta)} p_j(x)\mu'(x)\frac{G(x)-\zeta}{\| \partial\Delta(x) \|} d\mathrm{Vol} + o(\eta)$$
$$= -\int_{\theta_i \cap M_\Delta(\delta)} p_j(x)\mu'(x)\frac{G(x)-\zeta}{\| \partial\Delta(x) \|} d\mathrm{Vol} + o(\eta),$$

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 $t_n \to 0$,

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

and

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

as $\delta' \to \delta$. 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 $n \geqslant N$, and $|G \circ \psi_{d_x}(x_{-d_x}, \delta) - \zeta|$ is bounded by some absolute constant, $\sum_j p_j(x) = 1$ and $p_j(x) \geqslant 0$, we can let $\zeta \to 0$ to conclude that:

$$\lim_{n\to\infty} \frac{F_{\Delta + t_n G_n, \mu}(\delta) - F_{\Delta,\mu}(\delta)}{t_n} \leqslant \sum_{i=1}^{m}\sum_{j=1}^{J} -\int_{\theta_i \cap M_\Delta(\delta)} p_j(x)\mu'(x)\frac{G(x)}{\| \partial\Delta(x) \|} d\mathrm{Vol}.$$

The right side is given by:

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

On the other hand,

$$1(\Delta(x) + t_n G_n(x) \leqslant \delta) \geqslant 1(\Delta(x) + t_n(G(x) + \zeta) \leqslant \delta)$$

for some $\zeta > 0$. So,

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

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

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

$$= -\int_{\theta_i \cap \widetilde{M}_\Delta(\delta)} p_j(x)\mu'(x)\frac{G(x)+\zeta}{\| \partial\Delta(x) \|} d\text{Vol} - o(\eta)$$
$$= -\int_{\theta_i \cap M_\Delta(\delta)} p_j(x)\mu'(x)\frac{G(x)+\zeta}{\| \partial\Delta(x) \|} d\text{Vol} - o(\eta).$$

Let $\zeta \to 0$ and $\eta \to 0$, it follows that

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

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 $\partial_\Delta F_{\Delta,\mu}(\delta_n)[G] \to \partial_\Delta F_{\Delta,\mu}(\delta)[G]$ by uniform continuity of $\delta \mapsto \partial_\Delta F_{\Delta,\mu}(\delta)[G]$ on $\mathcal{D}$, which holds by Lemma E.3 because G, μ', and $\|\partial\Delta\|$ are continuous on $\overline{\mathcal{X}}$ and $\mathcal{D}$ excludes neighborhoods of the critical values of Δ in $\overline{\mathcal{X}}$.

Tho 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 $u \in \mathcal{U} = \{u \in (0,1) : \Delta^*_\mu(u) \in \mathcal{D},\ f_{\Delta,\mu}(\Delta^*_\mu(u)) \geqslant \varepsilon\}$. □

Proof of Lemma A.2 . To show Statement (a), Consider $t_n \to 0$ and $(G_n, H_n) \to (G, H) \in \mathbb{D}_0 := \mathbb{F}_0 \times \mathbb{H}$ as $n \to \infty$, such that $(\Delta + t_nG_n, \mu + t_nH_n) \in \mathbb{D}$. Let $\Delta_n := \Delta + t_nG_n$ and $\mu_n := \mu + t_nH_n$. 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 E.4,

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

Let $g_{\Delta,\delta} := 1(\Delta(x) \leqslant \delta)$. 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 $H_n \to H$ in $\mathbb{H}$. For the first term, we further decompose

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

The first two terms go to 0 by $\|H_n - H\|_{\mathcal{G}} \to 0$. Moreover, $H(g_{\Delta_n,\delta}) \to H(g_{\Delta,\delta})$ because $g_{\Delta_n,\delta}(X) = 1(\Delta_n(X) \leqslant \delta) \to g_{\Delta,\delta}(X) = 1(\Delta(X) \leqslant \delta)$ in the $L^2(\mu)$ norm, since $\Delta_n \to \Delta$ 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 $\delta \in \mathcal{D}$,

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

By an argument similar to the proof of Lemma E.4, 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). $\square$

E.3. **Proof of Lemma A.3.** We will denote the functions in the classes $\mathcal{F}_M$ and $\mathcal{F}_I$ by $\varphi_t(x)$ whenever we want to distinguish $x = (x_1, \dots, x_{d_x})$, the argument of the function, from $t := (t_1, \dots, t_{d_z})$, the index of the function in the class. Otherwise, we will use $\varphi(x)$. To analyze $\Lambda^-_{\Delta,\mu,\delta}$ it is convenient to introduce the operator $\Upsilon_{\Delta,\mu,\delta} : \widetilde{\mathbb{D}} \to \mathbb{R}$ defined by

$$\Upsilon_{\Delta,\mu,\delta}(\varphi) := \int \varphi(x) 1\{\Delta(x) \leqslant \delta\} d\mu(x),$$

since $\Lambda^-_{\Delta,\mu,\delta}(\varphi) = \Upsilon_{\Delta,\mu,\delta}(\varphi)/\Upsilon_{\Delta,\mu,\delta}(1)$.

Let $\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta)) := \cup_{\delta' \in B_\eta(\delta)} \widetilde{\mathcal{M}}_\Delta(\delta')$, where $\widetilde{\mathcal{M}}_\Delta(\delta) := \{x \in B(\mathcal{X}) : \Delta(x) = \delta\}$ and $B_\eta(\delta) := (\delta - \eta, \delta + \eta)$ for any $\delta \in \mathcal{V}$ and $\eta > 0$. When $\varphi_t \in \mathcal{F}_I$ we make the following technical assumption to deal with the discontinuity of the indicator functions:

AS.1. Define the set $\widetilde{\mathcal{Z}}_{k,\eta}(\delta, t_k) := \{x_{-k} : (x_k, x_{-k}) \in \widetilde{\mathcal{M}}_\Delta(B_\eta(\delta)), x_k = t_k\}$ for any $\eta > 0$, $\delta \in \mathcal{V}$, $k = 1, 2, ..., d_x$, and $t_k \in \mathbb{R}$. Then, for any $\epsilon > 0$, there exist $\eta_0 > 0$ such that for any $\eta < \eta_0$, $\int_{\widetilde{\mathcal{Z}}_{k,\eta}(\delta,t_k)} d\mu(x_{-k}) \leqslant \epsilon$ holds uniformly over all $\delta \in \mathcal{V}$, $t_k \in \mathbb{R}$ and $k = 1, 2, ..., d_x$.

The next result shows that $(\Delta, \mu, \delta) \mapsto \Upsilon_{\Delta,\mu,\delta}$ is Hadamard differentiable.

Lemma E.5 (Hadamard differentiability of $(\Delta, \mu, \delta) \mapsto \Upsilon_{\Delta,\mu,\delta}$)**.** *Assume that Assumptions S.1 and S.2 hold and $\delta \in \mathcal{D}$. Then,*

(a) The map $\Upsilon_{\Delta,\mu,\delta}(\varphi) : \widetilde{\mathbb{D}} \to \mathbb{R}$ is Hadamard-differentiable uniformly in $\varphi \in \mathcal{F}_M$ at (Δ, μ, δ) tangentially to $\widetilde{\mathbb{D}}_0$.

(b) If in addition Assumption AS.1 *holds, the map* $\Upsilon_{\Delta,\mu,\delta}(\varphi) : \widetilde{\mathbb{D}} \to \mathbb{R}$ *is Hadamard-differentiable uniformly in* $\varphi \in \mathcal{F}_I$ *at* (Δ, μ, δ) *tangentially to* $\widetilde{\mathbb{D}}_0$*.*

(c) The derivative map $\partial_{\Delta,\mu,\delta}\Upsilon_{\Delta,\mu,\delta}(\varphi) : \widetilde{\mathbb{D}} \to \mathbb{R}$ *is defined by:*

$$(G, H, K) \mapsto \partial_{\Delta,\mu,\delta}\Upsilon_{\Delta,\mu,\delta}(\varphi)[G, H, K] := \int_{\mathcal{M}_\Delta(\delta)} \varphi(x)\frac{K - G(x)}{\|\partial\Delta(x)\|}d\mathrm{Vol} + H(h_{\Delta,\delta,\varphi}),$$

where $h_{\Delta,\delta,\varphi} := \varphi(x)1\{\Delta(x) \leqslant \delta\}$*.*

Proof of Lemma E.5. Statements (a) and (b) follow by similar arguments. For brevity, we focus on the proof of Statement (b) and mention the changes needed for the proof of Statement (a), which is simpler.

To show Statement (b), consider $s_n \to 0$ and $(G_n, H_n, K_n) \to (G, H, K) \in \widetilde{\mathbb{D}}_0$ as $n \to \infty$, such that $(\Delta + s_nG_n, \mu + s_nH_n, \delta + s_nK_n) \in \widetilde{\mathbb{D}}$. Let $\Delta_n := \Delta + s_nG_n$, $\mu_n := \mu + s_nH_n$, and $\delta_n := \delta + s_nK_n$. Then, we can decompose

$$\Upsilon_{\Delta_n,\mu_n,\delta_n}(\varphi) - \Upsilon_{\Delta,\mu,\delta}(\varphi) = [\Upsilon_{\Delta_n,\mu_n,\delta_n}(\varphi) - \Upsilon_{\Delta_n,\mu,\delta_n}(\varphi)] + [\Upsilon_{\Delta_n,\mu,\delta_n}(\varphi) - \Upsilon_{\Delta,\mu,\delta}(\varphi)]. \quad \text{(E.13)}$$

The first term of (E.13) satisfies

$$\frac{\Upsilon_{\Delta_n,\mu_n,\delta_n}(\varphi) - \Upsilon_{\Delta_n,\mu,\delta_n}(\varphi)}{s_n} = H_n(h_{\Delta_n,\delta_n,\varphi}) = H(h_{\Delta,\delta,\varphi}) + o(1).$$

The first equality follows from linearity of $\mu \mapsto \Upsilon_{\Delta_n,\mu,\delta_n}(\varphi)$ and $h_{\Delta_n,\delta_n,\varphi} = \varphi(x)1\{\Delta_n(x) \leqslant \delta_n\}$. To show the second equality note that

$$H_n(h_{\Delta_n,\delta_n,\varphi}) - H(h_{\Delta,\delta,\varphi}) = H_n(h_{\Delta_n,\delta_n,\varphi}) - H_n(h_{\Delta,\delta,\varphi}) + [H_n - H](h_{\Delta,\delta,\varphi}),$$

where the second term goes to zero by the assumption $H_n \to H$ in $\widetilde{\mathbb{H}}$. For the first term, we further decompose

$$\begin{aligned}|H_n(h_{\Delta_n,\delta_n,\varphi}) - H_n(h_{\Delta,\delta,\varphi})| \leqslant{}& |H_n(h_{\Delta_n,\delta_n,\varphi}) - H(h_{\Delta_n,\delta_n,\varphi})| \\ &+ |H_n(h_{\Delta,\delta,\varphi}) - H(h_{\Delta,\delta,\varphi})| + |H(h_{\Delta_n,\delta_n,\varphi}) - H(h_{\Delta,\delta,\varphi})|.\end{aligned}$$

By definition of the space $\widetilde{\mathbb{H}}$, the first two terms go to 0 by $\|H_n - H\|_{\widetilde{\mathcal{G}}} \to 0$.

Moreover, $H(h_{\Delta_n,\delta_n,\varphi}) \to H(h_{\Delta,\delta,\varphi})$ because

$$h_{\Delta_n,\delta_n,\varphi}(X) = \varphi(X)1(\Delta_n(X) \leqslant \delta_n) \to h_{\Delta,\delta,\varphi}(X) = \varphi(X)1(\Delta(X) \leqslant \delta)$$

in the $L^2(\mu)$ norm, since $\Delta_n \to \Delta$ 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.

Next we show that the second term of (E.13) satisfies

$$\frac{\Upsilon_{\Delta_n,\mu,\delta_n}(\varphi) - \Upsilon_{\Delta,\mu,\delta}(\varphi)}{s_n} = \int_{\mathcal{M}_\Delta(\delta)} \varphi(x)\frac{K - G(x)}{\|\partial\Delta(x)\|}\mu'(x)d\mathrm{Vol} + o(1).$$

The proof follows the same steps as the proof of Lemma E.4 after noticing that we can write

$$\Upsilon_{\Delta_n,\mu,\delta_n}(\varphi) = \Upsilon_{\widetilde{\Delta}_n,\mu,\delta}(\varphi),$$

where $\widetilde{\Delta}_n = \Delta + s_n \widetilde{G}_n$ with $\widetilde{G}_n = G_n - K_n$, and replacing $\mu'(x)$ by $\widetilde{\mu}'(x) = \varphi(x)\mu'(x)$.

Specifically, following the notation in the proof of Lemma A.1,

$$\Upsilon_{\widetilde{\Delta}_n,\mu,\delta}(\varphi) = \sum_{i=1}^{m}\sum_{j=1}^{J}\int_{\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))\cap\theta_i} p_j(x)\varphi(x)\frac{1\{\delta \leqslant \Delta(x) \leqslant \delta - s_n\widetilde{G}_n(x)\}}{s_n}\widetilde{\mu}'(x)dx.$$

Without loss of generality, assume that θ_i intersects with $\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))$ at the axis $x_{k_i} = x_{d_x}$. When $\varphi(x) \in \mathcal{F}_I$, each component in the above summation satisfies:

$$\begin{aligned}
&\int_{\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))\cap\theta_i} p_j(x)\varphi(x)\frac{1\{\delta \leqslant \Delta(x) \leqslant \delta - s_n\widetilde{G}_n(x)\}\widetilde{\mu}'(x)}{s_n}dx \\
&= \int_{X_{i1}\times\ldots\times X_{i,d_x-1}}\int_{B_\eta(\delta)} \frac{(p_j\cdot\varphi\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta')}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta')|} \times \frac{1\{\delta \leqslant \delta' \leqslant \delta - s_n\widetilde{G}_n\circ\psi_{d_x}(x_{-d_x},\delta')\}}{s_n}d\delta' dx_{-d_x} \\
&= \int_{\widetilde{\mathcal{X}}^c_{d_x,\eta}(\delta,t_{d_x})}\int_{B_\eta(\delta)} \frac{(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta')}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta')|} \times \frac{1\{\delta \leqslant \delta' \leqslant \delta - s_n\widetilde{G}_n\circ\psi_{d_x}(x_{-d_x},\delta')\}}{s_n}d\delta' dx_{-d_x} \\
&+ \int_{\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x})\}}\int_{B_\eta(\delta)} \frac{(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta')}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta')|} \times \frac{1\{\delta \leqslant \delta' \leqslant \delta - s_n\widetilde{G}_n\circ\psi_{d_x}(x_{-d_x},\delta')\}}{s_n}d\delta' dx_{-d_x},
\end{aligned}$$

where $\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x}) := [X_{i1}\times\ldots\times X_{i,d_x-1}]\cap\widetilde{\mathcal{Z}}_{d_x,\eta}(\delta,t_{d_x})$ and $\widetilde{\mathcal{X}}^c_{d_x,\eta}(\delta,t_{d_x}) := X_{i1}\times\ldots\times X_{i,d_x-1}\setminus \widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x})$. When $\varphi(x)\in\mathcal{F}_M$, then we could simply let $\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x}) = \emptyset$ in the rest of the proof.

Partition $t = (t_x, t_y)$ corresponding to $Z = (X, Y)$. Although $x \mapsto \varphi(x) = 1(x \leqslant t_x)\mu(t_y \mid x)$ is a discontinuous function, $\delta \mapsto \varphi(x)\circ\psi_{d_x}(x_{-d_x},\delta)$ is continuous for those x such that $x_{-d_x} \in \widetilde{\mathcal{X}}^c_{d_x,\eta}(\delta,t_{d_x})$ and $\delta = \Delta(x)$. Accordingly, we partition the integral in two regions because the integrand is not necessarily continuous on $\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x})\times B_\eta(\delta)$. We use Assumption AS.1 to bound the integral in this region. Thus, for any $\epsilon > 0$, for η being small enough, the area of $\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x})$, defined as $\int_{\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x})}\mu'(x_{-d_x})dx_{-d_x}$, is less than or equal to ϵ by AS.1 uniformly over δ and t_{d_x}. Then, for large enough n, $k = 1, 2, ..., d_x$ and some arbitrarily small $\zeta > 0$, by continuity of the integrand,

$$\begin{aligned}
&\int_{\widetilde{\mathcal{X}}^c_{d_x,\eta}(\delta,t_{d_x})}\int_{B_\eta(\delta)} \frac{(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta')}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta')|}\times\frac{1\{\delta\leqslant\delta'\leqslant\delta-s_n\widetilde{G}_n\circ\psi_{d_x}(x_{-d_x},\delta')\}}{s_n}d\delta' dx_{-d_x}\\
&\qquad\leqslant-\int_{\widetilde{\mathcal{X}}^c_{d_x,\eta}(\delta,t_{d_x})}\frac{(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta)}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta)|}(\widetilde{G}\circ\psi_{d_x}(x_{-d_x},\delta)-\zeta)dx_{-d_x}+o(\eta)\\
&\qquad=-\int_{X_{i1}\times\ldots\times X_{i,d_x-1}}\frac{(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta)}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta)|}(\widetilde{G}\circ\psi_{d_x}(x_{-d_x},\delta)-\zeta)dx_{-d_x}\\
&\qquad\qquad+\int_{\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x})}\frac{(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta)}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta)|}(\widetilde{G}\circ\psi_{d_x}(x_{-d_x},\delta)-\zeta)dx_{-d_x}+o(\eta),
\end{aligned}$$

where $\widetilde{G}=G-K$.

The inequality above holds by continuity of the integrand $(x_{-d_x},\delta')\mapsto(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta')/|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta')|$ on $\widetilde{\mathcal{X}}^c_{d_x,\eta}(\delta,t_{d_x})\times B_\eta(\delta)$, and

$$\left|\int_{\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x})}\frac{(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta)}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta)|}(\widetilde{G}\circ\psi_{d_x}(x_{-d_x},\delta)-\zeta)dx_{-d_x}\right|\leqslant\int_{\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x})}C dx_{-d_x}\leqslant C\epsilon,$$

for

$$C:=\sup_{x_{-d_x}\in X_{i1}\times\ldots\times X_{i,d_x-1}}\left|\frac{(p_j\cdot\widetilde{\mu})\circ\psi_{d_x}(x_{-d_x},\delta)}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta)|}(\widetilde{G}\circ\psi_{d_x}(x_{-d_x},\delta)-\zeta)\right|,$$

which is bounded from above, because all components in C are bounded from above and $|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta)|$ is bounded away from zero. Similarly, for s_n large enough,

$$\left|\int_{\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x})}\int_{B_\eta(\delta)}(p_j\cdot\psi_{d_x})\times\frac{(1\{\delta\leqslant\Delta\leqslant\delta-s_n\widetilde{G}_n(x)\}\widetilde{\mu}')\circ\psi_{d_x}}{s_n|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}|}d\delta' dx_{-d_x}\right|\leqslant C\epsilon.$$

Therefore, combining the previous results

$$
\begin{aligned}
&\int_{\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))\cap\theta_i} p_j(x)\varphi(x)\frac{1\{\delta \leqslant \Delta(x) \leqslant \delta - s_n\widetilde{G}_n(x)\}}{s_n}\widetilde{\mu}'(x)dx \\
&\quad\leqslant -\int_{\widetilde{\mathcal{X}}^c_{d_x,\eta}(\delta,t_{d_x})} \frac{(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta)}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta)|}(\widetilde{G}\circ\psi_{d_x}(x_{-d_x},\delta)-\zeta)dx_{-d_x} + o(\eta) + C\epsilon \\
&\qquad = -\int_{X_{i1}\times\ldots\times X_{i,d_x-1}} \frac{(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta)}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta)|}(\widetilde{G}\circ\psi_{d_x}(x_{-d_x},\delta)-\zeta)dx_{-d_x} \\
&\qquad\quad + \int_{\widetilde{\mathcal{X}}_{d_x,\eta}(\delta,t_{d_x})} \frac{(p_j\cdot\widetilde{\mu}')\circ\psi_{d_x}(x_{-d_x},\delta)}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta)|}(\widetilde{G}\circ\psi_{d_x}(x_{-d_x},\delta)-\zeta)dx_{-d_x} + o(\eta) + C\epsilon \\
&\quad\leqslant -\int_{X_{i1}\times\ldots\times X_{i,d_x-1}} \frac{(p_j\cdot\widetilde{\mu})\circ\psi_{d_x}(x_{-d_x},\delta)}{|\partial_{x_{d_x}}\Delta\circ\psi_{d_x}(x_{-d_x},\delta)|}(\widetilde{G}\circ\psi_{d_x}(x_{-d_x},\delta)-\zeta)dx_{-d_x} + o(\eta) + 2C\epsilon \\
&\qquad = -\int_{\theta_i\cap\widetilde{\mathcal{M}}_\Delta(\delta)} p_j(x)\varphi(x)\cdot\widetilde{\mu}'(x)\frac{\widetilde{G}(x)-\zeta}{\|\partial\Delta(x)\|}d\mathrm{Vol} + o(\eta) + 2C\epsilon \\
&\qquad = -\int_{\theta_i\cap\mathcal{M}_\Delta(\delta)} p_j(x)\varphi(x)\cdot\widetilde{\mu}'(x)\frac{\widetilde{G}(x)-\zeta}{\|\partial\Delta(x)\|}d\mathrm{Vol} + o(\eta) + 2C\epsilon,
\end{aligned}
$$

where ζ, η and ϵ can be arbitrarily small for large enough n.

Similarly, we can show that

$$
\begin{aligned}
&\int_{\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))\cap\theta_i} p_j(x)\varphi(x)\frac{1\{\delta \leqslant \Delta(x) \leqslant \delta - s_n\widetilde{G}_n(x)\}}{s_n}\widetilde{\mu}'(x)dx \\
&\qquad \geqslant -\int_{\theta_i\cap\mathcal{M}_\Delta(\delta)} p_j(x)\varphi(x)\cdot\widetilde{\mu}'(x)\frac{\widetilde{G}(x)-\zeta}{\|\partial\Delta(x)\|}d\mathrm{Vol} - o(\eta) - 2C\epsilon
\end{aligned}
$$

Since we can choose η and ϵ to be arbitrarily small, we conclude that for any $\varphi \in \mathcal{F}_I$,

$$
\frac{\Upsilon_{\Delta_n,\mu_n,\delta_n}(\varphi) - \Upsilon_{\Delta,\mu,\delta}(\varphi)}{s_n} \to \int_{\mathcal{M}_\Delta(\delta)} \varphi(x)\frac{K - G(x)}{\|\partial\Delta(x)\|}\mu'(x)d\mathrm{Vol} + H(\varphi(x)1\{\Delta(x)\leqslant\delta\}).
$$

To show that the result holds uniformly in $\varphi \in \mathcal{F}_I$, we use the equivalence between uniform convergence and continuous convergence (e.g., Resnick (1987, p.2)). Take a sequence $\varphi^n \in \mathcal{F}_I$ that converges to $\varphi \in \mathcal{F}_I$ in the $L^1(\mu)$ norm, i.e., $\int_{\mathcal{X}} |\varphi^n - \varphi| d\mu \to 0$ as $n \to \infty$. Then, the preceding argument applies to this sequence and $\partial_{\Delta,\mu,\delta}\Upsilon_{\Delta,\mu,\delta}(\varphi^n)[K,G,H] \to \partial_{\Delta,\mu,\delta}\Upsilon_{\Delta,\mu,\delta}(\varphi)[K,G,H]$ by linearity of the map $\varphi \mapsto \partial_{\Delta,\mu,\delta}\Upsilon_{\Delta,\mu,\delta}(\varphi)[K,G]$.

□

Proof of Lemma A.3. Note that $\Lambda^-_{\Delta,\mu,\delta}(\varphi) = \Upsilon_{\Delta,\mu,\delta}(\varphi)/\Upsilon_{\Delta,\mu,\delta}(1)$, where $\Upsilon_{\Delta,\mu,\delta}(1) = \int 1(\Delta(x) \leqslant \delta)d\mu(x) = F_{\Delta,\mu}(\delta)$.

By Lemma E.5, $\Upsilon_{\Delta,\mu,\delta}(\varphi)$ and $\Upsilon_{\Delta,\mu,\delta}(1)$ are Hadamard-differentiable at (Δ,μ,δ) tangentially to $\widetilde{\mathbb{D}}_0$. Then, by the chain rule for Hadamard-differentiable mappings, $\Lambda^{-}_{\Delta,\mu,\delta}(\varphi)$ is Hadamard-differentiable at (Δ,μ,δ) tangentially to $\widetilde{\mathbb{D}}_0$ since $\Upsilon_{\Delta,\mu,\delta}(1) > 0$. The derivative map is obtained from

$$\partial_{\Delta,\mu,\delta}\Lambda^{-}_{\Delta,\mu,\delta}(\varphi) = \frac{\partial_{\Delta,\mu,\delta}\Upsilon_{\Delta,\mu,\delta}(\varphi)}{F_{\Delta,\mu}(\delta)} - \Lambda_{\Delta,\mu,\delta}(\varphi)\frac{\partial_{\Delta,\mu,\delta}\Upsilon_{\Delta,\mu,\delta}(1)}{F_{\Delta,\mu}(\delta)},$$

after replacing the expressions of $\partial_{\Delta,\mu,\delta}\Upsilon_{\Delta,\mu,\delta}(\varphi)$ and $\partial_{\Delta,\mu,\delta}\Upsilon_{\Delta,\mu,\delta}(1)$ from Lemma E.5 and grouping terms.

□

Appendix F. Sufficient Conditions for μ-Donsker Properties in Section 4

Lemma F.1 (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:*

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

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.

Proof of Lemma F.1. Since $\mathcal{V}$ is a union of finite number of closed intervals, for any $\zeta > 0$, we can construct a collection of closed intervals $\mathcal{I} := \{[a_i,b_i] : i = 1,2,...,r\}$ 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) $a_i \leqslant b_i \leqslant a_{i+1} \leqslant b_{i+1}$, for all $i = 1,2,...,r-1$, and (5) $r \leqslant \frac{C_0}{\zeta}$, 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 $\|\partial\Delta(x)\| \leqslant C$ for all $x \in \overline{\mathcal{X}}$ and $\|\partial\Delta(x)\| \geqslant c$ in $\widetilde{\mathcal{M}}_\Delta(B_\eta(\delta))$ for some small $\eta > 0$ and all $\delta \in \mathcal{D}$.

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

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

Moreover, using arguments similar to those used to show Lemma A.1, 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 $\widetilde{\Delta} \in \mathcal{F}$,

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

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

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

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

since $|b_i - a_i| < \zeta$ by construction. Therefore, for ζ small enough, $\{[1(\Delta_j + \zeta \leqslant a_i), 1(\Delta_j - \zeta \leqslant b_i)] : j = 1, 2, ..., q_\zeta, i = 1, 2, ..., r\}$, 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 $rq_\zeta \leqslant \frac{C_0}{\zeta} N(\zeta, \mathcal{F}, \|\cdot\|_\infty)$. Or equivalently, for ζ small enough,

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

Then by assumption,

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

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

Lemma F.2 (Sufficient conditions for $\widetilde{\mathcal{G}}$ 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:*

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

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 $\widetilde{\mathcal{G}}$ is μ-Donsker.

Proof of Lemma F.2. First, $\mathcal{F}_I$ and $\mathcal{F}_M$ are both μ-Donsker. By Lemma F.1, the class $\mathcal{F}$ is μ-Donsker. Since the class of the product of two functions from Donsker classes is Donsker, $\widetilde{\mathcal{G}}$ is μ-Donsker. □

Appendix G. 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 consider 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, $\mu_{c|d}$ denote the distribution of C conditional on $D = d$, μ_d denote the distribution of D, and $\pi_d(d) = \mathrm{P}(D = d)$. As above, $d_x = \dim(X)$, and $\mathcal{D}$ is a compact set consisting of regular values of Δ on $\overline{\mathcal{X}} := \cup_{d\in\mathcal{X}_d}\{d\} \times \overline{\mathcal{X}}_{c|d}$, where $\overline{\mathcal{X}}_{c|d}$ 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 $\overline{\mathcal{X}}_{c|d}$ 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 $\overline{\mathcal{X}}_{c|d}$ 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 $\overline{\mathcal{X}}_{c|d}$, $\mathcal{M}_{\Delta|d}(\delta) := \{c \in \overline{\mathcal{X}}_{c|d} : \Delta(d, c) = \delta\}$ 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′. $\widehat{\Delta}$, the estimator of Δ, 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 $a_n \to \infty$ as $n \to \infty$, 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 $H : g \mapsto H(g)$ uniformly continuous on $\mathcal{G} = \{1(f \leqslant \delta) : f \in \mathcal{F}, \delta \in \mathcal{V}\}$ with respect to the $L^2(\mu)$ norm, which are represented as:

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

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 $\widehat{\mu}(x) = \widehat{\mu}_d(d)\widehat{\mu}_{c|d}(c)$.

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

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

We generalize Lemmas A.1 and A.2 to the case where X includes discrete components.

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

Lemma G.1 (Properties of $F_{\Delta,\mu}$ and Δ^*_μ with discrete X)**.** *Suppose that* $\mathsf{S.1}'$ *and* $\mathsf{S.2}'$ *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:*

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

The map $\delta \mapsto f_{\Delta,\mu}(\delta)$ *is uniformly continuous on* $\mathcal{D}$.

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

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

(2) The map $\Delta^*_\mu(u) : \mathbb{D} \to \mathbb{R}$ *is Hadamard differentiable uniformly in* $u \in \mathcal{U}$ *at* (Δ, μ) *tangentially to* $\mathbb{D}_0$*, with derivative map* $\partial_{\Delta,\mu} \Delta^*_\mu(u) : \mathbb{D}_0 \to \mathbb{R}$ *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 $\mathcal{U} = \{\widetilde{u} \in [0,1] : \Delta^*_\mu(\widetilde{u}) \in \mathcal{D}, f_{\Delta,\mu}(\Delta^*_\mu(\widetilde{u})) > \varepsilon\}$ *for fixed* $\varepsilon > 0$.

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

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

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

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

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 $H_n \in \mathbb{H}$ such that $G_n \to G \in \mathbb{F}_0$ and $H_n \to H \in \mathbb{H}$. Let $\Delta_n = \Delta + t_n G_n$ and $\mu_n = \mu + t_n H_n$, where $t_n \to 0$ as $n \to \infty$.

As in the proof of Lemma A.2, 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 A.2 to each d and averaging over $d \in \mathcal{X}_d$, for any $\delta \in \mathcal{D}$

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

where we use that $\mathcal{X}_d$ is a finite set. By assumption $\mathsf{S}.4'$ and a similar argument to the proof of Lemma A.2,

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

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

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

By an argument similar to the proof of Lemma A.2, it can be shown that the convergence is uniform in $\delta \in \mathcal{D}$. This shows statement (1).

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 A.2. □

We are now ready to derive a functional central limit theorem for the empirical SPE-function. As in Theorem 4.1, let $r_n := a_n \wedge b_n$, the slowest of the rates of convergence of $\widehat{\Delta}$ and $\widehat{\mu}$, where $r_n/a_n \to s_\Delta \in [0,1]$ and $r_n/b_n \to s_\mu \in [0,1]$.

Theorem G.1 (FCLT for $\widehat{\Delta_\mu^*}(u)$ with discrete X)**.** *Suppose that* $\mathsf{S}.1'$-$\mathsf{S}.4'$ *hold, the convergence in* $\mathsf{S}.3'$ *and* $\mathsf{S}.4'$ *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* $\ell^\infty(\mathcal{U})$,

$$r_n(\widehat{\Delta}^*_{\widehat{\mu}}(u) - \Delta^*_\mu(u)) \rightsquigarrow \partial_{\Delta,\mu}\Delta^*_\mu(u)[s_\Delta G_\infty, s_\mu H_\infty], \tag{G.14}$$

as a stochastic process indexed by $u \in \mathcal{U}$, *where* $\mathcal{U}$ *is defined in Lemma G.1.*

Remark G.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 (G.14) under the same conditions as in Theorem 4.3, replacing $\mathsf{S}.1$-$\mathsf{S}.4$ by $\mathsf{S}.1'$-$\mathsf{S}.4'$. Accordingly, we do not repeat the statement here. □

Remark G.2 (CA with discrete covariates)**.** The results of the classification analysis can also be extended to the case where X contains discrete components following analogous arguments as for the SPE. We omit the details for the sake of brevity. $\square$

Proof of Theorem G.1. The result follows from Lemma G.1 and Lemma B.1. $\square$

Appendix H. 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 4.1, and approximated by bootstrap using Theorem 4.3. We first consider two simulation designs where the limit process in Theorem 4.1 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 $3,000$ simulations. The last design is calibrated to mimic the gender wage gap application.

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

$$\Delta^*_\mu(u) = 2(\sqrt{2u} - 1)1(u \leqslant 1/2) + 2(1 - \sqrt{2(1-u)})1(u > 1/2),$$

where 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 held fixed in the grid $\{-1, -0.9, \dots, 1\} \times \{-1, -0.9, \dots, 1\}$. Figure 1 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 4.1, 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 $\widehat{\Delta}^*_\mu(u) \overset{a}{\sim} N(\Delta^*_\mu(u), \exp[2\Delta^*_\mu(u)]/n)$, where $\overset{a}{\sim}$ denotes asymptotic approximation to the distribution.

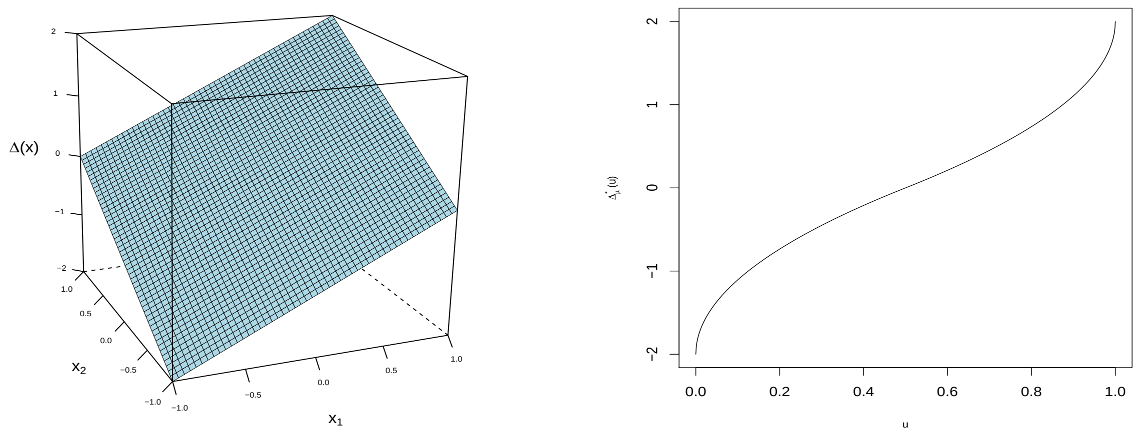

FIGURE 1. PE-function and SPE-function in Design 1. Left: PE function $x \mapsto \Delta(x)$. Right: SPE function $u \mapsto \Delta^*_\mu(u)$.

Table 1 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 indices $u \in \{0.1, 0.2, \dots, 0.9\}$. 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 indices. These asymptotic confidence intervals are not feasible in general, either 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 approximating this distribution by bootstrap. In this case the empirical bootstrap version of the empirical SPE is constructed from the bootstrap PE

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

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 quantile indices.

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

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

Table 1. Properties of Empirical SPE in Design 1

	Bias	Std. Dev.		Pointwise Coverage (%)	
u	($\times$ 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%.

with covariate X uniformly distributed on $\mathcal{X} = (-3, 3)$. Figure 2 plots $x \mapsto \Delta(x)$ on $\mathcal{X}$, and $u \mapsto \Delta^*_\mu(u)$ on $(0, 1)$.[1] 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.

To obtain an analytical expression of the limit $Z_\infty(u)$ of Theorem 4.1, 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 $u \notin \{1/6, 5/6\}$,

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

where

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

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

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

[2]The 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 $\delta > 2$.

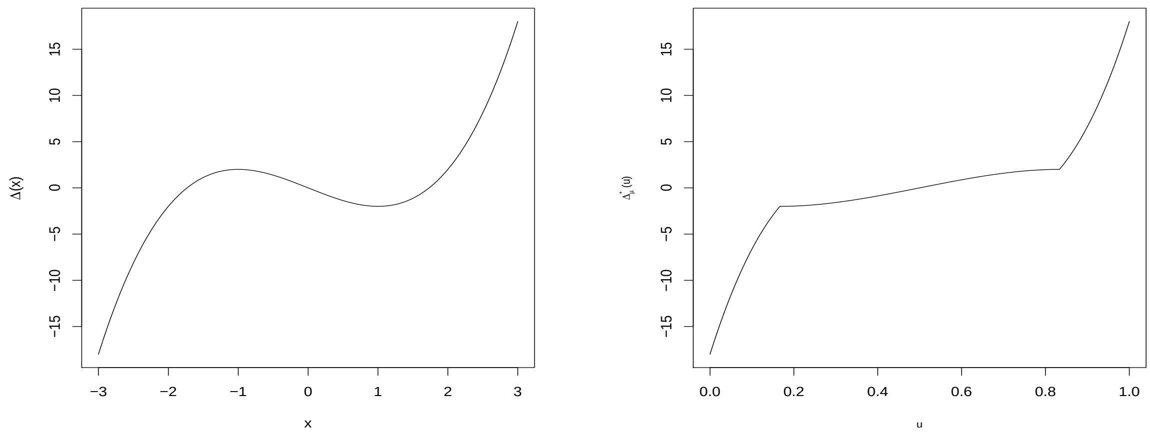

FIGURE 2. PE-function and SPE-function in Design 2. Left: PE function $x \mapsto \Delta(x)$. Right: SPE function $u \mapsto \Delta^*_\mu(u)$.

Table 2 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 indices $u \in \{1/12, 2/12, \ldots, 11/12\}$, where the values of X are held fixed in the grid $\{-3, -2.99, \ldots, 3\}$. The biases are small relative to dispersion except at the kinks $u = 1/6$ 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 $\breve{\Delta}_k(\delta)^2 - 1 = 0$. 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 provides coverages close to the nominal levels even at the kinks.

Design 3 (Calibration to CPS data)**.** This design is calibrated to the interactive linear model with additive error for the conditional expectation in the gender wage gap application of Section 3. More specifically, we generate log wages as

$$Y_i = P(T_i, W_i)'\beta + \sigma\varepsilon_i, \quad i = 1, \ldots, n,$$

where the covariates $X_i = (T_i, W_i)$ are fixed to the values in the 2015 CPS data set, $P(T, W) = (TW, (1-T)W)$, β and σ^2 are the least squares estimates of the regression coefficients and residual variance in the data set, $(\varepsilon_1, \ldots, \varepsilon_n)$ is a sequence of i.i.d. standard normal random variables independent of X_i, and $n = 32,523$, the sample size in the application. For each

TABLE 2. Properties of Empirical SPE in Design 2

	Bias	Std. Dev.		Pointwise Coverage (%)	
u	($\times$ 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%.

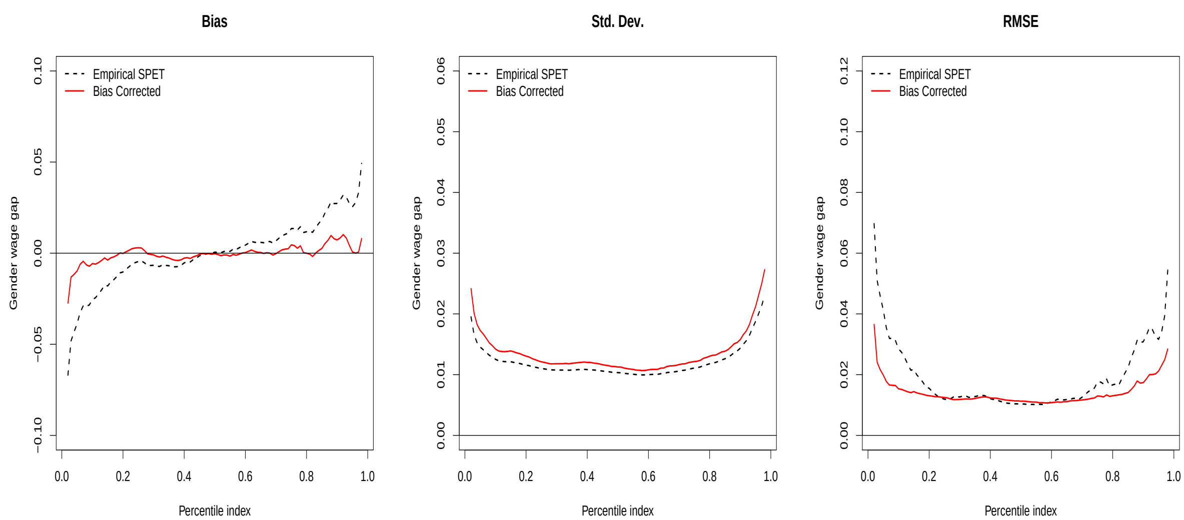

FIGURE 3. Bias, standard deviation and root mean square error of empirical and bias corrected SPE functions. Results obtained from 500 repetitions of a design calibrated to the CPS 2015 data.

simulated sample $\{(Y_i, X_i) : 1 \leqslant i \leqslant n\}$, we reestimate the model by least squares, obtain the empirical SPE-function on the treated over a grid of percentile indexes $\mathcal{U} = \{0.02, 0.03, \ldots, 0.98\}$, and construct a 90% uniform confidence band for the SPE-function using Algorithm 2.1 with standard exponential weights and $B = 200$. We repeat this procedure 500 times.

Figure 3 reports bias, standard deviation (Std. Dev.) and root mean squared error (RMSE) of the empirical and bias corrected SPE functions, see Remark 2.3. We find that the empirical SPE displays negative bias in the lower tail and positive bias in the upper tail, which are reduced by the bootstrap bias correction. The correction slightly increases dispersion, but reduces overall rmse for most percentiles, specially at the tails. Table 3 reports the empirical coverage of 90% confidence bands constructed around the empirical and bias corrected SPE functions. Here we find that the uncorrected bands undercover the entire SPE function, whereas the corrected bands have coverage above the nominal level. One possible reason for the overcoverage of the bootstrap corrected bands is that we keep the covariates fixed across samples, which is not accounted by the bootstrap procedure. To sum up, we find that the bootstrap corrections of Remark 2.3 reduce the bias of the empirical SPE and improve the coverage of the confidence bands in finite samples.

Table 3. Coverage of 90% Confidence Bands

	Uncorrected	Bootstrap Bias Corrected
Coverage	0.82	0.98

Notes: 500 simulations and 200 bootstrap repetitions.
DGP calibrated to CPS 2015.

Appendix I. Effect of Race on Mortgage Denials

To study the effect of race in the bank decisions of mortgage denials or racial mortgage denial gap, we use data on mortgage applications in Boston from 1990 (see Munnell, Tootell, Browne, and McEneaney (1996)). The Federal Reserve Bank of Boston collected these data in relation to the Home Mortgage Disclosure Act (HMDA), which 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 includes $2,380$ 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 for 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 banks take into account in the mortgage decisions. These include the monthly debt to income ratio; monthly housing expenses to income ratio; a categorial 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; and three indicators for self-employed, single, and high school graduate.

TABLE 4. Descriptive Statistics of Mortgage Applicants

	All	Black	White
Deny	0.12	0.28	0.09
Black	0.14	1.00	0.00
Debt-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 graduate	0.98	0.97	0.99
number of observations	2,380	339	2,041

Table 4 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.

Figure 4 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 of the main text using a logit model with $P(X) = X = (T, W)$ and $\widehat{\mu}$ equal to the empirical distribution of X in the whole sample. The confidence bands are constructed using Algorithm 2.1 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 monotonize 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 mortgage denial gap. Table 5 shows the results of the classification analysis, answering the question "who is affected the most and who the least?" The table shows that the 10% of the applicants *most affected* by

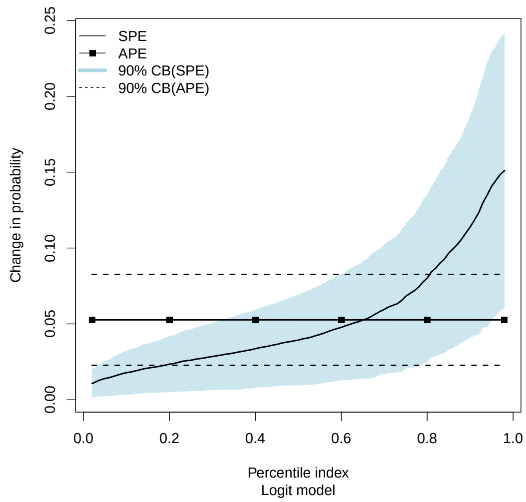

FIGURE 4. APE and SPE (introduced in this paper) of being black on the probability of mortgage denial. Estimates and 90% bootstrap uniform confidence bands (derived in this paper) based on a logit model are shown.

racial mortgage denial gap are *more likely* to have either of the following characteristics relative to the 10% of the *least affected* applicants: self employed, single, black, high debt to income ratio, high expense to income ratio, high loan to value ratio, medium or high loan-to-income ratio, bad consumer or credit scores, and credit problems.

References

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.

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. MCENEANEY (1996): "Mortgage Lending in Boston: Interpreting HMDA Data," *The American Economic Review*, 86(1), pp. 25–53.

RESNICK, S. I. (1987): *Extreme values, regular variation, and point processes*, vol. 4 of *Applied Probability. A Series of the Applied Probability Trust*. Springer-Verlag, New York.

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.

Table 5. Who is 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	10% Most Affected PE $> .11$		10% Least Affected PE $< .018$	
Deny	0.44	(0.03)	0.11	(0.04)
Black	0.37	(0.04)	0.07	(0.02)
Debt-to-income	0.39	(0.01)	0.25	(0.02)
Expenses-to-income	0.28	(0.01)	0.21	(0.02)
Bad consumer credit	4.64	(0.25)	1.31	(0.09)
Bad mortgage credit	1.99	(0.07)	1.37	(0.12)
Credit problems	0.45	(0.05)	0.05	(0.02)
Denied mortgage insurance	0.01	(0.01)	0.06	(0.04)
Medium loan-to-house	0.58	(0.06)	0.07	(0.04)
High loan-to-house	0.13	(0.03)	0.02	(0.01)
Self employed	0.18	(0.05)	0.05	(0.03)
Single	0.59	(0.05)	0.11	(0.06)
High school grad	0.93	(0.03)	1.00	(0.01)

Std. errors in parentheses obtained by bootstrap with 200 repetitions.