

RANDOM SETS FROM THE PERSPECTIVE OF METRIC STATISTICS

DAISUKE KURISU, YUTA OKAMOTO, AND TAISUKE OTSU

ABSTRACT. Since the seminal work by Beresteanu and Molinari (2008), the random set theory and related inference methods have been widely applied in partially identified econometric models. Meanwhile, there is an emerging field in statistics for studying random objects in metric spaces, called metric statistics. This paper clarifies a relationship between two fundamental concepts in these literatures, the Aumann and Fréchet means, and presents some applications of metric statistics to econometric problems involving random sets.

1. INTRODUCTION

Since the seminal work by Beresteanu and Molinari (2008), the random set theory and related inference methods have been widely applied in partially identified econometric models; see Molchanov and Molinari (2018) and Molinari (2020) for surveys in econometrics, and Molchanov (2017) for a comprehensive overview on the random set theory. The random set approach characterizes an identification region of interest by utilizing the Aumann (1965) mean for set valued random variables (SVRVs) and constructs its sample analog estimator by using Minkowski averages of SVRVs. Obviously the key ingredient for partial identification analysis using random sets is the Aumann mean, which is a natural and convenient extension of the conventional mean for Euclidean random variables to SVRVs.

On the other hand, in the recent statistics literature for analyzing complex data, there is an emerging and rapidly growing field, called *metric statistics* (or statistics for random objects); see Dubey et al. (2024) for an overview and references therein. Metric statistics is concerned with complex data situated in a metric space, and popular examples include distributional data, network data, symmetric positive definite matrices, trees, data on Riemannian manifolds, among others. In this literature, a fundamental notion to characterize the population mean for metric valued random objects is the Fréchet (1948) mean.

Given these literatures, it is natural to ask whether there is a relationship between these notions of population means, and also whether the methodologies of metric statistics can shed new light on econometric analysis of random sets. This is of course not the first paper addressing these issues. For example, Section 3.2 of Molchanov (2017) introduced the Fréchet mean and mentioned that it

Date: November 17, 2025.

Key words and phrases. Aumann mean, Fréchet mean, global Fréchet regression, metric statistics, random set theory.

The work of D. K. is partially supported by JSPS KAKENHI Grant Number 23K12456.

can be applied to SVRVs that situated in the Hausdorff metric space. Li et al. (2021) mentioned that their local regression smoother based on the Minkowski average can be interpreted as the sample Fréchet mean. Although these discussions are highly insightful, their main focuses are on the conventional random set analysis using the Aumann mean, and the analysis using the Fréchet mean is not pursued. On the other hand, to the best of our knowledge, there is no formal study in the literature of metric statistics on SVRVs. This paper is written to fill this gap.

In particular, this paper makes two contributions. First, we formally study the relationship between the Aumann and Fréchet means, and establish the equivalence of these notions of population means in the space of nonempty compact and convex sets equipped with the L^2 -metric based on their support functions. A key ingredient to establish such equivalence is an isometric embedding of a general metric space into a Hilbert space. Second, we apply or extend some methodologies of metric statistics to econometric problems that involve random sets. After introducing the global Fréchet regression (GFR) by Petersen and Müller (2019), we apply the GFR to the projection model of Euclidean covariates, and clarify the relation with the set valued best linear predictor by Beresteanu and Molinari (2008). Furthermore, we present some extensions of the GFR for an errors-in-variables model with set valued outcomes and a missing set valued data problem.

This paper is organized as follows. Section 2 discusses our main result, equivalence of the Aumann and Fréchet means. Then Section 3 presents applications of metric statistics for random sets.

2. MAIN RESULT

This section is devoted to discuss the relationship between the Aumann mean and the Fréchet mean of SVRVs. After introducing our basic setup and the Aumann mean (Section 2.1), we introduce a metric space based on the support function (Section 2.2). Then we introduce the Fréchet mean and present a key property, isometric embedding (Section 2.3). Based on these preparations, Section 2.4 establishes the equivalence of the Aumann and Fréchet means of SVRVs.

2.1. Setup and Aumann mean. We follow the notation in Beresteanu and Molinari (2008) and introduce our basic setup. Let $(\Omega, \mathcal{A}, \mu)$ be a measurable space, and $K(\mathbb{R}^d)$ be the collection of all nonempty closed subsets of \mathbb{R}^d . A random element $F : \Omega \rightarrow K(\mathbb{R}^d)$ is called an SVRV. Let $K_k(\mathbb{R}^d)$ be the set of nonempty compact subsets of \mathbb{R}^d , and $K_{kc}(\mathbb{R}^d)$ denote the set of nonempty compact and convex subsets of \mathbb{R}^d . For an SVRV $F \in K_k(\mathbb{R}^d)$, the Aumann mean is defined as follows.

Let $L^1 = L^1(\Omega, \mathbb{R}^d)$ denote the space of measurable random variables with values in \mathbb{R}^d s.t. $\|\xi\|_1 = \mathbb{E}[\|\xi\|] < \infty$, $\mathcal{S}(F)$ denote the set of all measurable selections (or points) from a set F , and $\mathcal{S}^1(F) = \mathcal{S}(F) \cap L^1$. The Aumann mean of an SVRV $F \in K_k(\mathbb{R}^d)$ is defined as

$$\mathbb{E}[F] = \{\mathbb{E}[f] : f \in \mathcal{S}^1(F)\}. \quad (1)$$

Similarly, the conditional Aumann mean of F given $X \in \mathcal{X}$ is defined as $\mathbb{E}[F|X] = \{\mathbb{E}[f|X] : f \in \mathcal{S}^1(F)\}$, where \mathcal{X} is a subset of \mathbb{R}^p . The Aumann mean is a natural generalization of the conventional mean

for Euclidean random variables and plays a fundamental role in random set theory and its econometric and statistical applications; see Molchanov (2017), Molchanov and Molinari (2018), and Molinari (2020).

Hereafter we focus on SVRVs on the set of nonempty compact and convex subsets, $K_{kc}(\mathbb{R}^d)$, and characterize the Aumann mean $\mathbb{E}[F]$ for $F \in K_{kc}(\mathbb{R}^d)$ by using the notion of the Fréchet mean for random objects in a metric space, which has been increasingly popular in recent literature on metric statistics; see Dubey et al. (2024) for example.

2.2. Support function and metric on $K_{kc}(\mathbb{R}^d)$. A key ingredient to clarify the relationship between the Aumann and Fréchet means is to choose a proper metric on the set of nonempty compact and convex subsets $K_{kc}(\mathbb{R}^d)$. To this end, we introduce the support function $s(p, F) = \sup_{f \in F} \langle p, f \rangle$ for $F \in K_{kc}(\mathbb{R}^d)$ over $p \in \mathbb{S}^{d-1}$, where $\mathbb{S}^{d-1} = \{x \in \mathbb{R}^d : \|x\| = 1\}$ is the unit sphere in \mathbb{R}^d . For an SVRV $F \in K_{kc}(\mathbb{R}^d)$, the Aumann mean $\mathbb{E}[F]$ is equivalently characterized by its support function.

For $F, G \in K_{kc}(\mathbb{R}^d)$, define

$$d_{kc}(F, G) = \left(\int_{\mathbb{S}^{d-1}} \{s(p, F) - s(p, G)\}^2 dp \right)^{1/2} =: \|s(\cdot, F) - s(\cdot, G)\|_{2, \mathbb{S}^{d-1}}.$$

This paper employs d_{kc} as a metric on $K_{kc}(\mathbb{R}^d)$ and establishes the equivalence of the Aumann and Fréchet means. We close this subsection by presenting some properties related to d_{kc} . Let $d_H(\cdot, \cdot)$ be the Hausdorff metric on $K_k(\mathbb{R}^d)$ defined by $d_H(F, G) = \max\{\sup_{f \in F} \inf_{g \in G} \|f - g\|, \sup_{g \in G} \inf_{f \in F} \|f - g\|\}$ for $F, G \in K_k(\mathbb{R}^d)$, $K_{kc}^B(\mathbb{R}^d) = \{F \in K_{kc}(\mathbb{R}^d) : \sup_{f \in F} \|f\| \leq B\}$ for a positive constant B , $L^2(\mathbb{S}^{d-1})$ denote the space of functions such that $\|f\|_{2, \mathbb{S}^{d-1}} < \infty$, and $\Psi : K_k(\mathbb{R}^d) \rightarrow L^2(\mathbb{S}^{d-1})$ be a map such that $\Psi(F) = s(\cdot, F)$.

Lemma 2.1.

- (i) d_{kc} is a metric on $K_{kc}(\mathbb{R}^d)$.
- (ii) Let $\{F_n\}_{n \geq 0} \subset K_{kc}^B(\mathbb{R}^d)$ be a sequence of compact convex sets. Then $\lim_{n \rightarrow \infty} d_{kc}(F_n, F_0) = 0$ if and only if $\lim_{n \rightarrow \infty} d_H(F_n, F_0) = 0$.
- (iii) $K_{kc}^B(\mathbb{R}^d)$ is a bounded closed convex subset of $K_{kc}(\mathbb{R}^d)$ with respect to d_{kc} .
- (iv) $\Psi(K_{kc}^B(\mathbb{R}^d))$ is a bounded closed convex subset of $L^2(\mathbb{S}^{d-1})$.

Lemma 2.1 establishes several topological properties of compact convex sets under the metric d_{kc} . In particular, Lemma 2.1 (ii) shows that convergence under the Hausdorff metric, which can be defined as the sup-norm distance between support functions on \mathbb{S}^{d-1} (see Lemma B.9), is equivalent to convergence under the L^2 -based distance d_{kc} . This equivalence motivates the use of d_{kc} , which is analytically more tractable, for statistical analysis of random compact convex sets. In Section 2.4, we show that the Aumann mean (and its sample counterpart, the normalized Minkowski sum) can be characterized as the Fréchet mean, that is, the minimizer of $\mathbb{E}[d_{kc}^2(\nu, F)]$, the expectation of the squared distance between a data descriptor $\nu \in K_{kc}^B(\mathbb{R}^d)$ and an SVRV $F \in K_{kc}^B(\mathbb{R}^d)$ (and its

sample version). Finally, Lemma 2.1 (iii) and (iv) guarantee the existence and uniqueness of the Fréchet mean.

2.3. Fréchet mean and isometric embedding. We now introduce the Fréchet mean for random objects. Let \mathcal{X} be a subset of \mathbb{R}^p , (\mathcal{M}, d) be a separable metric space, and $(Y, X) \in \mathcal{M} \times \mathcal{X}$ be a pair of random elements. The Fréchet mean $\mathbb{E}_\oplus[Y]$ of Y is defined as a minimizer of the population Fréchet function $Q(\nu) = \mathbb{E}[d^2(\nu, Y)]$, that is,

$$\mathbb{E}_\oplus[Y] \in \operatorname{argmin}_{\nu \in \mathcal{M}} Q(\nu).$$

Similarly, the conditional Fréchet mean $\mathbb{E}_\oplus[Y|X]$ of Y given X is defined as

$$\mathbb{E}_\oplus[Y|X] \in \operatorname{argmin}_{\nu \in \mathcal{M}} Q(\nu|X), \quad Q(\nu|X) = \mathbb{E}[d^2(\nu, Y)|X].$$

The Fréchet mean is a direct generalization of the conventional mean for the Euclidean space toward a general metric space, and its statistical analysis has been increasingly popular in recent literature. The key step to establish the relationship between the Aumann and Fréchet means is to consider an isometric embedding of a general metric space into a Hilbert space. To this end, we impose the following assumptions.

Assumption 2.1.

- (i) *There exist a Hilbert space \mathcal{H} equipped with an inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$, induced norm $\| \cdot \|_{\mathcal{H}}$, and a continuous injection $\Psi : \mathcal{M} \rightarrow \mathcal{H}$ such that $\Psi : \mathcal{M} \rightarrow \Psi(\mathcal{M})$ is isometry, i.e., $d(\alpha, \beta) = \|\Psi(\alpha) - \Psi(\beta)\|_{\mathcal{H}}$ for any $\alpha, \beta \in \mathcal{M}$.*
- (ii) *The set $\Psi(\mathcal{M})$ is a nonempty closed convex set in \mathcal{H} .*

Note that $(K_{kc}^B(\mathbb{R}^d), d_{kc})$ satisfies Assumption 2.1. Indeed, setting $\Psi(F) = s(\cdot, F)$ implies that $K_{kc}^B(\mathbb{R}^d)$ admits an isometric embedding into $L^2(\mathbb{S}^{d-1})$. Furthermore, by Lemma 2.1 (iv), the image $\Psi(K_{kc}^B(\mathbb{R}^d))$ satisfies Assumption 2.1 (ii). Let (\mathcal{M}, d) be a metric space. The metric d^2 is of negative type if, for all $n \geq 2$, $\nu_1, \dots, \nu_n \in \mathcal{M}$ and $\alpha_1, \dots, \alpha_n \in \mathbb{R}$ with $\sum_{i=1}^n \alpha_i = 0$, we have $\sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j d^2(\nu_i, \nu_j) \leq 0$. A sufficient condition for Assumption 2.1 (i) is provided in Proposition 3 in Sejdinovic et al. (2013) for example, which implies that if d^2 is of negative type, then this condition is satisfied; see also Schoenberg (1938). Many of the common metric spaces studied in metric statistics are known to satisfy Assumption 2.1. Examples include the 2-Wasserstein space for univariate probability distributions, the space of symmetric positive (semi)definite matrices endowed with the Frobenius, power, or log-Euclidean metric, the space of graph Laplacians equipped with the Frobenius metric, and the space of compositional data endowed with the Aitchison metric. See also Appendix C.1 in Kurisu et al. (2025b).

Under these assumptions, we obtain the following characterizations of the Fréchet mean. For a random element Z taking values in a Hilbert space \mathcal{H} , we define its expectation as the Riesz representation of the linear functional that maps $h \in \mathcal{H}$ to $\mathbb{E}[\langle h, Z \rangle_{\mathcal{H}}] \in \mathbb{R}$.

Proposition 2.1. *Suppose that Assumption 2.1 holds true and $\mathbb{E}[\|\Psi(Y)\|_{\mathcal{H}}] < \infty$. Then the following results hold.*

- (i) *The object $\Psi^{-1}(\mathbb{E}[\Psi(Y)])$ is well defined, and $\mathbb{E}_{\oplus}[Y] = \Psi^{-1}(\mathbb{E}[\Psi(Y)])$.*
- (ii) *$\Psi(\mathbb{E}_{\oplus}[Y]) = \mathbb{E}[\Psi(\mathbb{E}_{\oplus}[Y|X])]$.*

According to Proposition 2.1 (i), the Fréchet mean of Y is obtained as the pullback, through Ψ , of the expectation of $\Psi(Y)$ in the image space $\Psi(K_{kc}^B(\mathbb{R}^d))$. Proposition 2.1 (ii) can be interpreted as a law of iterated expectation for random objects, which naturally generalizes the corresponding result for Euclidean random variables.

2.4. Equivalence of Aumann and Fréchet means. We now apply the Fréchet mean to SVRVs in the metric space $(K_{kc}^B(\mathbb{R}^d), d_{kc})$. Let $(F, X) \in K_{kc}^B(\mathbb{R}^d) \times \mathcal{X}$ be a pair of random elements. For $F, G \subset \mathbb{R}^d$, let $F \oplus G = \{f + g : f \in F, g \in G\}$ denote the Minkowski sum of F and G . The population Fréchet mean $\mathbb{E}_{\oplus}[F]$ of F with respect to the metric d_{kc} is defined as a minimizer of the population Fréchet function $\mathcal{Q}(\nu) = \mathbb{E}[d_{kc}^2(\nu, F)]$, that is,

$$\mathbb{E}_{\oplus}[F] \in \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \mathcal{Q}(\nu).$$

Similarly, the population conditional Fréchet mean of F given X is defined as

$$\mathbb{E}_{\oplus}[F|X] \in \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \mathcal{Q}(\nu|X), \quad \mathcal{Q}(\nu|X) = \mathbb{E}[d_{kc}^2(\nu, F)|X].$$

The main results of this paper are presented as follows.

Proposition 2.2. *For an SVRV $F : \Omega \rightarrow K_{kc}^B(\mathbb{R}^d)$, the following results hold true.*

- (i) *$\mathbb{E}_{\oplus}[F]$ and $\mathbb{E}_{\oplus}[F|X]$ with respect to the metric d_{kc} uniquely exist.*
- (ii) *$\mathbb{E}_{\oplus}[F] = \mathbb{E}[F]$.*
- (iii) *$\mathbb{E}_{\oplus}[F|X] = \mathbb{E}[F|X]$ and $\mathbb{E}_{\oplus}[F] = \mathbb{E}_{\oplus}[\mathbb{E}_{\oplus}[F|X]]$.*

Proposition 2.2 (i) guarantees uniqueness of the Fréchet means $\mathbb{E}_{\oplus}[F]$ and $\mathbb{E}_{\oplus}[F|X]$. Proposition 2.2 (ii) and (iii) are our main results, equivalence of the Fréchet and conditional Fréchet means to the Aumann and conditional Aumann means, respectively. Furthermore, Proposition 2.2 (iii) also provides the law of iterated expectations for the Fréchet mean.

Note that an SVRV $F \in K_{kc}(\mathbb{R}^d)$ is a metric space-valued random element, which is called random object. See e.g., Marron and Alonso (2014) and Marron and Dryden (2021) for a review. By Proposition 2.2, the (conditional) Aumann mean of F can be defined as the (conditional) Fréchet mean of F under the metric d_{kc} . From this viewpoint, statistical analysis of SVRVs falls within the scope of metric statistics, a field that has seen remarkable development in recent years; see Dubey et al. (2024) for example.

Remark 2.1. *We can also show that the sample Fréchet mean for $\mathbb{E}_{\oplus}[F]$ based on SVRVs $F_1, \dots, F_n \in K_{kc}^B(\mathbb{R}^d)$ coincides with the conventional sample Minkowski mean, which is a sample counterpart of the Aumann mean. More precisely, the sample Fréchet mean $\mu_{\oplus, n}$ of F_1, \dots, F_n with respect to*

the metric d_{kc} is defined as a minimizer of the sample Fréchet function $\mathcal{Q}_n(\nu) = \frac{1}{n} \sum_{i=1}^n d_{kc}^2(\nu, F_i)$, that is

$$\mu_{\oplus, n} \in \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \mathcal{Q}_n(\nu).$$

The equivalence of the sample Minkowski and Fréchet means is presented as follows.

Corollary 2.1. *Let $F_1, \dots, F_n \in K_{kc}^B(\mathbb{R}^d)$ be SVRVs. Then there exists a unique sample Fréchet mean $\mu_{\oplus, n}$ with respect to d_{kc} and it coincides with the sample Minkowski mean, that is, $\mu_{\oplus, n} = \frac{1}{n} \bigoplus_{i=1}^n F_i$, where $\bigoplus_{i=1}^n F_i$ is the Minkowski sum of F_1, \dots, F_n .*

3. METRIC STATISTICS FOR RANDOM SETS

This section presents several methods developed within metric statistics that can be applied to the analysis of SVRVs. For $t \in \mathbb{R}$, define $tF = \{tf : f \in F\}$. Let $\mathcal{X} \subset \mathbb{R}^p$ be a compact set.

3.1. Linear regression as a weighted Fréchet mean. To motivate, consider the linear projection model for an Euclidean outcome $Y \in \mathbb{R}$ and Euclidean covariates $X \in \mathbb{R}^p$,

$$Y = \theta_0^* + (\theta_1^*)'(X - \mu) + \varepsilon, \quad \mathbb{E}[X\varepsilon] = 0,$$

where $\mu = \mathbb{E}[X]$ and the vector (θ_0^*, θ_1^*) solves

$$(\theta_0^*, \theta_1^*) = \operatorname{argmin}_{(\theta_0, \theta_1) \in \mathbb{R}^{p+1}} \mathbb{E}[(\mathbb{E}[Y|X] - \{\theta_0 + \theta_1'(X - \mu)\})^2].$$

Let $\sigma_{YX} = \mathbb{E}[Y(X - \mu)]$ and $\Sigma = \mathbb{E}[(X - \mu)(X - \mu)']$. Assume that Σ is invertible. Then we have $\theta_0^* = \mathbb{E}[Y]$ and $\theta_1^* = \Sigma^{-1}\sigma_{YX}$, and obtain the regression function

$$\begin{aligned} m(x) &:= \theta_0^* + (\theta_1^*)'(x - \mu) = \mathbb{E}[Y] + (x - \mu)\Sigma^{-1}\sigma_{YX} \\ &= \mathbb{E}[Y + (x - \mu)'\Sigma^{-1}(X - \mu)Y] = \mathbb{E}[w(x, X)Y], \end{aligned}$$

where $w(x, z) = 1 + (x - \mu)'\Sigma^{-1}(z - \mu)'$. Since $\mathbb{E}[w(x, X)] = 1$, the regression function $m(x)$ is characterized as the solution

$$m(x) = \operatorname{argmin}_{y \in \mathbb{R}} \mathbb{E}[w(x, X)d_E^2(y, Y)],$$

where d_E is the standard Euclidean metric.

By extending the above representation of $m(x)$, the global Fréchet regression (GFR) function for an outcome Y in a general metric space (\mathcal{M}, d) is defined as follows (Petersen and Müller, 2019):

$$m_{\oplus}(x) = \operatorname{argmin}_{\nu \in \mathcal{M}} \mathbb{E}[w(x, X)d^2(\nu, Y)].$$

Let $\{Y_i, X_i\}_{i=1}^n$ be an i.i.d. sample from the joint distribution of $(Y, X) \in \mathcal{M} \times \mathbb{R}^p$. Then the GFR estimator is given as the sample counterpart of m_{\oplus} , that is

$$\hat{m}_{\oplus}(x) = \operatorname{argmin}_{\nu \in \mathcal{M}} \frac{1}{n} \sum_{i=1}^n \hat{w}(x, X_i)d^2(\nu, Y_i), \quad (2)$$

where $\hat{w}(x, z) = 1 + (x - \bar{X})'\hat{\Sigma}^{-1}(z - \bar{X})'$, $\bar{X} = n^{-1} \sum_{i=1}^n X_i$, and $\hat{\Sigma} = n^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})'$.

3.2. Global Fréchet regression for random sets. We now apply the GFR for a pair of random elements $(F, X) \in K_{kc}^B(\mathbb{R}^d) \times \mathcal{X}$, which is defined as

$$m_{\oplus}(x) = \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \mathbb{E}[w(x, X)d_{kc}^2(\nu, F)].$$

In order to derive the explicit form of $m_{\oplus}(x)$, recall $\Psi : K_{kc}^B(\mathbb{R}^d) \ni F \mapsto s(\cdot, F) \in L^2(\mathbb{S}^{d-1})$, and define

$$m_{\oplus, \Psi}(x) = \operatorname{argmin}_{h \in L^2(\mathbb{S}^{d-1})} \mathbb{E}[w(x, X)\|h - \Psi(F)\|_{2, \mathbb{S}^{d-1}}].$$

Then we have

$$m_{\oplus, \Psi}(x) = \mathbb{E}[w(x, X)\Psi(F)] = \mathbb{E}[w(x, X)\Psi(\mathbb{E}_{\oplus}[F|X])] = \mathbb{E}[w(x, X)s(\cdot, \mathbb{E}_{\oplus}[F|X])],$$

where the second equality follows from Proposition 2.1.

Assume that $\inf_{z \in \mathcal{X}} w(x, z) \geq 1$, which guarantees $w(x, X)F \in K_{kc}^B(\mathbb{R}^d)$. Then from Proposition 2.2 (ii) and Lemma B.6 (i), we have

$$m_{\oplus, \Psi}(x) = \mathbb{E}[s(\cdot, w(x, X)\mathbb{E}_{\oplus}[F|X])] = s(\cdot, \mathbb{E}[w(x, X)F]) = \Psi(\mathbb{E}[w(x, X)F]),$$

which yields the expression of $m_{\oplus}(x)$ by the Aumann means:

$$\begin{aligned} m_{\oplus}(x) &= \mathbb{E}[w(x, X)F] = \mathbb{E}[F] \oplus (x - \mu)' \Sigma^{-1} \mathbb{E}[(X - \mu)F] \\ &= (1, (x - \mu)') \begin{pmatrix} 1 & 0 \\ 0 & \Sigma \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{E}[F] \\ \mathbb{E}[(X - \mu)F] \end{pmatrix} =: (1, (x - \mu)') \Theta. \end{aligned}$$

Note that Θ corresponds to the population set valued best linear predictor in Beresteanu and Molinari (2008). On the other hand, if $\inf_{z \in \mathcal{X}} w(x, z) < 1$, we have

$$m_{\oplus, \Psi}(x) \leq \mathbb{E}[s(\cdot, w(x, X)\mathbb{E}_{\oplus}[F|X])] = \Psi(\mathbb{E}[w(x, X)F]). \quad (3)$$

Specifically, it is possible to provide an example such that $m_{\oplus, \Psi}(x) < \Psi(\mathbb{E}[w(x, X)F])$ (see Appendix A.6). Therefore, for the GFR with a general weight function $w(x, z)$, we compute $m_{\oplus}(x)$ by projecting $m_{\oplus, \Psi}(x)$ on $\Psi(K_{kc}^B(\mathbb{R}^d))$, that is,

$$m_{\oplus}(x) = \Psi^{-1}(\tilde{m}_{\oplus, \Psi}(x)), \quad \tilde{m}_{\oplus, \Psi}(x) = \operatorname{argmin}_{h \in \Psi(K_{kc}^B(\mathbb{R}^d))} \|h - m_{\oplus, \Psi}(x)\|_{2, \mathbb{S}^{d-1}}.$$

The following result shows that $m_{\oplus}(x)$ is in fact the GFR function.

Proposition 3.1. *$m_{\oplus}(x) = \Psi^{-1}(\tilde{m}_{\oplus, \Psi}(x))$ is the GFR function, that is, $\Psi^{-1}(\tilde{m}_{\oplus, \Psi}(x))$ is the unique minimizer of $\mathbb{E}[w(x, X)d_{kc}^2(\nu, F)]$ over $\nu \in K_{kc}^B(\mathbb{R}^d)$. Additionally, if $\sup_{z \in \mathcal{X}} |w(x, z)| < \infty$ and $m_{\oplus, \Psi}(x) \in \Psi(K_{kc}(\mathbb{R}^d))$, then $m_{\oplus}(x) \subset \mathbb{E}[w(x, X)F]$.*

Figure 3.1 illustrates the relation between $m_{\oplus}(x)$ and $\mathbb{E}[w(x, X)F]$ when $d = 1$. The colored region corresponds to the set of support functions. Note that, in the one-dimensional case, F can be written as $F = [-a, b]$ with $b + a \geq 0$. Note also that the support function $s(p, F)$ of $F \in K_{kc}(\mathbb{R})$ is characterized by $s(1, F)$ and $s(-1, F)$ and satisfies $b = s(1, F) \geq -s(-1, F) = -a$. Hence the set of support functions $\Psi(K_{kc}(\mathbb{R})) = \{s(\cdot, F) : F \in K_{kc}(\mathbb{R})\}$ can be identified with $\{(a, b) \in \mathbb{R}^2 :$

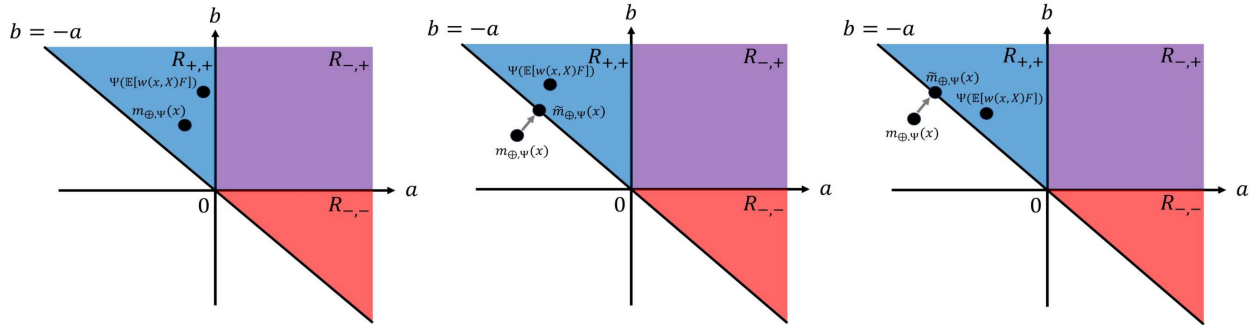


FIGURE 3.1. Illustration of the relation between $m_{\oplus}(x)$ and $\mathbb{E}[w(x, X)F]$ when $d = 1$.

$b + a \geq 0\}$. The blue region $R_{+,+}$, the purple region $R_{-,+}$, and the red region $R_{-,-}$ correspond to the sets K_1, K_2 , and K_3 such that $K_1 \sqcup K_2 \sqcup K_3 = K_{kc}(\mathbb{R})$ where $K_1 = \{[x, y] : 0 < x \leq y < \infty\}$, $K_2 = \{[x, y] : -\infty < x \leq 0 \leq y < \infty\}$, and $K_3 = \{[x, y] : -\infty < x \leq y < 0\}$. In each figure, black points correspond to $m_{\oplus, \Psi}$, $\tilde{m}_{\oplus, \Psi}$, and $\Psi(\mathbb{E}[w(x, X)F])$. The left figure represents the case $m_{\oplus, \Psi}(x) \in \Psi(K_{kc}(\mathbb{R}))$, in which case $m_{\oplus}(x) \subset \mathbb{E}[w(x, X)F]$. The middle figure corresponds to the case of $m_{\oplus, \Psi}(x) \notin \Psi(K_{kc}(\mathbb{R}))$ and $m_{\oplus}(x) \subset \mathbb{E}[w(x, X)F]$. The right figure corresponds to the case of $m_{\oplus, \Psi}(x) \notin \Psi(K_{kc}(\mathbb{R}))$ and $m_{\oplus}(x) \not\subset \mathbb{E}[w(x, X)F]$.

Let $(F, X) \in K_{kc}^B(\mathbb{R}^d) \times \mathcal{X}$ be a pair of random elements and $\{F_i, X_i\}_{i=1}^n$ be an i.i.d. sample from the joint distribution of (F, X) . The uniform convergence rate of the GFR estimator in (2) can be established as follows.

Proposition 3.2. *Assume that $d \leq 4$. For a given $B_0 > 0$, we have*

$$\sup_{\|x\| \leq B_0} d_{kc}(\hat{m}_{\oplus}(x), m_{\oplus}(x)) = \begin{cases} O_p(n^{-\frac{1}{2(\alpha-1)}}) & \text{for any } \alpha > 2 \text{ when } d = 1, \\ O_p(n^{-\frac{1}{2(1+(d-1)/4)}) & \text{when } d \in \{2, 3, 4\}. \end{cases}$$

In Proposition 3.2, since $\alpha > 2$ can be chosen arbitrarily, the convergence rate of the GFR estimator can be arbitrarily close to the parametric rate when $d = 1$. The restriction $d \leq 4$ is due to the complexity of the metric space $(K_{kc}(\mathbb{R}^d), d_{kc})$. Specifically, let $N(\varepsilon, \mathcal{F}, d)$ be the covering number, that is, the minimal number of balls of radius ε (with respect to the metric d) needed to cover the set \mathcal{F} . We can see $N(\varepsilon, K_{kc}^B(\mathbb{R}^d), d_{kc}) \leq D\varepsilon^{-2}$ for $d = 1$ but $N(\varepsilon, K_{kc}^B(\mathbb{R}^d), d_{kc}) \leq e^{D\varepsilon^{-(d-1)/2}}$ for $d \geq 2$ where D is a positive constant which is independent of ε . To show Proposition 3.2 for $d \geq 2$, we need to verify $\int_0^1 \sqrt{\varepsilon^{-(d-1)/2}} d\varepsilon < \infty$, which requires $d \leq 4$. See the proof of Proposition 3.2 for details. The same comment applies to Proposition 3.3 below.

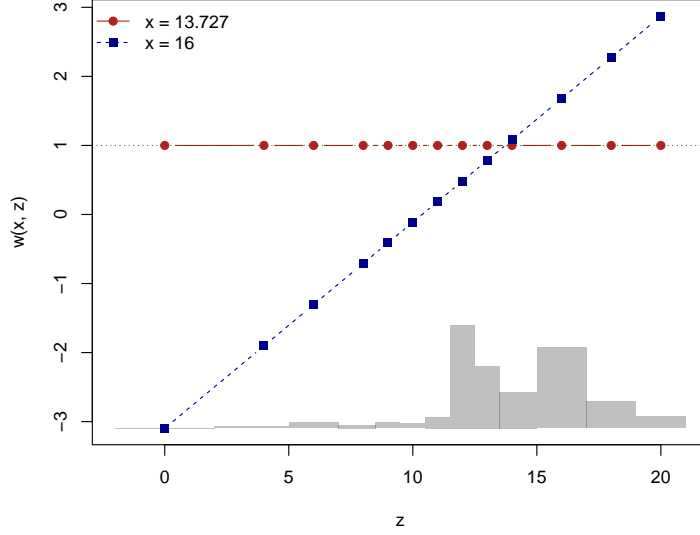


FIGURE 3.2. Weight functions $w(\mu, z)$ and $w(16, z)$ and histogram of X .

3.2.1. *Illustration.* As an illustration, consider the case of one-dimensional interval-valued random variable, i.e., $F \in K_{kc}^B(\mathbb{R})$. In this case, letting $F = [L, U]$, we have

$$m_{\oplus, \Psi}(x) = \mathbb{E}[w(x, X)s(p, F)] \text{ for } p \in \{-1, 1\} = \begin{cases} -\mathbb{E}[w(x, X)L] & \text{for } p = -1 \\ \mathbb{E}[w(x, X)U] & \text{for } p = 1 \end{cases}.$$

Also since $w(x, X)F = [\min\{w(x, X)L, w(x, X)U\}, \max\{w(x, X)L, w(x, X)U\}]$, we obtain

$$\Psi(\mathbb{E}[w(x, X)F]) = \begin{cases} -\mathbb{E}[\min\{w(x, X)L, w(x, X)U\}] & \text{if } p = -1 \\ \mathbb{E}[\max\{w(x, X)L, w(x, X)U\}] & \text{if } p = 1 \end{cases}.$$

Following Beresteanu and Molinari (2008), we consider the best linear prediction of the returns to education of log-wages for white men between the ages of 20 and 50. We artificially divide the support of the outcome variable into five equal intervals to construct interval data. We employ the March 2009 Current Population Survey data from Hansen (2022), and treat them as the population. Then we compare the GFR $m_{\oplus}(x) = \Psi^{-1}(\tilde{m}_{\oplus, \Psi}(x))$ and the Aumann mean $\mathbb{E}[w(x, X)F]$ at $x = \mu \approx 13.727$ and $x = 16$, corresponding to the Bachelor's degree. Note that $w(\mu, z) = 1 + (\mu - \mu)'\Sigma^{-1}(z - \mu) = 1$ (see Figure 3.2) so that $m_{\oplus}(x) = \mathbb{E}[w(x, X)F]$. In this case, the Aumann mean and the GFR are computed as

$$\mathbb{E}[w(\mu, X)F] = [9.624, 12.253] \quad \text{and} \quad m_{\oplus}(\mu) = [9.624, 12.253].$$

On the other hand, Figure 3.2 shows that $\inf_{z \in \mathcal{X}} w(16, z) < 1$. We can also confirm that $m_{\oplus, \Psi}(16)$ is

a proper support function, i.e., $m_{\oplus, \Psi}(16) \in \Psi(K_{kc}(\mathbb{R}))$. These suggest that $m_{\oplus}(x) \subset \mathbb{E}[w(x, X)F]$ (Proposition 3.1). Indeed, we obtain that

$$\mathbb{E}[w(16, X)F] = [9.891, 12.778] \quad \text{and} \quad m_{\oplus}(16) = [10.020, 12.649].$$

3.3. Errors-in-variables. The GFR approach can be naturally extended to errors-in-variables models. Suppose the object of interest is

$$m_{\oplus}(x) = \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \mathbb{E}[w(x, X)d_{kc}^2(\nu, F)].$$

However, Euclidean covariates X are mismeasured, and we instead observe $W = X + \varepsilon$, where ε is a measurement error such that $\mathbb{E}[\varepsilon|X, F] = 0$ and $\mathbb{E}[\varepsilon\varepsilon'] = \Sigma_{\varepsilon}$. In this setup, note that $\mathbb{E}[W] = \mu$, $\Sigma_W = \mathbb{E}[(W - \mathbb{E}[W])(W - \mathbb{E}[W])'] = \Sigma + \Sigma_{\varepsilon}$, and

$$\mathbb{E}[\{1 + (x - \mathbb{E}[W])'\Sigma_W^{-1}(W - \mathbb{E}[W])\}d_{kc}^2(\nu, F)] = \mathbb{E}[\{1 + (x - \mu)'(\Sigma + \Sigma_{\varepsilon})^{-1}(X - \mu)\}d_{kc}^2(\nu, F)].$$

Then the GFR estimator (2) with $\mathcal{M} = K_{kc}^B(\mathbb{R}^d)$, $d = d_{kc}$, and $Y_i = F_i$ is not consistent for $m_{\oplus}(x)$.

Suppose we observe instrumental variables $Z = X + v$ (or repeated measurements on X) such that $\mathbb{E}[v|X, F]$ and $\mathbb{E}[v\varepsilon'|X, F] = 0$. Then we can recover Σ as

$$\Sigma_{ZW} = \mathbb{E}[(Z - \mathbb{E}[Z])(W - \mathbb{E}[W])'] = \mathbb{E}[(X - \mu + v)(X - \mu + \varepsilon)'] = \Sigma.$$

Therefore, the GFR function $m_{\oplus}(x)$ can be estimated based on the alternative representation

$$m_{\oplus}(x) = \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \mathbb{E}[\tilde{w}(x, X)d_{kc}^2(\nu, F)],$$

where $\tilde{w}(x, z) = 1 + (x - \mu)'\Sigma_{ZW}^{-1}(z - \mu)$.

Other possible statistical methods for analyzing set-valued response variables with Euclidean covariates using the GFR or related regression models include Tucker et al. (2023) and Kurisu and Otsu (2025) for variable selection and model averaging, Bhattacharjee and Müller (2023) for an extension of the GFR to single index models, Petersen and Müller (2019) and Chen and Müller (2022) for extensions of the GFR to local linear regression.

3.4. Missing data. Consider the missing at random setup. Let $T \in \{0, 1\}$ be the indicator for missing data. We observe TF and X , where $(F, X) \in K_{kc}^B(\mathbb{R}^d) \times \mathcal{X}$ and assume F and T are independent given X . Let $\{F_i, X_i, T_i\}_{i=1}^n$ be an i.i.d. sample from the joint distribution of (F, X, T) . We are interested in the Fréchet mean of F , that is, $\mathbb{E}_{\oplus}[F] = \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \mathbb{E}[d_{kc}^2(\nu, F)]$.

Letting $e(x) = \mathbb{P}(T = 1|X = x)$, we can see that

$$\mathbb{E}\left[\frac{T}{e(X)}d_{kc}^2(\nu, F)\right] = \mathbb{E}\left[\frac{1}{e(X)}\mathbb{E}[Td_{kc}^2(\nu, F)|X]\right] = \mathbb{E}\left[\frac{1}{e(X)}\mathbb{E}[T|X]\mathbb{E}[d_{kc}^2(\nu, F)|X]\right] = \mathbb{E}[d_{kc}^2(\nu, F)].$$

Therefore, the Fréchet mean of F can be alternatively written as

$$\mathbb{E}_{\oplus}[F] = \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \mathbb{E}\left[\frac{T}{e(X)}d_{kc}^2(\nu, F)\right],$$

which can be estimated by

$$\hat{\mathbb{E}}_{\oplus}[F] = \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \frac{1}{n} \sum_{i=1}^n \frac{T_i}{\hat{e}(X_i)} d_{kc}^2(\nu, F_i) = \left(\frac{1}{n} \sum_{i=1}^n \frac{T_i}{\hat{e}(X_i)} \right)^{-1} \frac{1}{n} \bigoplus_{i=1}^n \frac{T_i}{\hat{e}(X_i)} F_i,$$

and $\hat{e}(x)$ is a nonnegative nonparametric estimator of $e(x)$.

One can see that $\hat{\mathbb{E}}_{\oplus}[F]$ is a modified version of inverse probability weighting estimator proposed in Kurisu et al. (2024):

$$\hat{\Theta}_1^{(\text{IPW})} = \frac{1}{n} \bigoplus_{i=1}^n \frac{T_i}{\hat{e}(X_i)} F_i. \quad (4)$$

See Appendix A.8 for the derivation of (4). The following result provides the convergence rate of $\hat{\mathbb{E}}_{\oplus}[F]$.

Proposition 3.3. *Suppose that the following conditions hold.*

- (i) *There exists a constant $\eta_0 \in (0, 1/2)$ such that $\eta_0 \leq e(x) \leq 1 - \eta_0$ for all $x \in \mathcal{X}$.*
- (ii) *$\sup_{x \in \mathcal{X}} |\hat{e}(x) - e(x)| = O_p(\rho_n)$, $\rho_n \rightarrow 0$ as $n \rightarrow \infty$.*
- (iii) *As $\delta \rightarrow 0$, $\int_0^1 \sqrt{\log N(\delta\varepsilon, B_{\delta'}(e), \|\cdot\|_{\infty})} d\varepsilon = O(\delta^{-\varpi_1})$ for some $\delta' > 0$ and $\varpi_1 \in (0, 1)$, where $\|e_1 - e_2\|_{\infty} = \sup_{x \in \mathcal{X}} |e_1(x) - e_2(x)|$ for $e_1, e_2 : \mathcal{X} \rightarrow \mathbb{R}$, and $B_{\gamma}(e)$ is the ball of radius $\gamma > 0$ centered at e .*

Then for $d \leq 4$, we have $d_{kc}(\hat{\mathbb{E}}_{\oplus}[F], \mathbb{E}_{\oplus}[F]) = O_p(n^{-\frac{1}{2(1+\max\{(d-1)/4, \varpi_1\})}} + \rho_n^{\varpi_2})$ for any $\varpi_2 \in (0, 1)$.

If the propensity score $e(x)$ belongs to a class of parametric models $\mathcal{M}_e = \{e(x; \varphi) : \varphi \in \Phi\}$ with a compact parameter space $\Phi \subset \mathbb{R}^q$, Condition (iii) is satisfied with $\int_0^1 \sqrt{\log N(\delta\varepsilon, B_{\delta'}(e), \|\cdot\|_{\infty})} d\varepsilon = O(-\log \delta)$ as $\delta \rightarrow 0$. In this case, it can be shown that convergence rate of $\hat{\mathbb{E}}_{\oplus}[F]$ is $O_p(n^{-\frac{1}{2(\alpha-1)}})$ for any $\alpha > 2$ when $d = 1$ and $O_p(n^{-\frac{1}{2(1+(d-1)/4}})$ when $d \in \{2, 3, 4\}$. For details on this point, see, for example, Section 4.1 of Kurisu et al. (2024).

In the recent literature, metric statistics has been extended to conduct causal inference for outcomes situated in a general metric space; see Kurisu et al. (2024), Zhou et al. (2025), Kurisu et al. (2025a,b), and Bhattacharjee et al. (2025). In sum, metric statistics can be a useful toolkit for econometric analysis on non-Euclidean economic data including SVRVs.

APPENDIX A. PROOFS

An SVRV F is called integrably bounded if $\sup\{\|f\| : f \in F\}$ has finite expectation.

A.1. Proof of Lemma 2.1.

A.1.1. *Proof of (i).* For any $F_1, F_2, F_3 \in K_{kc}(\mathbb{R}^d)$, it suffices to verify the following conditions. (a) $d_{kc}(F_1, F_2) = 0 \Leftrightarrow F_1 = F_2$, (b) $d_{kc}(F_1, F_2) = d_{kc}(F_2, F_1)$, and (c) $d_{kc}(F_1, F_2) \leq d_{kc}(F_1, F_3) + d_{kc}(F_3, F_2)$. Conditions (b) and (c) follow immediately from the definition of d_{kc} . Now we verify Condition (a). Observe that $d_{kc}(F_1, F_2) = 0 \Leftrightarrow s(p, F_1) \equiv s(p, F_2)$ for almost all $p \in \mathbb{S}^{d-1}$ and this implies $s(p, F_1) = s(p, F_2)$ for all $p \in \mathbb{S}^{d-1}$. Indeed, define $h(p) = s(p, F_1) - s(p, F_2)$ and assume that $s(p, F_1) = s(p, F_2)$ for all $p \in \mathbb{S}^{d-1}$. If there exists a point $p_0 \in \mathbb{S}^{d-1}$ such that $h(p_0) \neq 0$, then Lemma B.4 yields that there exists a neighborhood $U(p_0)$ of p_0 such that $h(p) \neq 0$ on $U(p_0)$. This contradicts the assumption. Then we have $s(p, F_1) = s(p, F_2)$ for all $p \in \mathbb{S}^{d-1}$. Since F_1 and F_2 are closed convex sets, we obtain $F_1 = F_2$.

A.1.2. *Proof of (ii).* First, we verify $\lim_{n \rightarrow \infty} d_H(F_n, F_0) = 0 \Rightarrow \lim_{n \rightarrow \infty} d_{kc}(F_n, F_0) = 0$. For any bounded closed convex sets, F, G , we have $d_{kc}(F, G) \leq \sqrt{\text{vol}(\mathbb{S}^{d-1})} \sup_{p \in \mathbb{S}^{d-1}} |s(p, F) - s(p, G)|$. Then Lemmas B.9 and B.10 yield $\lim_{n \rightarrow \infty} d_H(F_n, F_0) = 0 \Rightarrow \lim_{n \rightarrow \infty} d_{kc}(F_n, F_0) = 0$.

Next, we verify $\lim_{n \rightarrow \infty} d_{kc}(F_n, F_0) = 0 \Rightarrow \lim_{n \rightarrow \infty} d_H(F_n, F_0) = 0$. Note that $\sup_{p \in \mathbb{S}^{d-1}} |s(p, F_n)| \leq \sup_n (\sup_{f \in F_n} \|f\|) \leq B$ and hence the sequence $\{s(\cdot, F_n)\}_{n \geq 1}$ is uniformly bounded. Moreover, from the proof of Lemma B.4, we have $\sup_n |s(u, F_n) - s(v, F_n)| \leq B\|u - v\|$ for any $u, v \in \mathbb{S}^{d-1}$, and hence the sequence $\{s(\cdot, F_n)\}_{n \geq 1}$ is equicontinuous. Then from Arzelá-Ascoli theorem, there exist a subsequence $\{s(\cdot, F_{n_k})\}_{k \geq 1}$ and continuous function $t(\cdot)$ on \mathbb{S}^{d-1} such that $\lim_{k \rightarrow \infty} \sup_{p \in \mathbb{S}^{d-1}} |s(p, F_{n_k}) - t(p)| = 0$. This yields $\lim_{k \rightarrow \infty} \|s(\cdot, F_{n_k}) - t(\cdot)\|_{2, \mathbb{S}^{d-1}} \leq \text{vol}(\mathbb{S}^{d-1}) \lim_{k \rightarrow \infty} \sup_{p \in \mathbb{S}^{d-1}} |s(p, F_{n_k}) - t(p)| = 0$. Then we have $\|s(\cdot, F_0) - t(\cdot)\|_{2, \mathbb{S}^{d-1}} \leq \lim_{k \rightarrow \infty} \|s(\cdot, F_0) - s(\cdot, F_{n_k})\|_{2, \mathbb{S}^{d-1}} + \lim_{k \rightarrow \infty} \|s(\cdot, F_{n_k}) - t(\cdot)\|_{2, \mathbb{S}^{d-1}} = 0$. This yields $s(p, F_0) = t(p)$ for all $p \in \mathbb{S}^{d-1}$. Applying the same argument in the proof of Lemma 2.1 (i), we have $s(\cdot, F_0) = t(\cdot)$. Then we have $\lim_{n \rightarrow \infty} \sup_{p \in \mathbb{S}^{d-1}} |s(p, F_n) - s(p, F_0)| = 0$. Then from Lemma B.9, $\lim_{n \rightarrow \infty} d_{kc}(F_n, F_0) = 0 \Rightarrow \lim_{n \rightarrow \infty} d_H(F_n, F_0) = 0$ and Lemma B.10 yields that F is a bounded closed convex set.

A.1.3. *Proof of (iii).* The conclusion follows from Lemma 2.1 (ii) and the fact that $K_{kc}(\mathbb{R}^d)$ is closed with respect to d_H (Lemma B.10).

A.1.4. *Proof of (iv).* It is easy to verify the convexity. Now we show the closedness.

(Step 1) For a sequence of sets $\{F_n\} \subset K_{kc}^B(\mathbb{R}^d)$, assume that $\lim_{n \rightarrow \infty} \|s(\cdot, F_n) - t(\cdot)\|_{2, \mathbb{S}^{d-1}} = 0$. From the proof of Lemma 2.1 (ii), we have that $t(\cdot)$ is a continuous function with $\sup_{p \in \mathbb{S}^{d-1}} |t(p)| \leq \sup_n (\sup_{f \in F_n} \|f\|) \leq B$ and $\lim_{k \rightarrow \infty} \sup_{p \in \mathbb{S}^{d-1}} |s(p, F_{n_k}) - t(p)| = 0$ for some subsequence $\{s(\cdot, F_{n_k})\}_{k \geq 1}$.

(Step 2) Now we verify that $t(\cdot)$ is a support function of a bounded closed convex set $F \subset \mathbb{R}^d$. For this, we first extend $s(\cdot, F_{n_k})$ on \mathbb{S}^{d-1} to a function on \mathbb{R}^d . Define

$$\tilde{s}(x, F_{n_k}) = \begin{cases} \|x\|s\left(\frac{x}{\|x\|}, F_{n_k}\right) & x \neq 0 \\ 0 & x = 0 \end{cases}.$$

We also define $\tilde{t} : \mathbb{R}^d \rightarrow \mathbb{R}$ is the same way. Then $\lim_{k \rightarrow \infty} \sup_{p \in \mathbb{S}^{d-1}} |s(p, F_{n_k}) - t(p)| = 0$ implies $\lim_{k \rightarrow \infty} \sup_{x: \|x\| \leq R} |\tilde{s}(x, F_{n_k}) - \tilde{t}(x)| = 0$ for any $R > 0$.

Now we verify (a) positive homogeneity and (b) convexity of $\tilde{t}(\cdot)$ that characterize support functions (Schneider, 2014). For (a), pick any $\lambda \geq 0$ and $x \in \mathbb{R}^d$. We have

$$|\tilde{t}(\lambda x) - \lambda \tilde{t}(x)| \leq |\tilde{t}(\lambda x) - \tilde{s}(\lambda x, F_{n_k})| + |\tilde{s}(\lambda x, F_{n_k}) - \lambda \tilde{s}(x, F_{n_k})| + |\lambda \tilde{s}(x, F_{n_k}) - \lambda \tilde{t}(x)| \rightarrow 0,$$

as $k \rightarrow \infty$, which implies (a) as $\tilde{t}(\lambda x) = \lambda \tilde{t}(x)$. For (b), if \tilde{t} is subadditive (i.e., $\tilde{t}(x+y) \leq \tilde{t}(x) + \tilde{t}(y)$ for any $x, y \in \mathbb{R}^d$), then from the positive homogeneity of \tilde{t} , we can see that \tilde{t} is convex. Now we verify the subadditivity of \tilde{t} . Note that for each k , $\tilde{s}(\cdot, F_{n_k})$ is subadditive. Then we have $0 \leq \tilde{s}(x, F_{n_k}) + \tilde{s}(y, F_{n_k}) - \tilde{s}(x+y, F_{n_k})$. This yields $0 \leq \tilde{t}(x) + \tilde{t}(y) - \tilde{t}(x+y)$, which implies the subadditivity of \tilde{t} , and we obtain (b).

(Step 3) Now we verify that \tilde{t} is a support function of a bounded closed convex set F . Define $F = \bigcap_{p \in \mathbb{S}^{d-1}} \{x \in \mathbb{R}^d : \langle p, x \rangle \leq \tilde{t}(p)\}$. Then we can see that F is a closed convex set and $\tilde{t}(\cdot)$ is a support function of F . Moreover, for any $x \in F$ with $x \neq 0$, $\|x\| = \langle x/\|x\|, x \rangle \leq t(x/\|x\|) \leq \sup_n (\sup_{f \in F_n} \|f\|) \leq B$. Then F is a bounded closed convex set. In particular $\tilde{t}(\cdot) \equiv s(\cdot, F)$, $F \in K_{kc}^B(\mathbb{R}^d)$.

A.2. Proof of Proposition 2.1.

A.2.1. *Proof of (i).* Note that $\Psi(Y) \in \Psi(\mathcal{M})$, which is a closed convex set (Assumption 2.1 (ii)). Then from Lemma B.1, we have $\mathbb{E}[\Psi(Y)] \in \Psi(\mathcal{M})$. Since Ψ is injective (Assumption 2.1 (i)) and $\mathbb{E}[\Psi(Y)] \in \Psi(\mathcal{M})$, $\Psi^{-1}(\mathbb{E}[\Psi(Y)])$ is well defined. From the definition of $\mathbb{E}[\Psi(Y)]$, we have $\mathbb{E}[\langle h, \Psi(Y) \rangle_{\mathcal{H}}] = \langle h, \mathbb{E}[\Psi(Y)] \rangle_{\mathcal{H}}$ for $h \in \mathcal{H}$. This yields

$$\mathbb{E}[\Psi(Y)] = \operatorname{argmin}_{h \in \mathcal{H}} \mathbb{E}[\|h - \Psi(Y)\|_{\mathcal{H}}^2] = \operatorname{argmin}_{h \in \Psi(\mathcal{M})} \mathbb{E}[\|h - \Psi(Y)\|_{\mathcal{H}}^2].$$

Thus, we have $\mathbb{E}[\|\mathbb{E}[\Psi(Y)] - \Psi(Y)\|_{\mathcal{H}}^2] \leq \mathbb{E}[\|\Psi(\nu) - \Psi(Y)\|_{\mathcal{H}}^2]$ for any $\nu \in \mathcal{M}$. By the isometry of Ψ (Assumption 2.1 (i)), we have $\mathbb{E}[d^2(\Psi^{-1}(\mathbb{E}[\Psi(Y)]), Y)] \leq \mathbb{E}[d^2(\nu, Y)]$ for any $\nu \in \mathcal{M}$. Therefore, we obtain $\mathbb{E}_{\oplus}[Y] = \Psi^{-1}(\mathbb{E}[\Psi(Y)])$.

A.2.2. *Proof of (ii).* Applying the same argument in the proof of (i) to the conditional distribution of Y given X , we have $\Psi(\mathbb{E}_{\oplus}[Y|X]) = \mathbb{E}[\Psi(Y)|X]$. Then we have $\mathbb{E}[\Psi(\mathbb{E}_{\oplus}[Y|X])] = \mathbb{E}[\Psi(Y)] = \Psi(\mathbb{E}_{\oplus}[Y])$, where the last equality follows from (i).

A.3. Proof of Proposition 2.2.

A.3.1. *Proof of (i) and (ii).* Note that $\Psi(K_{kc}^B(\mathbb{R}^d)) \subset L^2(\mathbb{S}^{d-1})$ is closed and convex (Lemma 2.1 (iv)). Then from Proposition 2.1 (i), we have $s(\cdot, \mathbb{E}_\oplus[F]) = \Psi(\mathbb{E}_\oplus[F]) = \mathbb{E}[\Psi(F)] = \mathbb{E}[s(\cdot, F)]$. Since F an integrably bounded SVRV, Lemma B.11 yields $s(\cdot, \mathbb{E}_\oplus[F]) \equiv s(\cdot, \mathbb{E}[F])$, where $\mathbb{E}[F]$ is the Aumann mean of F . Then Lemma B.8 yields $\mathbb{E}_\oplus[F] = \mathbb{E}[F]$.

A.3.2. *Proof of (iii).* The first statement follows by applying a similar argument to the proof of (i). Note that $\mathbb{E}[F|X] \in K_{kc}^B(\mathbb{R}^d)$ and $\mathbb{E}[F] = \mathbb{E}[\mathbb{E}[F|X]]$. Thus, from (i) and Proposition 2.1 (ii), we have

$$\Psi(\mathbb{E}_\oplus[\mathbb{E}_\oplus[F|X]]) = \mathbb{E}[\Psi(\mathbb{E}[F|X])] = s(\cdot, \mathbb{E}[\mathbb{E}[F|X]]) = s(\cdot, \mathbb{E}[F]) = \Psi(\mathbb{E}_\oplus[F]),$$

which yields $\mathbb{E}_\oplus[F] = \mathbb{E}_\oplus[\mathbb{E}_\oplus[F|X]]$.

A.4. **Proof of Corollary 2.1.** From Proposition 2.1, we have $s(\cdot, \mu_{\oplus, n}) = \Psi(\mu_{\oplus, n}) = \frac{1}{n} \sum_{i=1}^n \Psi(F_i) = \frac{1}{n} \sum_{i=1}^n s(\cdot, F_i)$. Thus, Lemma B.6 (iii) yields $s(\cdot, \mu_{\oplus, n}) = s(\cdot, \frac{1}{n} \bigoplus_{i=1}^n F_i)$, and Lemma B.8 implies $\mu_{\oplus, n} = \frac{1}{n} \bigoplus_{i=1}^n F_i$.

A.5. **Proof of Proposition 3.1.** First, we verify that $m_\oplus(x)$ is the GFR function. Note that $\tilde{m}_{\oplus, \Psi}(x)$ exists and is unique from Lemmas 2.1 and B.2. From the definition of $\tilde{m}_{\oplus, \Psi}(x)$, we have

$$\begin{aligned} \mathbb{E}[w(x, X)d_{kc}^2(\nu, F)] &= \|\Psi(\nu) - \mathbb{E}[w(x, X)\Psi(F)]\|_{2, \mathbb{S}^{d-1}}^2 \\ &\quad + \int_{\mathbb{S}^{d-1}} \mathbb{E}[w(x, X)s^2(p, F)]dp - \|\mathbb{E}[w(x, X)\Psi(F)]\|_{2, \mathbb{S}^{d-1}}^2 \\ &\geq \|\tilde{m}_{\oplus, \Psi}(x) - \mathbb{E}[w(x, X)\Psi(F)]\|_{2, \mathbb{S}^{d-1}}^2 \\ &\quad + \int_{\mathbb{S}^{d-1}} \mathbb{E}[w(x, X)s^2(p, F)]dp - \|\mathbb{E}[w(x, X)\Psi(F)]\|_{2, \mathbb{S}^{d-1}}^2 \\ &= \mathbb{E}[w(x, X)d_{kc}^2(\Psi^{-1}(\tilde{m}_{\oplus, \Psi}(x)), F)], \end{aligned}$$

for any $\nu \in K_{kc}^B(\mathbb{R}^d)$. This yields $\Psi^{-1}(\tilde{m}_{\oplus, \Psi}(x)) \in \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \mathbb{E}[w(x, X)d_{kc}^2(\nu, F)]$ so that $\Psi^{-1}(\tilde{m}_{\oplus, \Psi}(x))$ is the unique GFR function.

Next, we verify $m_\oplus(x) \subset \mathbb{E}[w(x, X)F]$. Note that $w(x, X)F \in K_{kc}^{B'}(\mathbb{R}^d)$ where $B' = B \sup_{z \in \mathcal{X}} |w(x, z)|$. Then Lemma 2.1 (iv), Proposition 2.1 (i), and Proposition 2.2 (ii) yield $\mathbb{E}[w(x, X)F] = \mathbb{E}_\oplus[w(x, X)F] \in K_{kc}^{B'}(\mathbb{R}^d) \subset K_{kc}(\mathbb{R}^d)$. Since $m_{\oplus, \Psi}(x) \in \Psi(K_{kc}(\mathbb{R}^d))$, $m_{\oplus, \Psi}(x) = \tilde{m}_{\oplus, \Psi}(x)$ is the support function of $m_\oplus(x) = \Psi^{-1}(m_{\oplus, \Psi}(x)) \in K_{kc}(\mathbb{R}^d)$ and hence (3) yields $m_\oplus(x) = s(\cdot, m_\oplus(x)) \leq s(\cdot, \mathbb{E}[w(x, X)F])$. Therefore, from Lemma B.7, we have $m_\oplus(x) \subset \mathbb{E}[w(x, X)F]$.

A.6. **An example such that $\tilde{m}_{\oplus, \Psi}(x) < \Psi(\mathbb{E}[w(x, X)F])$.** Let $(F, X) \in \{[-1, 2], [1, 6], \{0\}\} \times \{-2, 0, 2\}$ and let

$$\begin{aligned} \mathbb{P}(F = [-1, 2], X = -2) &= \mathbb{P}(F = [1, 6], X = 2) = 1/4, \\ \mathbb{P}(F \neq [-1, 2], X = -2) &= \mathbb{P}(F \neq [1, 6], X = 2) = 0, \\ \mathbb{P}(F = \{0\}, X = 0) &= 1/2, \quad \mathbb{P}(F \neq \{0\}, X = 0) = 0. \end{aligned}$$

Note that $\mathbb{S}^0 = \{-1, 1\}$. Define $g_1(p) = \mathbb{E}[w(2, X)s(p, F)]$ and $g_2(p) = s(p, \mathbb{E}[w(2, X)F])$. Then we have $g_1(p) < g_2(p)$ for $p \in \mathbb{S}^0$. Indeed,

$$g_1(p) = -\frac{1}{4}s(p, [-1, 2]) + \frac{3}{4}s(p, [1, 6]) = \begin{cases} 4 & \text{if } p = 1 \\ -1 & \text{if } p = -1 \end{cases},$$

$$g_2(p) = s\left(p, -\frac{1}{4}[-1, 2] \oplus \frac{3}{4}[1, 6]\right) = s\left(p, \left[\frac{1}{4}, \frac{19}{4}\right]\right) = \begin{cases} \frac{19}{4} & \text{if } p = 1 \\ -\frac{1}{4} & \text{if } p = -1 \end{cases}.$$

Note that $\|g\|_{2, \mathbb{S}^0}^2 = g^2(-1) + g^2(1)$. Then we have $0 = \|g_1 - s(\cdot, [1, 4])\|_{2, \mathbb{S}^0}^2 < \|g_1 - g_2\|_{2, \mathbb{S}^0}^2 = 18/16$. This implies $s(\cdot, [1, 4]) = \tilde{m}_{\oplus, \Psi}(2) \neq \Psi(\mathbb{E}[w(2, X)F]) = s(\cdot, [1/4, 19/4])$. Likewise, we have $\tilde{m}_{\oplus, \Psi}(0) = \Psi(\mathbb{E}[w(0, X)F]) = s(\cdot, \mathbb{E}[F]) = s(\cdot, [0, 2])$ and $s(\cdot, [-1, 0]) = \tilde{m}_{\oplus, \Psi}(-2) \neq \Psi(\mathbb{E}[w(-2, X)F]) = s(\cdot, [-9/4, 5/4])$.

A.7. Proof of Proposition 3.2. We first show the result for $d \geq 2$ and then show the result for $d = 1$.

A.7.1. *Proof for the case $d \geq 2$.* Define

$$Q(\nu, x) = \mathbb{E}[w(x, X)d_{kc}^2(\nu, F)], \quad Q_n(\nu, x) = \frac{1}{n} \sum_{i=1}^n \hat{w}(x, X_i)d_{kc}^2(\nu, F_i).$$

Following the proof of Theorem 2 in Petersen and Müller (2019), it suffices to verify the following conditions.

- (a) Almost surely, for all $\|x\| \leq B_0$, the objects $m_{\oplus}(x)$ and $\hat{m}_{\oplus}(x)$ exist and are unique. Additionally, for any $\varepsilon > 0$,

$$\inf_{\|x\| \leq B_0} \inf_{d_{kc}(\nu, m_{\oplus}(x)) > \varepsilon} Q(\nu, x) - Q(m_{\oplus}(x), x) > 0,$$

and there exists $\zeta = \zeta(\varepsilon) > 0$ such that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left(\inf_{\|x\| \leq B_0} \inf_{d_{kc}(\nu, \hat{m}_{\oplus}(x)) > \varepsilon} Q_n(\nu, x) - Q_n(\hat{m}_{\oplus}(x), x) \geq \zeta \right) = 1.$$

- (b) Let $B_\delta(m_{\oplus}(x)) \subset K_{kc}^B(\mathbb{R}^d)$ be the ball of radius δ centered at $m_{\oplus}(x)$ and $N(\varepsilon, B_\delta(m_{\oplus}(x)), d_{kc})$ be its covering number using ball of size ε . As $\delta \rightarrow 0$, it holds

$$\int_0^1 \sup_{\|x\| \leq B_0} \sqrt{1 + \log N(\delta\varepsilon, B_\delta(m_{\oplus}(x)), d_{kc})} d\varepsilon = O(\delta^{-\frac{d-1}{4}}).$$

- (c) There exist $\tau > 0$ and $D > 0$, possibly depending on B_0 , such that

$$\inf_{\|x\| \leq B_0} \inf_{d_{kc}(\nu, m_{\oplus}(x)) < \tau} \{Q(\nu, x) - Q(m_{\oplus}(x), x) - Dd_{kc}^2(\nu, m_{\oplus}(x))\} \geq 0.$$

Verification of (a). Applying the same argument in the proof of Proposition 3.1, we can see that $m_{\oplus}(x)$ and $\hat{m}_{\oplus}(x)$ exist and are unique for all $\|x\| \leq B_0$, hence proving Condition (a).

Verification of (b). For the remainder of this proof, for any $F \in K_{kc}^B(\mathbb{R}^d)$ and $\gamma > 0$, $B_\gamma(F)$ refers to the d_{kc} ball of radius γ centered at F . Lemma 2.7.12 of van der Vaart and Wellner (2023) implies $N(\varepsilon, K_{kc}^B(\mathbb{R}^d), d_{kc}) \leq e^{D\varepsilon^{-(d-1)/2}}$, where D is independent of ε . Then for $F \in K_{kc}^B(\mathbb{R}^d)$ and $\delta \in (0, 1)$, we can take a collection of compact convex sets in $K_{kc}(\mathbb{R}^d)$, $C_{\delta\varepsilon}(F) = \{g_u : u \in U\}$ with $U \subset \mathbb{R}$ such that $|U| = N(\delta\varepsilon, B_1(F), d_{kc}) \leq e^{D(\delta\varepsilon)^{-(d-1)/2}}$. Then we have shown that $\sup_{F \in K_{kc}^B(\mathbb{R}^d)} \log N(\delta\varepsilon, B_\delta(F), d_{kc}) \leq D(\delta\varepsilon)^{-(d-1)/2}$. Observe that $\sup_{\|x\| \leq B_0} \log N(\delta\varepsilon, B_\delta(m_\oplus(x)), d_{kc}) \leq D(\delta\varepsilon)^{-(d-1)/2}$, so for any $\delta \in (0, 1)$, we have

$$\int_0^1 \sup_{\|x\| \leq B_0} \sqrt{1 + \log N(\delta\varepsilon, B_\delta(m_\oplus(x)), d_{kc})} d\varepsilon \leq \int_0^1 \sqrt{1 + D(\delta\varepsilon)^{-(d-1)/2}} d\varepsilon = O(\delta^{-\frac{d-1}{4}}).$$

Verification of (c). For $f, g \in L^2(\mathbb{S}^{d-1})$, let $\langle f, g \rangle_{2, \mathbb{S}^{d-1}} = \int_{\mathbb{S}^{d-1}} f(x)g(x)dx$. From Lemma B.3, $m_\oplus(x)$ is characterized by $\langle m_\oplus, \Psi(x) - \tilde{m}_\oplus, \Psi(x) - \tilde{m}_\oplus \rangle_{2, \mathbb{S}^{d-1}} \leq 0$ for all $h \in \Psi(K_{kc}^B(\mathbb{R}^d))$ and $\|x\| \leq B_0$. Then we have

$$\begin{aligned} Q(\nu, x) - Q(m_\oplus(x), x) &= d_{kc}^2(\nu, m_\oplus(x)) - 2\langle m_\oplus, \Psi(x) - \tilde{m}_\oplus, \Psi(\nu) - \tilde{m}_\oplus \rangle_{2, \mathbb{S}^{d-1}} \\ &\geq d_{kc}^2(\nu, m_\oplus(x)). \end{aligned}$$

Consequently, we can take $D = 1$ and τ arbitrary to satisfy Condition (c).

A.7.2. Proof for the case $d = 1$. Verification of (a) and (c) is the same as in $d \geq 2$. One can see that $N(\varepsilon, K_{kc}^B(\mathbb{R}), d_{kc}) \leq D\varepsilon^{-2}$ where D is independent of ε . By applying a similar argument to verify (b) when $d \geq 2$, we have

$$\int_0^1 \sup_{\|x\| \leq B_0} \sqrt{1 + \log N(\delta\varepsilon, B_\delta(m_\oplus(x)), d_{kc})} d\varepsilon = O(-\log \delta).$$

Following the proof of Theorem 2 in Petersen and Müller (2019), this yields $\sup_{\|x\| \leq B_0} d_{kc}(\hat{m}_\oplus(x), m_\oplus(x)) = O_p(n^{-\frac{1}{2(\alpha-1)}})$ for any $\alpha > 2$.

A.8. Derivation of (4). For $\alpha, \beta \in K_{kc}^B(\mathbb{R}^d)$, let $\gamma_{\alpha, \beta} : [0, 1] \rightarrow K_{kc}^B(\mathbb{R}^d)$ denote the geodesic from α to β under the metric d_{kc} , and $\gamma_{\alpha, \beta}(\rho)$ be the end point of the (extended) geodesic $\rho \odot \gamma_{\alpha, \beta}$. See Kurisu et al. (2024) for a more detailed discussion on geodesic metric spaces and extensions of geodesics. Since $F \in K_{kc}^B(\mathbb{R}^d)$ can be isometrically embedded into $L^2(\mathbb{S}^{d-1})$ by using the map $\Psi(F) = s(\cdot, F)$, we have

$$\gamma_{\{0\}, F_i} \left(\frac{T_i}{\hat{e}(X_i)} \right) = \Psi^{-1} \left(\Psi(\{0\}) + \frac{T_i}{\hat{e}(X_i)} (\Psi(F_i) - \Psi(\{0\})) \right) = \Psi^{-1} \left(\frac{T_i}{\hat{e}(X_i)} \Psi(F_i) \right) = \frac{T_i}{\hat{e}(X_i)} F_i.$$

Therefore, we have

$$\hat{\Theta}_1^{(\text{IPW})} = \operatorname{argmin}_{\nu \in K_{kc}^B(\mathbb{R}^d)} \frac{1}{n} \sum_{i=1}^n d_{kc}^2 \left(\nu, \gamma_{\{0\}, F_i} \left(\frac{T_i}{\hat{e}(X_i)} \right) \right) = \frac{1}{n} \bigoplus_{i=1}^n \frac{T_i}{\hat{e}(X_i)} F_i.$$

A.9. **Proof of Proposition 3.3.** Observe that

$$\begin{aligned}
& d_{kc}(\hat{\mathbb{E}}_{\oplus}[F], \mathbb{E}_{\oplus}[F]) \\
&= \left\| s \left(\cdot, \left(\frac{1}{n} \sum_{i=1}^n \frac{T_i}{\hat{e}(X_i)} \right)^{-1} \frac{1}{n} \bigoplus_{i=1}^n \frac{T_i}{\hat{e}(X_i)} F_i \right) - s(\cdot, \mathbb{E}_{\oplus}[F]) \right\|_{2, \mathbb{S}^{d-1}} \\
&\leq \frac{B}{\eta_0} \sqrt{\text{vol}(\mathbb{S}^{d-1})} \left| \left(\frac{1}{n} \sum_{i=1}^n \frac{T_i}{\hat{e}(X_i)} \right)^{-1} - 1 \right| + \left\| s \left(\cdot, \frac{1}{n} \bigoplus_{i=1}^n \frac{T_i}{\hat{e}(X_i)} F_i \right) - s(\cdot, \mathbb{E}_{\oplus}[F]) \right\|_{2, \mathbb{S}^{d-1}} \\
&\leq \frac{B}{\eta_0} \sqrt{\text{vol}(\mathbb{S}^{d-1})} \left\{ \frac{1}{\eta_0^2} \sup_{x \in \mathcal{X}} |\hat{e}(x) - e(x)| + \left| \frac{1}{n} \sum_{i=1}^n \frac{T_i}{e(X_i)} - 1 \right| \right\} (1 + o_p(1)) + d_{kc}(\hat{\Theta}_1^{(\text{IPW})}, \mathbb{E}_{\oplus}[F]) \\
&= O_p(\rho_n) + O_p(n^{-1/2}) + d_{kc}(\hat{\Theta}_1^{(\text{IPW})}, \mathbb{E}_{\oplus}[F]).
\end{aligned}$$

Therefore, the conclusion follows from a similar argument in the proof of Theorem A.4 in Kurisu et al. (2024).

APPENDIX B. AUXILIARY RESULTS

We first present some results on Hilbert space.

Lemma B.1. *If C is a nonempty closed convex subset of a Hilbert space \mathcal{H} and Z is a random element taking values in C with $\mathbb{E}[\|Z\|_{\mathcal{H}}] < \infty$, then $\mathbb{E}[Z] \in C$.*

Proof. Let $\mu_Z = \mathbb{E}[Z]$ be the expectation of Z and suppose $\mu_Z \notin C$. Since C is closed, there is an open ball $B_{\delta}(\mu_Z) = \{h \in \mathcal{H} : \|h - \mu_Z\|_{\mathcal{H}} < \delta\}$ such that $B_{\delta}(\mu_Z) \cap C = \emptyset$. By Hahn-Banach separation theorem, there exists an $h_0 \in \mathcal{H}$ and $a \in \mathbb{R}$ such that $\langle h_0, \mu_Z \rangle_{\mathcal{H}} < a \leq \langle h_0, h \rangle_{\mathcal{H}}$ for all $h \in C$. Since $a \leq \langle h_0, h \rangle_{\mathcal{H}}$ for all $h \in C$, we have $a \leq \mathbb{E}[\langle h_0, Z \rangle_{\mathcal{H}}]$. This implies $a \leq \langle h_0, \mu_Z \rangle_{\mathcal{H}}$ and contradicts with $\langle h_0, \mu_Z \rangle_{\mathcal{H}} < a$. \square

Lemma B.2. *Let C be a nonempty closed convex set of a Hilbert space \mathcal{H} . For any $y \in \mathcal{H}$, the projection $\pi_C(y) := \text{argmin}_{h \in C} \|y - h\|_{\mathcal{H}}$ exists and is unique.*

Proof. First, we show the existence of a minimizer. Define $r = \inf_{h \in C} \|y - h\|_{\mathcal{H}}$. Since C is nonempty, r is finite. Indeed, for any $h \in C$, it holds $r \leq \|y - h\|_{\mathcal{H}} < \infty$. Choose a minimizing sequence $\{h_n\}_{n \geq 1} \subset C$ such that $\|y - h_n\|_{\mathcal{H}} \rightarrow r$. This sequence is bounded. Indeed, for any $h_0 \in C$, we have $\|h_n\|_{\mathcal{H}} \leq \|h_0\|_{\mathcal{H}} + \|y - h_n\|_{\mathcal{H}} \leq \|y\|_{\mathcal{H}} + (r + 1)$ for sufficiently large n . Hence $\{h_n\}_{n \geq 1}$ has a convergent subsequence $h_{n_k} \rightarrow h^* \in \mathcal{H}$. Since C is closed, it holds $h^* \in C$. Then we have $\|y - h^*\|_{\mathcal{H}} = \lim_{k \rightarrow \infty} \|y - h_{n_k}\|_{\mathcal{H}} = r$ and h^* attains the minimum. Thus, a minimizer exists.

Next, we show the uniqueness. Consider $f(h) = \|y - h\|_{\mathcal{H}}^2$, which is strictly convex on \mathcal{H} . Indeed, for any $h_1, h_2 \in \mathcal{H}$ and $t \in (0, 1)$, it holds $f(th_1 + (1-t)h_2) \leq t\|y - h_1\|_{\mathcal{H}}^2 + (1-t)\|y - h_2\|_{\mathcal{H}}^2$ and the equality holds if and only if $h_1 = h_2$. Indeed,

$$t\|y - h_1\|_{\mathcal{H}}^2 + (1-t)\|y - h_2\|_{\mathcal{H}}^2 - \|t(y - h_1) + (1-t)(y - h_2)\|_{\mathcal{H}}^2 = t(1-t)\|h_1 - h_2\|_{\mathcal{H}}^2.$$

Now suppose $h_1^*, h_2^* \in C$ are distinct minimizers. Then $m = (h_1^* + h_2^*)/2 \in C$ by the convexity of C , but strict convexity of f yields $\|y - m\|_{\mathcal{H}}^2 < \frac{1}{2}\|y - h_1^*\|_{\mathcal{H}}^2 + \frac{1}{2}\|y - h_2^*\|_{\mathcal{H}}^2 = r^2$, contradicting the minimality of r . Hence the minimizer is unique. \square

Lemma B.3. *For a Hilbert space \mathcal{H} , let $C \subset \mathcal{H}$ be a nonempty closed convex set and $y \in \mathcal{H}$. Then for every $h \in C$, $\langle y - \pi_C(y), h - \pi_C(y) \rangle_{\mathcal{H}} \leq 0$.*

Proof. Now we show Lemma B.3. Note that $\pi_C(y) \in C$. Fix $h_0 \in C$ and for $t \in (0, 1)$, define $\gamma(t) = (1 - t)\pi_C(y) + th_0 \in C$. Then we have

$$\|y - \pi_C(y)\|_{\mathcal{H}}^2 \leq \|y - \gamma(t)\|_{\mathcal{H}}^2 = \|y - \pi_C(y)\|_{\mathcal{H}}^2 - 2t\langle y - \pi_C(y), h_0 - \pi_C(y) \rangle_{\mathcal{H}} + t^2\|h_0 - \pi_C(y)\|_{\mathcal{H}}^2.$$

This yields $0 \leq -2t\langle y - \pi_C(y), h_0 - \pi_C(y) \rangle_{\mathcal{H}} + t^2\|h_0 - \pi_C(y)\|_{\mathcal{H}}^2$. Recall $t > 0$. Therefore, we have $0 \leq -2\langle y - \pi_C(y), h_0 - \pi_C(y) \rangle_{\mathcal{H}} + t\|h_0 - \pi_C(y)\|_{\mathcal{H}}^2$. Letting $t \downarrow 0$ yields $\langle y - \pi_C(y), h_0 - \pi_C(y) \rangle_{\mathcal{H}} \leq 0$. \square

Hereafter, we summarize some results on compact convex sets and their support functions used in the main text.

Lemma B.4. *If $F \subset \mathbb{R}^d$ is a bounded set (i.e., $\sup_{f \in F} \|f\| < \infty$), then the support function $s(p, F)$, $p \in \mathbb{S}^{d-1}$ is Lipschitz continuous and uniformly continuous on \mathbb{S}^{d-1} .*

Proof. For any $u, v \in \mathbb{R}^d$ and any $f \in F$, we have $\langle u, f \rangle - \langle v, f \rangle \leq \langle u - v, f \rangle \leq (\sup_{f \in F} \|f\|)\|u - v\|$. Then we have $s(u, F) - s(v, F) \leq (\sup_{f \in F} \|f\|)\|u - v\|$. Likewise, we have $s(v, F) - s(u, F) \leq (\sup_{f \in F} \|f\|)\|u - v\|$, which implies $|s(u, F) - s(v, F)| \leq (\sup_{f \in F} \|f\|)\|u - v\|$, and the conclusion follows. \square

Lemma B.5. *Let $s(\cdot, F)$ be the support function of $F \in K_{kc}(\mathbb{R}^d)$. Then we have $F = \bigcap_{p \in \mathbb{S}^{d-1}} \{x \in \mathbb{R}^d : \langle p, x \rangle \leq s(p, F)\} =: F_S$.*

Proof. First, we show that if $f \in F$, then $f \in F_S$. For any $f \in F$ and $p \in \mathbb{S}^{d-1}$, we have $\langle p, f \rangle \leq \sup_{f \in F} \langle p, f \rangle = s(p, F)$, which implies $f \in F_S$ and hence $F \subset F_S$. Next, we show that if $f \in F_S$, then $f \in F$. Suppose $f_0 \notin F$. Then by Hahn-Banach separation theorem, there exist $u \in \mathbb{R}^d$ and $a \in \mathbb{R}$ such that $\langle u, f \rangle \leq a < \langle u, f_0 \rangle$ for any $f \in F$. Thus, we have $s(u, F) < \langle u, f_0 \rangle$ and hence $f_0 \notin F_S$. Therefore, we obtain $F_S \subset F$, and the conclusion follows. \square

Lemma B.6. *The following results hold true.*

(i) *(Positive homogeneity) For any $p \in \mathbb{R}^d$, $t \geq 0$, and subset $F \subset \mathbb{R}^d$,*

$$s(tp, F) = ts(p, F), \quad s(p, tF) = ts(p, F).$$

(ii) *(Additivity) For any $p \in \mathbb{R}^d$ and subsets $F, G \subset \mathbb{R}^d$,*

$$s(p, F \oplus G) = s(p, F) + s(p, G).$$

(iii) (Convexity) For any $p \in \mathbb{R}^d$, $t \in [0, 1]$, and subsets $F, G \subset \mathbb{R}^d$,

$$s(p, (tF) \oplus ((1-t)G)) = ts(p, F) + (1-t)s(p, G).$$

Proof. (i) We only show the first result since the second result can be shown in the same manner. Observe that for any $f \in F$, $\langle tp, f \rangle = t\langle p, f \rangle$. Then we have $s(pt, F) = ts(p, F)$.

(ii) We first show $s(p, F \oplus G) \leq s(p, F) + s(p, G)$. Observe that for any $f \in F$ and $g \in G$,

$$\langle p, f + g \rangle = \langle p, f \rangle + \langle p, g \rangle = \sup_{f \in F} \langle p, f \rangle + \sup_{g \in G} \langle p, g \rangle = s(p, F) + s(p, G).$$

Then we have $s(p, F \oplus G) = \sup_{h \in F \oplus G} \langle p, h \rangle \leq s(p, F) + s(p, G)$. Next we show $s(p, F) + s(p, G) \leq s(p, F \oplus G)$. If $s(p, F) = -\infty$ or $s(p, G) = -\infty$, then the inequality follows immediately. Now we assume that $s(p, F) > -\infty$ and $s(p, G) > -\infty$. Fix any positive ε . From the definition of the supremum, there exist $f_\varepsilon \in F$ and $g_\varepsilon \in G$ such that $\langle p, f_\varepsilon \rangle \geq s(p, F) - \varepsilon$ and $\langle p, g_\varepsilon \rangle \geq s(p, G) - \varepsilon$. Hence we have $\langle p, f_\varepsilon + g_\varepsilon \rangle \geq s(p, F) + s(p, G) - 2\varepsilon$. This yields $s(p, F \oplus G) = \sup_{h \in F \oplus G} \langle p, h \rangle \geq s(p, F) + s(p, G) - 2\varepsilon$. Letting $\varepsilon \downarrow 0$, we obtain $s(p, F) + s(p, G) \leq s(p, F \oplus G)$.

(iii) The result follows from (i) and (ii). \square

Lemma B.7. For $F, G \in K_{kc}(\mathbb{R}^d)$, $s(p, F) \leq s(p, G)$ for all $p \in \mathbb{S}^{d-1}$ if and only if $F \subset G$.

Proof. From the definition of the support function, the verification of $F \subset G \Rightarrow s(p, F) \leq s(p, G)$ for all $p \in \mathbb{S}^{d-1}$ is straightforward. Now we verify $s(p, F) \leq s(p, G)$ for all $p \in \mathbb{S}^{d-1} \Rightarrow F \subset G$. From Lemma B.5, we have $K = \bigcap_{p \in \mathbb{S}^{d-1}} \{x \in \mathbb{R}^d : \langle p, x \rangle \leq s(p, K)\}$ for $K \in \{F, G\}$. Since $s(p, F) \leq s(p, G)$ for all $p \in \mathbb{S}^{d-1}$, we obtain $\{x \in \mathbb{R}^d : \langle p, x \rangle \leq s(p, F)\} \subset \{x \in \mathbb{R}^d : \langle p, x \rangle \leq s(p, G)\}$ for all $p \in \mathbb{S}^{d-1}$. This yields $F \subset G$. \square

Lemma B.8. Let $s(\cdot, F)$ and $s(\cdot, G)$ be the support functions of $F, G \in K_{kc}(\mathbb{R}^d)$, respectively. Then $F = G$ if and only if $s(p, F) = s(p, G)$ for $p \in \mathbb{S}^{d-1}$.

Proof. We only verify $s(p, F) = s(p, G)$ for all $p \in \mathbb{S}^{d-1} \Rightarrow F = G$ since the converse is obvious. Note that $s(p, F) = s(p, G)$ for all $p \in \mathbb{S}^{d-1} \Leftrightarrow s(p, F) \leq s(p, G)$ and $s(p, F) \geq s(p, G)$ for all $p \in \mathbb{S}^{d-1}$. Then Lemma B.7 yields $F \subset G$ and $F \supset G$, which implies $F = G$. \square

Lemma B.9. For any $F, G \in K_{kc}(\mathbb{R}^d)$, we have $d_H(F, G) = \sup_{p \in \mathbb{S}^{d-1}} |s(p, F) - s(p, G)|$.

Proof. We first show $\sup_{p \in \mathbb{S}^{d-1}} |s(p, F) - s(p, G)| \leq d_H(F, G)$. Let $B = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$. Fix any positive ε and any $f \in F$. From the definition of d_H , there exists $g_{f,\varepsilon} \in G$ such that $\|f - g_{f,\varepsilon}\| \leq d_H(F, G) + \varepsilon$. Hence for any $p \in \mathbb{S}^{d-1}$, we have

$$\begin{aligned} \langle p, f \rangle &= \langle p, g_{f,\varepsilon} \rangle + \langle p, f - g_{f,\varepsilon} \rangle \leq \langle p, g_{f,\varepsilon} \rangle + \|p\| \|f - g_{f,\varepsilon}\| \\ &\leq \langle p, g_{f,\varepsilon} \rangle + d_H(F, G) + \varepsilon \leq s(p, G) + d_H(F, G) + \varepsilon. \end{aligned}$$

This yields $s(p, F) \leq s(p, G) + d_H(F, G) + \varepsilon$. Letting $\varepsilon \downarrow 0$, we have $s(p, F) \leq s(p, G) + d_H(F, G)$. Likewise, we have $s(p, G) \leq s(p, F) + d_H(F, G)$. Then we have $|s(p, F) - s(p, G)| \leq d_H(F, G)$, which yields $\sup_{p \in \mathbb{S}^{d-1}} |s(p, F) - s(p, G)| \leq d_H(F, G)$.

Next we show $d_H(F, G) \leq \sup_{p \in \mathbb{S}^{d-1}} |s(p, F) - s(p, G)|$. Let $\delta = \sup_{p \in \mathbb{S}^{d-1}} |s(p, F) - s(p, G)|$. Then for any $p \in \mathbb{S}^{d-1}$ and any $f \in F$, we have $\langle p, f \rangle \leq s(p, F) \leq s(p, G) + \delta$. This yields

$$\begin{aligned} F &\subset \bigcap_{p \in \mathbb{S}^{d-1}} \left\{ x \in \mathbb{R}^d : \langle p, x \rangle \leq s(p, G) + \delta \right\} \\ &= \bigcap_{p \in \mathbb{S}^{d-1}} \left\{ x \in \mathbb{R}^d : \langle p, x \rangle \leq s(p, G \oplus (\delta B)) \right\} \quad (\text{Lemma B.6 (i), (ii)}) \\ &= G \oplus (\delta B) \quad (\text{Lemma B.5}) \end{aligned}$$

and hence for any $f \in F$, it holds $\inf_{g \in G} \|f - g\| \leq \delta$. Likewise, we have $G \subset F \oplus (\delta B)$ and thus $\inf_{f \in F} \|g - f\| \leq \delta$ for any $g \in G$. Therefore, we have $d_H(F, G) \leq \delta = \sup_{p \in \mathbb{S}^{d-1}} |s(p, F) - s(p, G)|$, and the conclusion follows. \square

Lemma B.10 (Theorem 3 in Vitale (1985)). *Let $\{F_n\}_{n \geq 1} \subset K_{kc}(\mathbb{R}^d)$ be a sequence of compact convex sets. If $\lim_{n \rightarrow \infty} d_H(F_n, F) = 0$, then $F \subset \mathbb{R}^d$ is a compact convex set.*

Lemmas 2.1 (ii) and B.10 imply that $(K_{kc}(\mathbb{R}^d), d_{kc})$ (or $(K_{kc}(\mathbb{R}^d), d_H)$) is a complete metric space.

Lemma B.11 (Lemma 2.2 in Artstein (1974)). *If an SVRV $F : \Omega \rightarrow K_{kc}(\mathbb{R}^d)$ is integrably bounded, then we have $\mathbb{E}[s(p, F)] = s(p, \mathbb{E}[F])$ for $p \in \mathbb{S}^{d-1}$.*

REFERENCES

- Artstein, Z. (1974). On the calculus of closed set-valued functions. *Indiana University Mathematics Journal*, 24(5):433–441.
- Aumann, R. J. (1965). Integrals of set valued functions. *Journal of Mathematical Analysis and Applications*, 12:1–12.
- Beresteanu, A. and Molinari, F. (2008). Asymptotic properties for a class of partially identified models. *Econometrica*, 76(4):763–814.
- Bhattacharjee, S., Li, B., Wu, X., and Xue, L. (2025). Doubly robust estimation of causal effects for random object outcomes with continuous treatments. *arXiv preprint arXiv:2506.22754*.
- Bhattacharjee, S. and Müller, H.-G. (2023). Single index Fréchet regression. *Annals of Statistics*, 51(4):1770–1798.
- Chen, Y. and Müller, H.-G. (2022). Uniform convergence of local Fréchet regression with applications to locating extrema and time warping for metric space valued trajectories. *Annals of Statistics*, 50(3):1573–1592.
- Dubey, P., Chen, Y., and Müller, H.-G. (2024). Metric statistics: Exploration and inference for random objects with distance profiles. *Annals of Statistics*, 52(2):757–792.
- Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Annales de l’Institut Henri Poincaré*, 10:215–310.
- Hansen, B. E. (2022). *Econometrics*. Princeton University Press.

- Kurisu, D. and Otsu, T. (2025). Model averaging for global Fréchet regression. *Journal of Multivariate Analysis*, 207:105416.
- Kurisu, D., Zhou, Y., Otsu, T., and Müller, H.-G. (2024). Geodesic causal inference. *arXiv preprint arXiv:2406.19604*.
- Kurisu, D., Zhou, Y., Otsu, T., and Müller, H.-G. (2025a). Geodesic synthetic control methods for random objects and functional data. *arXiv preprint arXiv:2505.00331*.
- Kurisu, D., Zhou, Y., Otsu, T., and Müller, H.-G. (2025b). Regression discontinuity designs for functional data and random objects in geodesic spaces. *arXiv preprint arXiv:2506.18136*.
- Li, Q., Molchanov, I., Molinari, F., and Peng, S. (2021). Local regression smoothers with set-valued outcome data. *International Journal of Approximate Reasoning*, 128:129–150.
- Marron, J. S. and Alonso, A. M. (2014). Overview of object oriented data analysis. *Biometrical Journal*, 56(5):732–753.
- Marron, J. S. and Dryden, I. L. (2021). *Object Oriented Data Analysis*. Chapman and Hall/CRC.
- Molchanov, I. (2017). *Theory of Random Sets*. Springer, 2nd edition.
- Molchanov, I. and Molinari, F. (2018). *Random Sets in Econometrics*, volume 60. Cambridge University Press.
- Molinari, F. (2020). Microeconometrics with partial identification. *Handbook of Econometrics*, 7:355–486.
- Petersen, A. and Müller, H.-G. (2019). Fréchet regression for random objects with euclidean predictors. *Annals of Statistics*, 47(2):691–719.
- Schneider, R. (2014). *Convex Bodies: the Brunn-Minkowski Theory*, volume 151. Cambridge university press, 2nd edition.
- Schoenberg, I. J. (1938). Metric spaces and positive definite functions. *Transactions of the American Mathematical Society*, 44(3):522–536.
- Sejdinovic, D., Sriperumbudur, B., Gretton, A., and Fukumizu, K. (2013). Equivalence of distance-based and RKHS-based statistics in hypothesis testing. *Annals of Statistics*, 41(5):2263–2291.
- Tucker, D. C., Wu, Y., and Müller, H.-G. (2023). Variable selection for global Fréchet regression. *Journal of the American Statistical Association*, 118(542):1023–1037.
- van der Vaart, A. and Wellner, J. (2023). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer Cham, 2nd edition.
- Vitale, R. A. (1985). L^p metrics for compact, convex sets. *Journal of Approximation Theory*, 45(3):280–287.
- Zhou, Y., Kurisu, D., Otsu, T., and Müller, H.-G. (2025). Geodesic difference-in-differences. *arXiv preprint arXiv:2501.17436*.

(D. Kurisu) CENTER FOR SPATIAL INFORMATION SCIENCE, THE UNIVERSITY OF TOKYO, 5-1-5, KASHIWANOHA, KASHIWA-SHI, CHIBA 277-8568, JAPAN.

Email address: daisukekurisu@ccsis.u-tokyo.ac.jp

(Y. Okamoto) GRADUATE SCHOOL OF ECONOMICS, KYOTO UNIVERSITY, YOSHIDA HONMACHI, SAKYO, KYOTO
606-8501, JAPAN.

Email address: `okamoto.yuuta.57w@st.kyoto-u.ac.jp`

(T. Otsu) DEPARTMENT OF ECONOMICS, LONDON SCHOOL OF ECONOMICS, HOUGHTON STREET, LONDON, WC2A
2AE, UK.

Email address: `t.otsu@lse.ac.uk`