Electronic Journal of Statistics
Vol. 19 (2025) 1722–1783
ISSN: 1935-7524
https://doi.org/10.1214/25-EJS2371

Wasserstein-Fréchet integration of conditional distributions

Álvaro Gajardo*[1], Hans-Georg Müller[1] and Hang Zhou[1]

[1] *Department of Statistics, University of California, Davis,*
e-mail: aegajardo@ucdavis.edu*;* hgmueller@ucdavis.edu*;* hgzhou@ucdavis.edu

Abstract: We propose a generalized notion of Wasserstein-Fréchet integral of conditional distributions for the classical situation of a joint distribution between two scalar random variables X and Y by viewing the space of probability distributions as a metric space and defining the Wasserstein-Fréchet integral of conditional distributions as a Fréchet integral. Within this general framework we illustrate various special cases, focusing on the case where one adopts the 2-Wasserstein metric, which however is only one possible choice of metric to implement the proposed method. We demonstrate that this choice often leads to a useful and interpretable notion of the conditional distribution of Y in situations where Y varies systematically with X when one is interested in the residual distribution of Y after the systematic effect is removed. We provide convergence results for the estimated Wasserstein-Fréchet integral of conditional distributions for several commonly encountered data generating mechanisms that are of statistical relevance. These include scatterplot data (X_i, Y_i), $i = 1, \dots, n$; data where one has a sample of fully observed (conditional) densities along with predictors X_i; and data where one encounters conditional densities that are not fully observed and instead one has samples of the data that they generate.

MSC2020 subject classifications: Primary 62G05; secondary 62G20.
Keywords and phrases: conditional distribution, distributional data, Fréchet integral, optimal transport, residual distribution, scatterplot data, Wasserstein metric.

Received May 2022.

1. Introduction and preliminaries

The problem of defining a notion of mean when dealing with random objects that reside in a metric space is of key statistical interest (Bhattacharya and Patrangenaru, 2005; Agueh and Carlier, 2011). For a metric space $\mathcal{M}$ with a probability measure P, the Fréchet mean or barycenter (Fréchet, 1948) on the metric space $(\mathcal{M}, d)$ is defined as $\bar{\nu} = \operatorname{argmin}_{\nu_0 \in \mathcal{M}} E_{\mathsf{P}}(d^2(\nu, \nu_0))$, where ν is a random element in $\mathcal{M}$. Here $\bar{\nu}$ in general is not unique and may consist of more than one element. In the following, expectations and moments of distributions will always be taken with regard to the measure P. Whether the Fréchet mean is unique depends on both the space and the measure P. It extends the notion of center from the usual mean or expectation in Euclidean space to more

*This research was done while Álvaro Gajardo was a PhD student at the University of California, Davis.

general metric spaces and random objects. Extensions to conditional means or conditional barycenters for situations in which the response is a random object lying in a metric space and where the predictor is an Euclidean vector have been studied recently (Petersen and Müller, 2019a). Another useful tool is Fréchet integrals, where the Fréchet mean is taken over the arguments of a random function that takes values in a metric space. These generalized integrals were found to be instrumental for various scenarios (Petersen and Müller, 2016a; Dubey and Müller, 2020; Lin and Müller, 2021).

In the following, we use the abbreviation pdf for probability density function and assume that pairs of random variables (X, Y) have a joint distribution with well-defined joint, conditional and marginal pdfs in the traditional sense, where the joint distribution of (X, Y) has the pdf $f_{X,Y}(x, y)$; the marginal pdf of X and Y are denoted by f_X, f_Y and the conditional pdf of $Y \mid X$ by $f_{Y|X} = f_{X,Y}(x, y)/f_X(x)$. The key idea of our approach is as follows: We start by observing that the marginal density f_Y is the center of the (random) conditional pdfs $f_{Y|X}(\cdot, X)$, viewed as elements of the space of pdfs $\mathcal{F} = \{g : \mathcal{Y} \to \mathbb{R}\colon \int_{\mathcal{Y}} g(s)ds = 1, g \geq 0\}$, since

$$f_Y(y) = \int_{\mathcal{X}} f_{Y|X}(y, x) f_X(x)\, dx = E(f_{Y|X}(y, X)),$$

where f_X denotes the density of X. Accordingly, the classical marginal f_Y is the solution to an optimization problem over $\mathcal{F}$,

$$f_Y = \operatorname*{argmin}_{g \in \mathcal{F}} E[d^2_{L^2}(f_{Y|X}(\cdot, X), g)], \tag{1}$$

where d_{L^2} denotes the L^2 distance in $\mathcal{F}$ with respect to the Lebesgue measure. To see this, observe that for any $g \in \mathcal{F}$ and using properties of the L^2 inner product, similarly to Proposition 1 in Petersen and Müller (2019a),

$$E[d^2_{L^2}(f_{Y|X}(\cdot, X), g)] = E[\|f_{Y|X}(\cdot, X) - f_Y\|^2_{L^2}] + \|f_Y - g\|^2_{L^2},$$

where the minimizer is uniquely achieved at the density $g = f_Y$. We conclude that the classical marginal density f_Y is the (population) Fréchet mean of the densities $f_{Y|X}(\cdot, X)$ when viewed as random elements of the metric space $(\mathcal{F}, d_{L^2})$ of density functions endowed with the L^2 metric.

This observation suggests that using the L^2 metric in (1) is a special case of a more general principle, where the metric d_{L^2} may be replaced by another metric d in the space of distributions. Implementing this idea leads to the proposed Fréchet integral of conditional distributions,

$$f_{Y,d} = \operatorname*{argmin}_{g \in \mathcal{F}} E[d^2(f_{Y|X}(\cdot, X), g)], \tag{2}$$

which depends on the choice of the metric d, with $f_Y = f_{Y,d_{L^2}}$. Of particular interest is the choice of the 2-Wasserstein metric $d = d_{\mathcal{W}}$, due to its ability to measure convergence in law while taking into account the geometry of the intrinsic space. In the following we refer to Fréchet integrals of conditional distributions with respect to the Wasserstein metric as Wasserstein-Fréchet integrals of conditional distributions and denote these by $f_{Y,d_{\mathcal{W}}}$.

Wasserstein-Fréchet integrals of conditional distributions can be expected to capture the underlying geometry of $Y|X = x$ in scenarios of statistical interest such as when conditional distributions form location-scale families or warped/deformed distributions that are derived from a common template measure (Panaretos and Zemel, 2019). Data examples illustrating that Wasserstein-Fréchet integrals of conditional distributions provide a useful alternative to the usual marginal distributions are provided in Section 6.1 below. For example, the right panel of Figure 2 below displays a comparison of the regular marginal density and the density of the Wasserstein-Fréchet integral of the conditional distribution for weekday bike rentals during 2019 at a station in the Chicago Divvy bike system. In contrast to the unimodal Wasserstein-Fréchet integral, the regular marginal density is seen to be bimodal, which turns out not to be a reasonable representation of the random fluctuations around the mean trend. A similar phenomenon can be seen in Figure 1 for a simulated example. Since the classical notion of marginal measure implicitly utilizes minimization of the L^2 distance, in many situations it is not suited to track the underlying geometry of the space of conditional distributions $f_{Y|X}(\cdot, x)$, $x \in \mathcal{X}$, even in simple cases such as when $Y|X = x$ belongs to the Gaussian ensemble with a mean that varies with x. These examples demonstrate that since the space of density functions $\mathcal{F}$ is a nonlinear subset of L^2, adopting the L^2 metric in this space is not always a statistically meaningful choice.

Estimation of a population level Fréchet mean in Wasserstein space proceeds by finding the empirical barycenter of a collection of iid observations of probability measures $\nu_1, \ldots, \nu_k$ (Le Gouic and Loubes, 2017), where the empirical barycenter of $\nu_1, \ldots, \nu_k$ is given by the minimizer of the Fréchet functional $\mu \to \sum_{j=1}^{k} d^2(\nu_j, \mu)$, which exists and is unique under absolute continuity of at least one of the measures μ_j (Agueh and Carlier, 2011). In this framework, Wasserstein-Fréchet integrals of conditional distributions are of interest when the probability measures $\nu = \nu(X)$ are related to a predictor X, where measures $\nu(X)$ have a (random) density $f_{Y|X}(\cdot, X)$; there is a connection with the population barycenter for parametric families studied in Bigot and Klein (2018), where the probability distribution family is defined through a transformation $\Phi(\theta)$ that is known in advance and inherently tied to the conditional distribution $Y|X = x$. However, in practical situations such as scatterplot data (X_i, Y_i), $i = 1, \ldots, n$, the available information does not correspond to a sample of probability measures but instead to a sample of observations coming from the joint distribution of (X, Y), and the fact that population level random measures are not directly observed needs to be taken into account when constructing suitable empirical estimates; the empirical barycenter estimation approach in Bigot and Klein (2018) is infeasible due to the unavailability of the sample of distributions.

In the following, we consider the space of probability measures on $\mathcal{Y} \subseteq \mathbb{R}$, denoted by $\mathcal{P}(\mathcal{Y})$. For the subspace $\mathcal{P}_p(\mathcal{Y}) \subset \mathcal{P}(\mathcal{Y})$ of probability measures on $\mathcal{Y}$ with finite p-moments, $p \geq 1$, a metric that has attracted much interest in the literature is the p-Wasserstein metric, which for two probability measures $\nu_1, \nu_2 \in \mathcal{P}_p(\mathcal{Y})$ is defined by $d_{\mathcal{W}_p}(\nu_1, \nu_2) = \inf E(|R_1 - R_2|^p)^{1/p}$, where the infimum is taken with respect to both scalar random variables R_1 and R_2 such that $R_1 \sim \nu_1$ and $R_2 \sim \nu_2$ (Villani, 2003). Of particular interest is the choice of the 2-Wasserstein metric $d = d_{\mathcal{W}}$, due to its ability to measure convergence in law while taking into account the geometry of the intrinsic space. While we study only the case of one-dimensional distributions here, for the case of multivariate distributions, it is noteworthy that the solutions to the optimal transport problem under the $\mathcal{W}_2$-metric (Villani, 2003; Ambrosio, Gigli and Savaré, 2005) are maximal monotone operators (Gangbo and McCann, 1996). The convergence of the empirical estimator is thus associated with problems relating to the convergence of convex functions (Hallin et al., 2021; del Barrio, Sanz and Hallin, 2024; Segers, 2022; del Barrio, González-Sanz and Loubes, 2024).

When dealing with data in a simple regression problem, one typically observes bivariate scatterplot data $(X_1, Y_1), \ldots, (X_n, Y_n)$ coming independently from the joint distribution $F_{X,Y}$ of scalar random variables X and Y, where we assume throughout that the density $f_{Y|X}(\cdot, x)$ of the conditional distribution of $Y|X = x$ exists. In this setting, a classical modeling assumption is to connect the response Y with the predictor X by postulating an additive measurement error model $Y = m(X) + \epsilon$, where ϵ is an error term that satisfies $E(\epsilon|X) = 0$ and $m(x) = E(Y|X = x)$ is the (unknown) conditional mean or regression function. To give a motivating example for the notion of Wasserstein-Fréchet integral of conditional distributions, suppose for the moment that the conditional distribution of $Y|X = x$ belongs to a location-scale family. This includes important situations such as Gaussian linear regression where the error term $\epsilon|X = x$ belongs to the Gaussian ensemble so that the distribution of $Y|X = x$ is recovered once $m(x)$ is available. This observation motivates to target a notion of center for the conditional distributions $Y|X = x$ that is able to track the underlying geometry of the probability model connecting Y with X, and therefore that of the error term ϵ conditionally on X.

A classical approach consists is to first estimate the unknown regression function $m(x)$ nonparametrically, e.g., by employing local constant or linear kernel estimators (Fan and Gijbels, 1996), followed by obtaining residuals $Y - m(X)$ at predictor levels $X = X_i$, and using these to recover the distribution of $\epsilon|X = x$. However, this multi-step process involves bandwidth or tuning parameter selection for the mean estimation step and requires careful analysis of bias and variance, as the estimate of the regression function m impacts the non-parametric distribution estimation step for the error term. Instead, we approach this problem more directly through the proposed Wasserstein-Fréchet integral of conditional distributions, obliviating the need to estimate the regression function m. For location-scale families, the proposed approach captures the distribution of the error term, as demonstrated in the following. Since the 2-Wasserstein metric is inherently connected to quantile functions (Villani, 2003), the proposed Wasserstein-Fréchet integral of conditional distributions is also related to the problem of conditional quantile estimation (Parzen, 1979; Falk, 1984, 1985;

Stute, 1986; Samanta, 1989; Bhattacharya and Gangopadhyay, 1990; Jones and Hall, 1990; Li and Racine, 2008; Yu and Jones, 1998).

The organization of the paper is as follows. We introduce the notion of Wasserstein-Fréchet integrals of conditional distributions in Section 2, where we also provide various examples and comparisons with traditional marginals. Empirical estimates along with theoretical convergence results are provided in Section 3 for different data generating mechanisms such as (1) when one observes scatterplot data $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{iid}{\sim} F_{X,Y}$; (2) when one has available fully observed conditional densities $f_i = f_{Y|X}(\cdot, X_i)$; and (3) when one has a dense sample $Y_{i1}, \ldots, Y_{im_i}$ coming from f_i which introduces further estimation errors that need to be accounted for. In Section 4 we study the optimal transport from the estimated conditional densities f_i to that of the Wasserstein-Fréchet integral of conditional distributions and provide theoretical justifications. In Section 5 we present simulation results for various settings and in Section 6 we illustrate the proposed estimates for data from bike pickups in the Divvy bike system in Chicago, for Covid-19 cases across states in the United States and for data on child development from the ECHO cohort study. Proofs of the theoretical results along with additional simulation settings can be found in the Appendix.

Some notes on notation: Throughout, we use $\mathcal{P}(\mathcal{Y})$ to denote the space of probability measures defined on $\mathcal{Y} \subset \mathbb{R}$, and $\mathcal{P}_p(\mathcal{Y})$ to refer to the subset of $\mathcal{P}(\mathcal{Y})$ containing probability measures with finite p-moments. The p-Wasserstein metric is denoted by $d_{\mathcal{W}_p}$, where we use the simplified notation $d_{\mathcal{W}}$ for the special case $d_{\mathcal{W}_2}$. We denote the space of density functions corresponding to the measures in $\mathcal{P}(\mathcal{Y})$ by $\mathcal{F}$, and $d^2_{L^2}$ represents the L^2-metric, defined as $d^2_{L^2}(f, g) = \int \{f(u) - g(u)\}^2 \, du$. We do not distinguish between the space of probability measures $\mathcal{P}(\mathcal{Y})$ and the space of density functions $\mathcal{F}$ when referring to the Wasserstein space.

2. Population Wasserstein-Fréchet integral of conditional distributions

For a metric space $(\mathcal{M}, d)$, an interval $\mathcal{T} \subset \mathbb{R}$ and an object-valued function $h : \mathcal{T} \to \mathcal{M}$, the Fréchet integral of the metric-space valued curve h is (Petersen and Müller, 2016a)

$$h^* = \underset{\omega \in \mathcal{M}}{\arg\inf} \int_{\mathcal{T}} d^2(h(t), \omega) dt.$$

Its existence can be guaranteed if $\mathcal{M}$ is compact and uniqueness if $\mathcal{M}$ is a Hadamard space; in the case of non-uniqueness the Fréchet integral h^* is a set of minimizers, in analogy to the Fréchet mean. We show below that the classical marginal density of the scalar random variable Y can be characterized as a generalized Fréchet integral (Dubey and Müller, 2020) when employing the L^2 metric in the space of density functions.

Suppose that one has scatterplot data $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{iid}{\sim} (X, Y)$, where X and Y are scalar random variables. Let $f_{Y|X}(\cdot, x)$ be the density of the conditional distribution of $Y|X = x$ with corresponding quantile function $Q_{Y|X}(t, x)$, $t \in (0, 1)$, where the support $\mathcal{X}$ of X is a compact interval. Suppose that $(\mathcal{F}_d, d)$ is a metric space with a suitable metric d, where $\mathcal{F}_d \subset \mathcal{F}$ is an appropriate subspace consisting of density functions of probability measures for which the underlying metric d is well defined. For example, when one employs the p-Wasserstein metric $d = d_{\mathcal{W}_p}$, it is natural to consider the metric space of density functions of probability measures with finite p-th moments. Let $\nu(x)$ be the probability measure with density function $f_{Y|X}(\cdot, x)$, $x \in \mathcal{X}$, which is assumed to reside in $\mathcal{F}_d$. The generalized Fréchet integral of $\nu(x)$, $x \in \mathcal{X}$, is

$$\nu_d^* = \underset{\omega \in \mathcal{F}_d}{\arg\inf} \int_{\mathcal{X}} d^2(\nu(x), \omega) dF(x), \tag{3}$$

where here and in the following we write F for F_X and f for f_X for the distribution function and density of the random variable X, which are assumed to exist on the support $\mathcal{X}$. The integral in (3) is understood in the Lebesgue-Stieltjes sense and can be written as $\nu^* = \underset{\omega \in \mathcal{F}_d}{\arg\inf} \int_{\mathcal{X}} d^2(\nu(x), \omega) f(x) dx$, in analogy to the generalized Fréchet integral introduced in Dubey and Müller (2020), where one can find further details about conditions for existence and uniqueness. We denote by F_ω, f_ω the distribution function and pdf of a probability measure ω and by $\mathcal{C}(\mathcal{Y})$ the space of density functions of absolutely continuous probability measures on $\mathcal{Y}$ with respect to the Lebesgue measure. If $\mathcal{F}_d = \mathcal{C}(\mathcal{Y})$ and d is the L^2 distance between corresponding density functions, namely $d^2(\nu(x), \omega) = \int_{\mathbb{R}} (f_{Y|X}(s, x) - f_\omega(s))^2 ds$, then $f^*_{\nu_{L^2}} = f_Y$.

Proposition 1. *Let $\mathcal{F}_d$ be the space of absolutely continuous probability measures with respect to the Lebesgue measure, endowed with the L^2 metric between density functions. Then the L^2 generalized marginal $\nu^*_{L^2}$ of $\nu(x)$, $x \in \mathcal{X}$, as in* (3)*, corresponds to the classical marginal distribution of Y.*

If we consider the 2-Wasserstein space $(\mathcal{F}_d, d) = (\mathcal{P}_2(\mathcal{Y}), d_{\mathcal{W}})$, then (3) leads to the notion of the *Wasserstein-Fréchet Integral*, which is a probability distribution that we denote by ν^*. Proposition 2 shows that the Wasserstein-Fréchet Integral ν^* has the quantile function

$$Q^*(t) = \int_{\mathcal{X}} Q_{Y|X}(t, x) dF(x), \; t \in (0, 1). \tag{4}$$

Proposition 2. *For the choice $(\mathcal{F}_d, d) = (\mathcal{P}_2(\mathcal{Y}), d_{\mathcal{W}})$ corresponding to the 2-Wasserstein space, if $\mathcal{Y} = [0, 1]$ is compact or $\mathcal{Y} = \mathbb{R}$ is unbounded and there exists a function $g \in L^2([0, 1])$ such that $Q_{Y|X}(t, x) \leq g(t)$ for all x, then the Wasserstein-Fréchet Integral ν_d^* in* (3) *is unique and has corresponding quantile function $Q(\nu_d^*) = Q^*$* (4)*.*

If f is continuous on the compact set $\mathcal{X}$, the condition $Q_{Y|X}(t, x) \leq g(t)$ in Proposition 2 is automatically satisfied when $\mathcal{Y}$ is compact; for the unbounded case $\mathcal{Y} = \mathbb{R}$ it is implied if $\sup_{x \in \mathcal{X}} |Q_{Y|X}(\cdot, x)| \in L^2([0, 1])$. The latter is a mild condition and holds in the Gaussian setting of Example 1 below whenever $\mu(\cdot)$ and $\sigma(\cdot)$ are continuous.

The proposed Wasserstein-Fréchet Integral is readily seen to keep track of the underlying geometry of the relationship between Y and X, as demonstrated in the following examples. We remark that since the Wasserstein-Fréchet Integral can be alternatively viewed as a barycenter in 2-Wasserstein space, some of its properties are well known (see for example Álvarez Esteban et al. (2016); Bigot and Klein (2018)).

Example 1. Consider a Gaussian linear regression setting, where $Y|X = x \sim N(\mu(x), \sigma^2(x))$, $\sigma^2(x) > 0$, $x \in \mathcal{X}$, and let Φ be the cdf of a standard normal random variate. Then $Q_{Y|X}(t, x) = \Phi^{-1}(t)\sigma(x) + \mu(x)$, $t \in (0, 1)$ and the Wasserstein-Fréchet Integral is given by

$$\nu^* = N\left(\int_{\mathcal{X}} \mu(x)dF(x), \left(\int_{\mathcal{X}} \sigma(x)dF(x)\right)^2\right),$$

where the mean and variance of ν^* correspond to the Euclidean barycenter of the random variables Y and $\sigma(X)$, respectively. This can be seen from the fact that $E(Y) = E(E(Y|X)) = E(\mu(X)) = \int_{\mathcal{X}} \mu(x)dF(x)$ for the mean and similarly for the variance.

Example 2. Consider the more general model $Y = g(X) + \epsilon$, where $g(X) = E(Y|X = x)$ is the regression function and ϵ is a random variable that may depend on X and is such that $E(\epsilon|X) = 0$. Proposition 2 and some simple calculations show that the quantile function of the Wasserstein-Fréchet Integral ν^* is given by

$$Q^*(t) = \int_{\mathcal{X}} q_{\epsilon|X}(t, x)dF(x) + E(g(X)), \quad t \in (0, 1),$$

where $q_{\epsilon|X}(\cdot, x)$ is the conditional quantile function associated with the distribution $\epsilon|X = x$. For the special case when the error ϵ is independent of X it follows that

$$Q^*(t) = q_\epsilon(t) + E(g(X)), \quad t \in (0, 1),$$

which shows that if the error distribution is a location-scale family, then the Wasserstein-Fréchet Integral belongs to the same family.

Example 3. Consider a location-scale family model for the conditional distribution $Y|X = x$, i.e., $f_{Y|X}(y, x) = \frac{1}{\sigma(x)} f_0\left(\frac{y-\mu(x)}{\sigma(x)}\right)$, where f_0 corresponds to a density function over $\mathbb{R}$ such that $\int_{\mathbb{R}} z f_0(z)dz = 0$ and $\int_{\mathbb{R}} z^2 f_0(z)dz = 1$, and $\mu(X)$ and $\sigma(X) > 0$ are (measurable) functions of X. Then the quantile function corresponding to the Wasserstein-Fréchet Integral is given by

$$Q^*(t) = Q_0(t) \int_{\mathcal{X}} \sigma(x)dF(x) + \int_{\mathcal{X}} \mu(x)dF(x), \quad t \in (0, 1),$$

where Q_0 is the quantile function corresponding to the density f_0, since for a location-scale model it holds that $Q_{Y|X}(t,x) = Q_0(t)\sigma(x) + \mu(x)$ by Proposition 2. Thus, the Wasserstein-Fréchet Integral belongs to the same location-scale family, with pdf

$$f_{\nu^*}(y) = \frac{1}{E(\sigma(X))} f_0\left(\frac{y - E(\mu(X))}{E(\sigma(X))}\right),\ y \in \mathcal{Y}.$$

In the context of the time-warping or curve registration problem in functional data analysis (Sakoe and Chiba, 1978; Kneip and Gasser, 1992; Wang and Gasser, 1997; Liu and Müller, 2004; Bigot and Charlier, 2011), one may consider an observation Y and a time shift random variable θ such that $f_{Y|\theta}(y,\theta) = f_0(y-\theta)$, and $\theta \sim g$ for some density g. This can be viewed in the framework of the location-scale model for the conditional distribution of $Y|X$ when taking $X = \theta$, $\sigma(X) = 1$ and $\mu = \text{id}$. Then the Wasserstein-Fréchet Integral has the density $f_{\nu^*}(\nu) = f_0(\nu - E(\theta))$, which is a preferred notion of center compared to the Euclidean marginal $f_E(y) = \int_{\mathcal{X}} f_0(\nu - x)dF_\theta(x)$, as in general the latter does not lie in the family of time-warped densities (Bigot and Charlier, 2011; Bigot, 2013) and does not adequately reflect the underlying shape of the template density f_0, which usually is the target of interest.

Example 4. Considering a convex combination of Gaussian measures, let $m > 0$, $\pi = (\pi_1, \dots, \pi_m)$ such that $\sum_{j=1}^m \pi_j = 1$ with $\pi_j \geq 0$, and introduce random means $\mu_1(X), \dots, \mu_m(X)$
$\in \mathbb{R}$ and random standard deviations $\sigma_1(X), \dots, \sigma_m(X) > 0$. Suppose that the distribution of $Y|X = x$ satisfies $Q_{Y|X}(t,x) = \sum_{j=1}^m \pi_j Q_j(t)$, $t \in (0,1)$, where Q_j is the quantile function of a Gaussian random variable $\mathcal{N}(\mu_j(x), \sigma_j^2(x))$. Then

$$\nu^* \sim \mathcal{N}\left(\sum_{j=1}^m \pi_j E(\mu_j(X)), \Big[\sum_{j=1}^m \pi_j E(\sigma_j(X))\Big]^2\right).$$

This situation can be viewed equivalently as an empirical Wasserstein barycenter (Agueh and Carlier, 2011; Panaretos and Zemel, 2019) of Gaussian measures $\nu_j \sim N(E[\mu_j(X)], E[\sigma_j(X)])$, $j = 1, \dots, p$, which has been well studied for the Gaussian case (Takatsu, 2011; Agueh and Carlier, 2011) and also for location-scatter families (Álvarez Esteban et al., 2016).

The following example shows that the proposed Wasserstein-Fréchet Integral measure remains in the distribution class of $Y|X$ for several well known probability models, which elucidates a closedness property for the proposed Wasserstein-Fréchet Integral and demonstrates the utility of the 2-Wasserstein geometry.

Example 5. The Wasserstein-Fréchet Integral ν^* remains in the distribution class corresponding to $Y|X$ for the following probability models:

- If $Y|X = x \sim \text{Lognormal}(\mu(x), \sigma)$ has a Log-normal distribution with parameters $\mu(x)$ and $\sigma > 0$, then

$$\nu^* \sim \text{Lognormal}(\log(\int_{\mathcal{X}} \exp(\mu(x)) dF(x)), \sigma).$$

- If $Y|X = x \sim \mathcal{E}(\theta(x))$ has an exponential distribution with inverse scale parameter $\theta(x) > 0$, then $\nu^* \sim \mathcal{E}(\int_{\mathcal{X}} \theta(x) dF(x))$.
- If $Y|X = x \sim \text{Gumbel}(\mu(x), \beta(x))$ has a Gumbel distribution with location and scale parameters $\mu(x)$ and $\beta(x) > 0$, respectively, then $\nu^* \sim \text{Gumbel}(\int_{\mathcal{X}} \mu(x) dF(x), \int_{\mathcal{X}} \beta(x) dF(x))$.
- If $Y|X = x \sim \text{Gompertz}(\lambda, \kappa(x))$ has a Gompertz distribution with shape and scale parameters $\lambda > 0$ and $\kappa(x) > 0$, respectively, then $\nu^* \sim \text{Gompertz}(\lambda, \{\int_{\mathcal{X}} [1/\kappa(x)] dF(x)\}^{-1})$.

In contrast, when d is chosen as the L^2 metric between corresponding density functions, the classical marginal measure typically does not track the underlying geometry of the distribution of $Y|X = x$ in important cases such as Gaussian location-scale families, as the classical marginal is then given by $f_Y(s) = f_{\nu^*_{L^2}}(s) = E\{\phi[(s - \mu(X))/\sigma(X)]/\sigma(X)\}$.

We next show that the proposed generalized Wasserstein-Fréchet Integral adapts to the geometry induced by other possible metrics for probability measures such as the 1-Wasserstein distance and the Hellinger distance, or equivalently the L^2 distance between the square-root densities viewed as elements of the unit Hilbert sphere in the context of the Fisher-Rao metric. For these important scenarios the generalized Fréchet integral in (3) admits a closed form.

For the 1-Wasserstein space, where $\mathcal{F}_d = \mathcal{P}_1(\mathcal{Y})$ and d is chosen as the 1-Wasserstein metric, which is equivalent to the L^1 distance between corresponding cdfs (Villani, 2003), Fubini's theorem shows

$$\begin{aligned}
\nu_1^* &= \underset{\omega \in \mathcal{P}_1(\mathcal{Y})}{\arg\inf} \int_{\mathcal{X}} \int_{\mathbb{R}} |F_{Y|X}(s,x) - F_\omega(s)| ds dF(x) \\
&= \underset{\omega \in \mathcal{P}_1(\mathcal{Y})}{\arg\inf} \int_{\mathbb{R}} E\left(|F_{Y|X}(s,X) - F_\omega(s)|\right) ds.
\end{aligned}$$

This is easily shown to be minimized when $F_\omega(s) = \text{median}(F_{Y|X}(s,X))$, $s \in \mathbb{R}$. That $\text{median}(F_{Y|X}(s,X))$ is a valid cdf is a consequence of it being contained in $(0,1)$, non-decreasing in $s \in \mathbb{R}$ and satisfying $\lim_{s\to-\infty} F_\omega(s) = 0$ and $\lim_{s\to+\infty} F_\omega(s) = 1$. In a classical linear regression setting, where $Y = \beta_0 + \beta_1 X + \epsilon$, $\epsilon \sim N(0, \sigma^2)$, the 1-Wasserstein-Fréchet Integral ν_1^* has cdf $F_{\nu_1^*}(s) = \Phi[(s - \beta_0 - \beta_1 \text{median}(X))/\sigma]$.

Example 6. Consider $\mathcal{F}_d = \mathcal{P}_1(\mathcal{Y})$ and the metric d to be the 1-Wasserstein metric so that the probability measure ν_1^* has cdf $F_1^*(s) = \text{median}(F_{Y|X}(s,X))$, $s \in \mathbb{R}$. The following relations hold.

- If $Y|X = x \sim \text{Lognormal}(\mu(x), \sigma)$ has a Log-normal distribution with parameters $\mu(x)$ and $\sigma > 0$, then $\nu_1^* \sim \text{Lognormal}(\text{median}(\mu(X)), \sigma)$.

- If $Y|X = x \sim \mathcal{E}(\lambda(x))$ has an exponential distribution with rate parameter $\lambda(x) > 0$, then $\nu_1^* \sim \mathcal{E}(\text{median}(\lambda(X)))$.
- If $Y|X = x \sim \text{Weibull}(\lambda(x), \kappa)$ has a Weibull distribution with scale and shape parameters $\lambda(x) > 0$ and $\kappa > 0$, respectively, then $\nu_1^* \sim \text{Weibull}(\text{median}(\lambda(X)), \kappa)$.
- If $Y|X = x \sim \text{Gompertz}(\lambda(x), \kappa)$ has a Gompertz distribution with shape and scale parameters $\lambda(x) > 0$ and $\kappa > 0$, respectively, then $\nu_1^* \sim \text{Gompertz}(\text{median}(\lambda(X)), \kappa)$.
- If $Y|X = x \sim \text{Gumbel}(\mu(x), \beta)$ has a Gumbel distribution with location and scale parameters $\mu(x)$ and $\beta > 0$, respectively, then

$$\nu_1^* \sim \text{Gumbel}(\text{median}(\mu(X)), \beta).$$

If d is the Hellinger distance on the space $\mathcal{F}_d$ of absolutely continuous probability measures with respect to Lebesgue measure on $\mathcal{Y}$, i.e., $d_{\text{H}}(f_1, f_2) = \|\sqrt{f_1} - \sqrt{f_2}\|_{L^2}$ with $f_1, f_2 \in \mathcal{F}_d$, then the Hellinger marginal in the sense of (3) is given by

$$\nu_{\text{H}}^* = \arg\inf_{\omega \in \mathcal{F}_d} E(d_{\text{H}}^2(f_{Y|X}(\cdot, X), g_\omega)),$$

where g_ω is the density function of the probability measure ω. With $\langle \cdot, \cdot \rangle_{L^2}$ denoting the L^2 inner product, the well known property $d_{\text{H}}^2(f_{Y|X}(\cdot, X), g_\omega) = 2 - 2\langle [f_{Y|X}(\cdot, X)]^{1/2}, g_\omega^{1/2} \rangle_{L^2}$ reveals that

$$\nu_{\text{H}}^* = \arg\sup_{\omega \in \mathcal{F}_d} \langle E([f_{Y|X}(\cdot, X)]^{1/2}), g_\omega^{1/2} \rangle_{L^2}.$$

For $h(\cdot) = E([f_{Y|X}(\cdot, X)]^{1/2})$, observing $\|g_\omega^{1/2}\|_{L^2} = 1$, since g_ω is a density function, and the inequality $\langle h, g_\omega^{1/2} \rangle_{L^2} \leq \|h\|_{L^2}$, one finds that $g_\omega = h^2 / \|h\|_{L^2}^2$ achieves the upper bound over the class of density functions. Therefore, the density corresponding to the Hellinger marginal ν_{H}^* is given by $f_{\text{H}}(\cdot) = h(\cdot)^2 / \int_{\mathcal{Y}} h^2(y) dy$.

In the following, we adopt the 2-Wasserstein metric in the space of probability distributions over $\mathcal{Y}$ as it successfully keeps track of the underlying geometry inherent to the model between the response Y and predictor X, especially when deformations are of key importance in applications (Bolstad et al., 2003).

3. Empirical Wasserstein-Fréchet integrals of conditional distributions and quantiles

3.1. Preliminaries

We now consider the estimation of the Wasserstein-Fréchet Integral ν^* which can be equivalently characterized through its corresponding quantile function $Q^* = Q(\nu^*)$ given by the relation $Q^*(t) = \int_{\mathcal{X}} Q_{Y|X}(t, x) dF(x)$, $t \in (0, 1)$, as shown in Proposition 2. Without loss of generality, suppose that $\mathcal{X} = [0, 1]$. As we assume throughout that densities exist, one can write alternatively $Q^*(t) = \int_{\mathcal{X}} Q_{Y|X}(t,x)f(x)dx$, $t \in (0,1)$. The latter suggests empirical estimates as follows. Suppose that an estimate $\hat{Q}_{Y|X}(\cdot,x)$ of $Q_{Y|X}(\cdot,x)$ is available such that it is a valid quantile function for all predictor levels $x \in \mathcal{X}$, and similarly for a density estimate $\hat{f}$ of f over $\mathcal{X}$. Then, a natural estimate of the quantile function Q^* corresponding to the Wasserstein-Fréchet Integral that is guaranteed to reside in the space of quantile functions is obtained by plugging-in the empirical counterparts into the expression for Q^*. This leads to $\hat{Q}^*(t) = \int_{\mathcal{X}} \hat{Q}_{Y|X}(t,x)\hat{f}(x)dx$, $t \in (0,1)$.

Several smoothing approaches have been extensively studied in the literature for estimating the conditional quantile function $Q_{Y|X}(\cdot,x)$, $x \in \mathcal{X}$. However, many of the known approaches are not directly applicable as they do not yield estimates that reside in quantile space. For example, local linear estimators of the quantile regression function $Q_{Y|X}(\cdot,x)$ (Yu and Jones, 1998) are not guaranteed to produce valid quantile functions. To address this issue, Nadaraya-Watson kernel type estimators have been proposed in the literature for estimating the conditional distribution function $F_{Y|X}(\cdot,x)$ that are guaranteed to be non-decreasing and lying in $[0,1]$, and which are then inverted to obtain the corresponding (conditional) quantile function (Stute, 1986; Hall, Wolff and Yao, 1999; Li and Racine, 2008). In this approach one first estimates the conditional cdf $\hat{F}_{Y|X}(\cdot,x)$ of the (conditional) distribution $Y|X=x$ by a Nadaraya-Watson kernel type estimator given by

$$\hat{F}_{Y|X}(y,x) = \frac{n^{-1}\sum_{i=1}^n 1_{\{Y_i \le y\}} K_h(X_i - x)}{\tilde{f}(x)},\ y \in \mathcal{Y}, \tag{5}$$

where $\tilde{f}(x) = n^{-1}\sum_{i=1}^n K_h(X_i - x)$ is the standard kernel density estimator of $f(x)$, K is a kernel function corresponding to a symmetric density function with compact support $[-1,1]$, $K_h(\cdot) = K(\cdot/h)/h$ and $h > 0$ is a bandwidth. It is clear that (5) is a valid cdf as it is non-decreasing with values in $[0,1]$. This estimate corresponds to the one introduced in Hall, Wolff and Yao (1999) when taking the weights $p(x)$ in this paper to be identically one. Then, a direct estimator of the conditional quantile function is constructed by inverting $\hat{F}_{Y|X}(\cdot,x)$ via

$$\hat{Q}_{Y|X}(t,x) = \inf\{y \in \mathcal{Y} : \hat{F}_{Y|X}(y,x) \ge t\}, \quad t \in [0,1], \tag{6}$$

which is a valid and quantile function; see also Yu and Jones (1998); Fan, Hu and Truong (1994); Fan and Gijbels (1996). A parametric version is provided by quantile regression, which employs M-estimation (Koenker and Bassett, 1978) but may suffer from the problem of crossing quantiles (Chernozhukov, Fernández-Val and Galichon, 2010) and requires parametric assumptions.

We develop here an intrinsic estimation framework for Wasserstein-Fréchet integral of conditional distributions, for which we adopt a double-kernel smoothing approach. For a situation with compact support $\mathcal{X}$ of X, we utilize the boundary-corrected kernel density estimator $\hat{f}$ for f of Petersen and Müller (2016b), which adjusts for boundary effects while producing bona-fide density estimates, i.e., $\int_{\mathcal{X}} \hat{f}(x)dx = 1$ and $\hat{f}(x) \geq 0$, $x \in \mathcal{X}$. Let κ_2 be a kernel function corresponding to a density function over $[-1, 1]$ that is symmetric around 0. Then this boundary-corrected kernel density estimator for the density f of the predictor X is

$$\hat{f}(x) = \sum_{i=1}^{n} \kappa_2 \left(\frac{X_i - x}{h} \right) w(x, h) \Big/ \int_0^1 \kappa_2 \left(\frac{X_i - s}{h} \right) w(s, h) ds,$$

where the weights

$$\begin{aligned} w(x,h) =& (\int_{-x/h}^{1} \kappa_2(u)du)^{-1} 1_{\{x \in [0,h)\}} \\ &+ \left(\int_{-1}^{(1-x)/h} \kappa_2(u)du \right)^{-1} 1_{\{x \in (1-h,1]\}} + 1_{\{x \in [h,1-h]\}} \end{aligned}$$

are constructed in such a ways as to correct for the boundary bias.

Suppose that $\mathcal{Y}$ is compact and without loss of generality consider $\mathcal{Y} = [0, 1]$. For the estimation of the conditional density $f_{Y|X}(\cdot, x) = f_{Y,X}(\cdot, x)/f(x)$, we further adjust the classical double kernel based quotient type estimate (Stone, 1977; Samanta, 1989). For bandwidths $h_1, h < 0.5$, the estimate $\hat{f}_{Y|X}(y, x)$ of $f_{Y|X}(y, x)$ is given by

$$\hat{f}_{Y|X}(y,x) = \frac{\sum_{i=1}^{n} \kappa_1 \left(\frac{Y_i - y}{h_1} \right) w_1(y, h_1) \kappa_2 \left(\frac{X_i - x}{h} \right) w(x, h)}{\sum_{i=1}^{n} \kappa_2 \left(\frac{X_i - x}{h} \right) w(x, h) \int_0^1 \kappa_1 \left(\frac{Y_i - s}{h_1} \right) w_1(s, h_1) ds}, \ y \in \mathcal{Y}, \ x \in \mathcal{X},$$

where w_1 is defined analogously as w but replacing the kernel κ_2 by κ_1. This allows to correct for boundary bias in both $\mathcal{X}$ and $\mathcal{Y}$. The estimated conditional cdf can be directly obtained by integration, $\hat{F}_{Y|X}(y, x) = \int_0^y \hat{f}_{Y|X}(s, x)ds$, $y \in \mathcal{Y}$. It is easy to see that the estimated conditional density $\hat{f}_{Y|X}(\cdot, x)$ is a valid (conditional) density function since $\int_0^1 \hat{f}_{Y|X}(y, x)dy = 1$ for all x and $\hat{f}_{Y|X}(\cdot, x) \geq 0$, whence $\hat{F}_{Y|X}(\cdot, x)$ is a valid (conditional) cdf for all x. Then $\hat{Q}_{Y|X}(\cdot, x)$ is obtained by inverting the estimated conditional cdf $\hat{F}_{Y|X}(\cdot, x)$ using (6). Finally, the estimate of the quantile function $Q^* = Q(\nu^*)$ of the Wasserstein-Fréchet Integral ν^* is $\hat{Q}^*(t) = \int_{\mathcal{X}} \hat{Q}_{Y|X}(t, x)\hat{f}(x)dx$, $t \in (0, 1)$.

3.2. Estimation of the Wasserstein-Fréchet integral for scatterplot data

Suppose that the available data consists of a scatterplot $(X_1, Y_1), \ldots, (X_n, Y_n)$ sampled from the joint distribution of (X, Y). The following result shows that the Wasserstein-Fréchet Integral ν^* can be consistently recovered in the 2-Wasserstein metric. Denote by $\|g\|_\infty = \sup_{x \in \mathcal{X}} |g(x)|$ the supremum norm for a function $g : \mathcal{X} \to \mathbb{R}$. We require that

(S1) $f_{Y|X} : \mathcal{Y} \times \mathcal{X} \to \mathbb{R}^+$ is positive and continuously differentiable.
(S2) The density function f of X is continuously differentiable and satisfies $\inf_{x \in \mathcal{X}} f(x) \geq M$ for some $M > 0$.

Assumptions (S1)-(S2) are mild regularity conditions. If f is continuously differentiable and strictly positive, (S2) is immediately satisfied due to compactness of $\mathcal{X}$. Denote by $\mathcal{K}$ the space of kernel functions $\kappa_0 : \mathbb{R} \to \mathbb{R}$ that correspond to a continuous density function that is symmetric around 0 and has compact support $[-1, 1]$.

Theorem 3.1. *Suppose that* $\mathcal{Y} = [0, 1]$, $\mathcal{X} = [0, 1]$, $\kappa_1, \kappa_2 \in \mathcal{K}$ *and the regularity conditions (S1)-(S2) hold. If* $h = h_n = n^{-1/3}$*, then*

$$d^2_{\mathcal{W}}(\hat{\nu}^*, \nu^*) = O_p(h_1^2 + n^{-2/3}).$$

Choosing $h_1 = n^{-1/3}$ in Theorem 3.1 leads to the rate $O_p(n^{-2/3})$, which is due to the estimation of f in the L^2 norm, and is optimal for the class of differentiable density functions that are bounded away from zero (Petersen and Müller, 2016b). Since the W_1 metric is the weakest of all W_p metrics, $p \geq 1$, (Villani, 2009, Remark 6.6, page 107), it follows that the L^1-distance between the estimated and true Wasserstein-Fréchet Integral cdfs can also be consistently recovered.

Considering an extension to the case $\mathcal{Y} = \mathbb{R}$ where the support of the density $f_{Y|X}(\cdot, x)$ is unbounded, there is no need for boundary correction in this case and therefore we slightly adjust the estimators as follows. For simplicity, suppose that the kernel function κ_1 has unbounded support $\mathbb{R}$ (such as is the case for a Gaussian kernel). The estimate of $f_{Y|X}(y, x)$ is given by

$$\hat{f}_{Y|X}(y, x) = \frac{\sum_{i=1}^n \frac{1}{h_1} \kappa_1\left(\frac{Y_i - y}{h_1}\right) \kappa_2\left(\frac{X_i - x}{h}\right) w(x, h)}{\sum_{i=1}^n \kappa_2\left(\frac{X_i - x}{h}\right) w(x, h)},$$

and then the estimated conditional cdf $\hat{F}_{Y|X}(y, x) = \int_{-\infty}^y \hat{f}_{Y|X}(s, x) ds$, $y \in \mathbb{R}$ is

$$\hat{F}_{Y|X}(y, x) = \sum_{i=1}^n G\left(\frac{y - Y_i}{h_1}\right) \kappa_2\left(\frac{X_i - x}{h}\right) w(x, h) \Big/ \sum_{i=1}^n \kappa_2\left(\frac{X_i - x}{h}\right) w(x, h),$$

where $G(z) = \int_{-\infty}^z \kappa_1(s) ds$, $z \in \mathbb{R}$. Clearly $\hat{F}_{Y|X}(y, x)$ is non-decreasing in y for all x, and satisfies $\hat{F}_{Y|X}(y, x) \to 1$ as $y \to \infty$ and $\hat{F}_{Y|X}(y, x) \to 0$ as $y \to -\infty$, so that it is a valid (conditional) cdf. We require the following regularity conditions in the unbounded support case.

(S1') There exists $\delta \in (0, 1)$ such that $\int_{\mathcal{X}} \int_\delta^{1-\delta} Q^2_{Y|X}(t, x) dt dx < \infty$.
(S2') It holds that $\sup_{x \in \mathcal{X}, y \in \mathcal{Y}} |\frac{\partial^2}{\partial x^2} F_{Y,X}(y, x)| < \infty$ and

$$\sup_{x \in \mathcal{X}, y \in \mathcal{Y}} f_{Y|X}(y, x) < \infty.$$

(K1') κ_1 corresponds to a continuous density function with unbounded support $\mathbb{R}$ that is symmetric around 0 and satisfies $\int_{\mathbb{R}}|u|\kappa_1(u)du < \infty$.

Assumption (S1') is a mild condition that is satisfied if $\int_{\mathcal{X}} E(Y^2|X = x)dx < \infty$, which in turns holds if $E(Y^2|X = x)$ is continuous in x. In this case any $\delta \in (0, 1)$ can be taken in (S1'). Theorem 3.2 shows that ν^* can be consistently estimated in the interior, i.e., in terms of the δ-truncated 2-Wasserstein distance defined by $d^2_{\mathcal{W}_\delta}(\hat{\nu}^*, \nu^*) = \int_\delta^{1-\delta}(\hat{Q}^*(t) - Q^*(t))^2 dt$, with δ as in (S1').

Theorem 3.2. *Suppose that $\mathcal{Y} = \mathbb{R}$, $\mathcal{X} = [0, 1]$, $\kappa_2 \in \mathcal{K}$ and the regularity conditions (S1), (S2), (S2') and (K1') hold. Let $\delta \in (0, 1)$ be defined as in (S1') and $h = h_n = n^{-1/3}$. Suppose that $h_1 n^{p-1/3} \to \infty$ as $n \to \infty$ for some $p > 2/3$. Then*

$$d^2_{\mathcal{W}_\delta}(\hat{\nu}^*, \nu^*) = O_p(h_1^2 + n^{-2/3}).$$

Again choosing the bandwidth sequence as $h_1 = n^{-1/3}$ in Theorem 3.2 leads to the rate $O_p(n^{-2/3})$. For situations in which the conditional densities $f_i = f_{Y|X}(\cdot, X_i)$ can be assumed to be fully observed across subjects, one can readily utilize this additional information when constructing estimates of the Wasserstein-Fréchet Integral, as follows.

3.3. Estimation of the Wasserstein-Fréchet integral for fully observed and for estimated conditional densities

For another perspective, suppose that the conditional densities $f_i = f_{Y|X}(\cdot, X_i)$ are fully observed, in which case the available data is the i.i.d. sample

$$\{(X_1, f_1), \ldots, (X_n, f_n)\}.$$

Denoting by Q_i the quantile function corresponding to f_i, the Q_i form an i.i.d. sample of random quantile functions with $E(Q_1(t)) = Q^*(t)$, $t \in (0, 1)$, where $Q^* = Q(\nu^*)$ is the quantile function of the Wasserstein-Fréchet Integral ν^*. Then the law of large numbers suggests the estimate

$$\hat{Q}^*(t) = n^{-1} \sum_{i=1}^{n} Q_i(t), \quad t \in (0, 1),$$

for $Q(\nu^*)$. The next result shows that ν^* can be consistently recovered in the 2-Wasserstein metric at the $\sqrt{n}$-rate for predictors $X \in \mathbb{R}^p$, $p \geq 1$.

Theorem 3.3. *Suppose that $\mathcal{Y} = [0, 1]$ and $\mathcal{X} = [0, 1]^p$, $p \geq 1$. If the densities $f_i(\cdot) = f_{Y|X}(\cdot, X_i)$, $i = 1, \ldots, n$, are fully observed, then*

$$d^2_{\mathcal{W}}(\hat{\nu}^*, \nu^*) = O_p(n^{-1}).$$

We remark that when the densities f_i have unbounded support $\mathbb{R}$, the result still holds provided that $\int_0^1 \mathrm{Var}(Q_1(t))dt < \infty$, which is a mild regularity condition. For example, in the location-scale model of Example 3, this is satisfied when $\mathrm{Var}(\mu(X)) < \infty$ and $\mathrm{Var}(\sigma(X)) < \infty$.

However, for situations in which the conditional densities f_i remain latent and are not directly observed, which is the most common scenario in data analysis, these estimates are not feasible and Theorem 3.3 does not apply. A common framework is to instead assume that one has an increasing sample of observations coming from such densities f_i, so that consistent estimation of the individual f_i is possible (Bigot et al., 2018; Petersen and Müller, 2016b). This introduces further estimation errors that need to be accounted for in the final estimate of the Wasserstein-Fréchet Integral. To this end, suppose that the available data consist of a random sample $Y_{i1}, \dots, Y_{in_i} \overset{iid}{\sim} f_i$ for each f_i, where $f_i(\cdot) = f_{Y|X}(\cdot, X_i)$, i.e., a random sample from the conditional distribution $f_{Y|X}(\cdot, X_i)$ at predictor level X_i is available. We then estimate the individual density f_i by

$$\hat{f}_i(y) = \sum_{j=1}^{n_i} \kappa_1\left(\frac{Y_{ij} - y}{h_1}\right) w_1(y,h) \Big/ \sum_{j=1}^{n_i} \int_0^1 \kappa_1\left(\frac{Y_{ij} - s}{h_1}\right) w_1(s,h) ds,$$

where $y \in \mathcal{Y} = [0,1]$. Let $\hat{F}_i(y) = \int_0^y \hat{f}_i(\nu) d\nu$ be the estimated cdf obtained from the density estimate $\hat{f}_i$. By the law of large numbers, $n^{-1}\sum_{i=1}^n F_i^{-1}(t)$ converges in probability to $E(F_1^{-1}(t)) = E[E(F_1^{-1}(t)|X_1)] = E[Q_{Y|X}(t, X_1)] = Q^*(t)$ at the $\sqrt{n}$-rate. Thus, if $\hat{F}_i^{-1}(t)$ can be shown to be close to $F_i^{-1}(t)$, a natural estimate of $Q^*(t)$, $t \in (0,1)$, is $\hat{Q}^*(t) = n^{-1}\sum_{i=1}^n \hat{F}_i^{-1}(t)$.

Theorem 3.4. *Suppose that $\mathcal{Y} = [0,1]$, $\mathcal{X} = [0,1]^p$, $p \geq 1$, $\kappa_1 \in \mathcal{K}$ and the regularity conditions (S1) and (S2) hold, furthermore that $Y_{i1}, \dots, Y_{in_i} \overset{iid}{\sim} f_i$, where $f_i(\cdot) = f_{Y|X}(\cdot, X_i)$ and $n_i \geq 1$, $i = 1, \dots, n$. Suppose that $\min_{i=1,\dots,n} n_i \geq m(n)$, where $m(n) \to \infty$ as $n \to \infty$. If $m(n)h_1^4 = O(1)$ as $n \to \infty$ and $f_{Y|X}(\cdot, x)$ is twice-continuously differentiable for all x, then*

$$d_{\mathcal{W}}^2(\hat{\nu}^*, \nu^*) = O_p\left(n^{-1} + m(n)^{-1}\right).$$

Thus, if $m(n) = n$ which then requires $h_1 = O(n^{-1/4})$, potentially undersmoothing the individual density estimates $\hat{f}_i$, the overall n^{-1} convergence rate is maintained even for the more realistic case where the densities need to be estimated.

4. Optimal transport to the Wasserstein-Fréchet integrals

Next we study optimal transport maps (Villani, 2003) $T_i : [0,T] \to [0,T]$ from the individual (conditional) measures ν_i with corresponding density functions $f_i = f(\nu_i)$ to the Wasserstein-Fréchet Integral ν^*, which minimizes the transport cost

$$T_i = \underset{T_0:[0,T]\to[0,T]}{\arg\min} \int_0^T (y - T_0(y))^2 f_i(y) dy,$$

where T_0 pushes ν_i forward to ν^*, i.e., $T_0 \# \nu_i = \nu^*$ (Panaretos and Zemel, 2019). This problem has a well known solution given by $T_i(y) = Q^*(F_i(y))$, $y \in \mathcal{Y}$. With available empirical estimates $\hat{F}_i$ of F_i and $\hat{Q}^*$ of Q^*, a natural estimate of the optimal transport map T_i is then given by $\hat{T}_i^*(y) = \hat{Q}^*(\hat{F}_i(y))$, $y \in \mathcal{Y}$. The next result shows that the optimal transport maps across all subjects can be uniformly recovered when the conditional densities f_i are fully observed. Since in this case $F_i = \int_0^y f_i(s)ds$ is fully observed, the empirical optimal transport estimate is given by $\hat{T}_i^*(y) = \hat{Q}^*(F_i(y))$, $y \in \mathcal{Y}$.

Theorem 4.1. *Suppose that $\mathcal{Y} = [0,1]$, $\mathcal{X} = [0,1]^p$, $p \geq 1$, and there exists $L > 0$ such that $Q_{Y|X}(\cdot, x)$ is L-Lipschitz for all x. If the densities $f_i(\cdot) = f_{Y|X}(\cdot, X_i)$, $i = 1, \ldots, n$, are fully observed, then*

$$\max_{i=1,\ldots,n} \sup_{y \in \mathcal{Y}} |\hat{T}_i^*(y) - T_i^*(y)| = O_p\left((\log n/n)^{1/2}\right).$$

Ranking the individual observations $Y_1, \ldots, Y_n$ while adjusting for each corresponding covariate level X_i, $i = 1, \ldots, n$, is also of interest. In the context of time-varying functional data but without additional covariates, Chen, Dawson and Müller (2020) proposed to study the rank dynamics of a process $Y_0(\cdot)$ observed on a compact time window by targeting the cross-sectional percentile $F_t(Y_0(t))$, where $F_t(y) = P(Y_0(t) \leq y)$. Since here the responses Y_i are scalar, a similar idea would be to rank the Y_i based on the classical marginal cdf $F_Y(y) = \int_0^y f_Y(s)ds$. Instead, we propose to utilize the cdf of the Wasserstein-Fréchet Integral, where we obtain percentiles of the observations $Y_1, \ldots, Y_n$ through the Wasserstein-Fréchet Integral distribution functions $F^*(Y_i)$, $i = 1, \ldots, n$. This approach to ranking incorporates the corresponding covariate levels X_i in a natural way and removes the effect of the conditional mean. It basically ranks the responses across all subjects by how their response relates to the mean, i.e., to what degree it is above, at or below the conditional mean.

The following results justify this new approach. It shows that when the (conditional) densities are fully observed, the cdf of the Wasserstein-Fréchet Integral can be uniformly recovered across subjects.

Theorem 4.2. *If the conditions of Theorem 4.1 and (S1), (S2) are satisfied,*

$$\sup_{y \in \mathcal{Y}} |\hat{F}^*(y) - F^*(y)| = O_p\left(\left(\frac{\log n}{n}\right)^{1/2}\right).$$

For situations in which the (conditional) densities f_i remain unobserved and one only has available a sample of observations generated from each distribution, the optimal transports across all subjects can still be uniformly recovered, as per the following result.

Theorem 4.3. *Suppose that $\mathcal{Y} = [0,1]$, $\mathcal{X} = [0,1]^p$, $p \geq 1$, $\kappa_1 \in \mathcal{K}$, $n_i \geq m(n)$ with $m(n) \to \infty$ as $n \to \infty$, (S1) holds and $f_{Y|X}(\cdot, x)$ is twice-continuously differentiable for all x. If $m(n)h_1^4 = O(1)$, $m(n)^{2\gamma_0 - 1}h_1^2 \log m(n) \to \infty$ and*

$n/m(n)^{\rho_1} = o(1)$ as $n \to \infty$ for some $\gamma_0 > 1/2$ and $\rho_1 > 0$, then

$$\max_{i=1,\dots,n} \sup_{y\in\mathcal{Y}} |\hat{T}_i^*(y) - T_i^*(y)| = O_p\left(\left(\frac{\log n}{n}\right)^{1/2} + \left(\frac{\log m(n)}{m(n)}\right)^{1/2}\right).$$

Thus, if $h_1 = m(n)^{-\upsilon}$ for some $\upsilon > 1/4$, then one can take any $\gamma_0 > 1/2 + \upsilon$ in Theorem 4.3 to fulfill the condition $m(n)^{2\gamma_0-1}h_1^2 \log m(n) \to \infty$ as $n \to \infty$. If $m(n) = n^\rho$ has a polynomial growth rate for some $\rho > 0$, then taking any $\rho_1 > \rho^{-1}$ satisfies the condition $n/m(n)^{\rho_1} = o(1)$ as $n \to \infty$. The rate $O_p((\log n/n)^{1/2})$ is achieved provided that $m = m(n)$ grows faster than n. Since the uniform recovery of the optimal transports is naturally tied to the uniform estimation of Q^*, Lemma A.8 in the Appendix provides the rate of convergence for $\sup_{t\in(0,1)}|\hat{Q}^*(t) - Q^*(t)|$, which is a stronger metric compared to the 2-Wasserstein metric that governs the result in Theorem 3.4.

The next result shows that the Wasserstein ranking of the observations

$$Y_1, \dots, Y_n$$

can be uniformly recovered when the f_i are unknown as long as one has available an increasing sample of observations that are generated by each f_i.

Theorem 4.4. *Under the regularity conditions of 4.3*

$$\sup_{y\in\mathcal{Y}} |\hat{F}^*(y) - F^*(y)| = O_p\left(\left(\frac{\log n}{n}\right)^{1/2} + \left(\frac{\log m(n)}{m(n)}\right)^{1/2}\right),$$

where $\hat{F}^*(y) = \inf_{t\in[0,1]}\{\hat{Q}^*(t) > y\}$.

The fact that $\hat{F}^*(y)$ is a valid cdf, i.e., right-continuous on $\mathcal{Y}$ with $\hat{F}^*(0) = 0$ and $\hat{F}^*(1) = 1$, is a consequence of the non-decreasing property of $\hat{Q}^*$ (Feng et al., 2012). Here we take the convention that $\inf_{t\in[0,1]} \emptyset = 1$ as $\mathcal{Y} = [0, 1]$

5. Simulations

We consider the following random mechanism that generates fully observed conditional densities: First generate predictors $X_1, \dots, X_n \overset{iid}{\sim} U(0, 1)$. Then, conditional on X_i, the conditional density $f_i(\cdot) = f_{Y|X}(\cdot, X_i)$ corresponds to the density function of a truncated normal random variate with support $[0, 1]$ and parameters $\mu = a_0 + b_0X_i$ and $\sigma = a_1 + b_1X_i$, where the scalars a_0, b_0, a_1, b_1 are such that $a_1 + b_1x > 0$ for any $x \in [0, 1]$. Since f_i is assumed to be fully observed, we equivalently consider its corresponding quantile function Q_i. The estimated quantile function of the Wasserstein-Fréchet Integral is given by $\hat{Q}^*(t) = n^{-1}\sum_{i=1}^n Q_i(t)$, which is computed over a dense grid in $[0, 1]$, while its true counterpart is obtained numerically using that $Q^*(t) = \int_{\mathcal{X}} Q_{Y|X}(t, x)f(x)dx$, where f is the uniform density in $[0, 1]$ and $Q_{Y|X}(\cdot, x)$ corresponds to the quantile function of a truncated normal variate with support $[0, 1]$ and parameters $\mu = a_0 + b_0x$ and $\sigma = a_1 + b_1x$. We set $a_0 = 0.1$, $b_0 = 0.9$, $a_1 = 0.1$ and $b_1 = 0.1$.

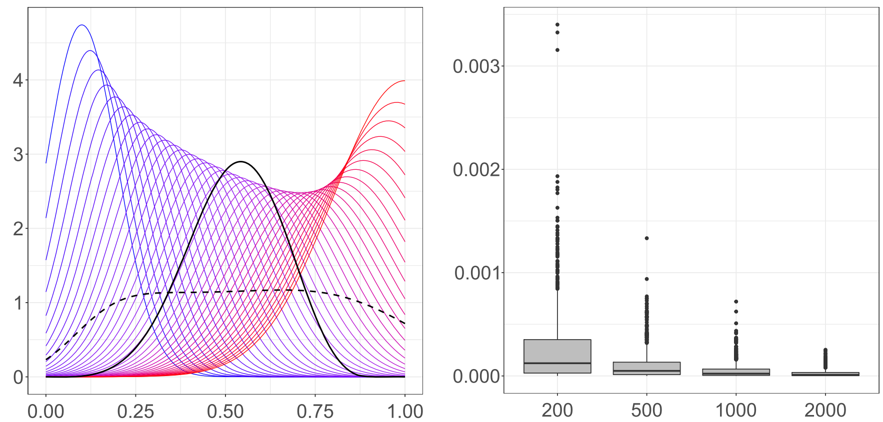

FIG 1. *Left panel: Conditional truncated normal density functions $f_{Y|X}(\cdot, x)$ with support $[0,1]$ and parameters $\mu = a_0 + b_0 x$ and $\sigma = a_1 + b_1 x$ over a dense grid of increasing x values in $[0,1]$ (from blue to red) as outlined in the simulation setting for fully observed conditional densities. The density of the Wasserstein-Fréchet Integral measure is shown in solid black while the classical marginal is dashed. Right panel: Boxplots of squared Wasserstein distances $\mathcal{D}_n^2$ for 1000 simulations for increasing sample sizes in the same simulation setting. Here $\mathcal{D}_n^2 = d_{\mathcal{W}}^2(\hat{\nu}^*, \nu^*) = \int_0^1 (\hat{Q}^*(t) - Q^*(t))^2 dt$ is estimated numerically over a dense grid in $[0,1]$.*

The left panel of Figure 1 shows the density function of the truncated conditional Gaussian variates at different predictor levels x along with the density function corresponding to the Wasserstein-Fréchet Integral as well as the classical marginal of Y. It is clearly seen that the classical marginal measure does not track the underlying geometry of the random mechanism that generates conditional densities, in contrast to the Wasserstein-Fréchet Integral, which captures this geometry. To assess the finite sample performance of the empirical estimates, we compute the squared 2-Wasserstein distance $\mathcal{D}_n^2 = d_{\mathcal{W}}^2(\hat{\nu}^*, \nu^*) = \int_0^1 (\hat{Q}^*(t) - Q^*(t))^2 dt$ between the estimated Wasserstein-Fréchet Integral measure $\hat{\nu}^*$ and its true counterpart ν^* for increasing sample sizes n. The right panel of Figure 1 shows the resulting boxplots of $\mathcal{D}_n^2$ across 1000 simulations. These boxplots shrink towards zero as n increases, which demonstrates the convergence of the estimated Wasserstein-Fréchet Integral measure towards its true counterpart. Figure 11 in the Appendix illustrates the individual optimal transports from the conditional densities f_i to the Wasserstein-Fréchet Integral density. For low values of the predictor X, points $y \in \mathcal{Y}$ closer to 0 are pushed strongly to the right, while for higher values of X the points y they are pushed to the left.

We also investigate in the Appendix the situation where only scatterplot data (X_i, Y_i), $i = 1, \ldots, n$, are available and the conditional densities f_i are latent and not directly observed. We explore a scenario where Y_i comes from a mixture Gaussian distribution and X has a beta distribution.

6. Data applications

6.1. Daily bike rental data

This dataset contains the bike pick-up times and locations for bike rentals and is publicly available at https://www.divvybikes.com/system-data. We illustrate the proposed Wasserstein-Fréchet Integral for the Divvy bike trip records in Chicago during weekdays of 2019. These data are publicly available at https://www.divvybikes.com/system-data and contain the individual pick-up times for bike rentals at several locations in the Divvy bike system in Chicago. The data have been analyzed previously from a replicated point process perspective in Gervini and Khanal (2019); Gajardo and Müller (2021). We consider the response Y to be the total number of bike trips during non-holiday weekdays that originate at the station located near the intersection of Clinton St and Washington Blvd. As predictor X we take the observed daily temperature in Chicago from the station "Northerly Island" which is publicly available at https://www.ncdc.noaa.gov/cdo-web/.

Thus we have available scatterplot data $(X_i, Y_i)_{i=1}^n$ for $n = 222$ weekdays excluding holidays, where we disregard days with very low temperature below $-5°$C. The left panel of Figure 2 shows the scatterplot between total daily bike pickups and daily temperature which seems to suggest a near-linear regression relationship between the two variables, where both the (conditional) mean and variance appear to vary with the predictor level x. One would expect the conditional density function $f_{Y|X}(\cdot, x)$ to be unimodal. The right panel of Figure 2 displays the estimated Wasserstein-Fréchet Integral along with the classical marginal in terms of their density functions. It is clearly seen that the Wasserstein-Fréchet Integral is unimodal and resembles a Gaussian shape with some asymmetry, while the classical marginal is spread out and has a bimodal appearance, which is not unexpected as it reflects the entire distribution of the Y without taking into account the shifting means of the conditional distributions. The domain for X was chosen as $\mathcal{X} = [-5, 29.4]$ and we used a Gaussian kernel for κ_1 and an Epanechnikov kernel for κ_2. The bandwidth sequences h and h_1 were taken as 10% of the observed range for X and Y, respectively. The Wasserstein-Fréchet Integral is seen to be useful in summarizing the distribution of the measurements, in contrast to the classical marginal.

6.2. COVID-19 cases

As a second data application, we consider the COVID-19 daily confirmed case trajectories across states in the United States. These data are publicly available at https://github.com/CSSEGISandData/COVID-19 (accessed on December 7, 2021) from the COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. This dataset has been extensively analyzed from different perspectives, including functional data analysis and point processes (see, e.g., Carroll et al., 2020). We obtained the daily

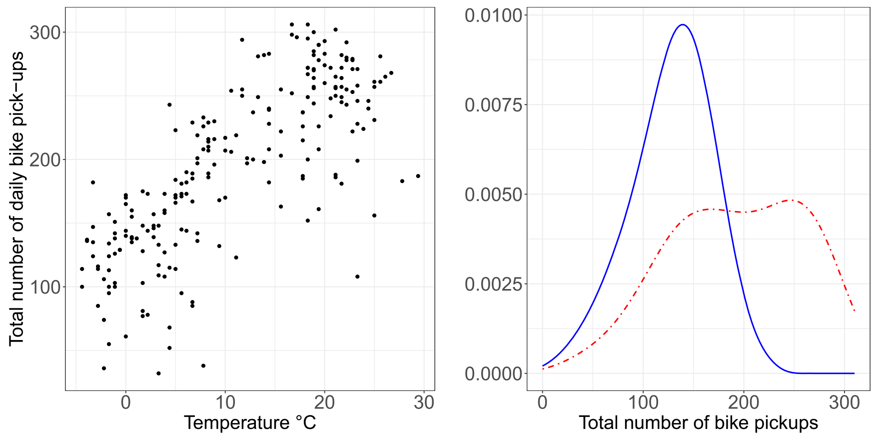

FIG 2. *The left panel shows the scatterplot of (X, Y), where X is the daily temperature in degree Celsius and Y corresponds to the total number of bike pick-ups during weekdays, excluding holidays, in* 2019 *at a bike station in the Divvy bike system in Chicago. The right panel shows the estimated Wasserstein-Fréchet Integral (blue solid line) and classical marginal densities (red dashed line) of the random fluctuations in the y-direction.*

confirmed cases for the time period between September 1, 2020 and March 31, 2021, and converted these into histograms of the distribution of cases over this time domain, using the same approach as described in Gajardo and Müller (2022).

As associated covariate we took the total number of cases per capita during the time window for each state, where the state population for 2019 was obtained from www.census.gov/programs-surveys/popest/data/data-sets.html. This covariate provides an indicator for the intensity of the infections in the respective state. The conditional densities f_i were obtained in a pre-smoothing step, which was performed by employing the frechet R package (Chen et al., 2020). The smoothing bandwidth employed for this smoothing step was constrained to be below 4% of the domain to avoid oversmoothing. To further mitigate boundary effects in the pre-smoothing step, we also used the daily confirmed cases prior to and after the selected time domain when obtaining the density estimates.

Figure 3 illustrates the individual conditional densities f_i for each state along with the classical marginal distribution and the proposed Wasserstein-Fréchet Integral, both visualized as densities. Again the Wasserstein-Fréchet Integral emerges as a preferred summary measure, as it provides a reasonable quantification of the variation in these data, whereas the classical marginal again is bimodal and spread out, which makes it much less useful as a data summary.

6.3. Child development cohort data

The Environmental Influences on Child Health Outcomes (ECHO) study (dash.

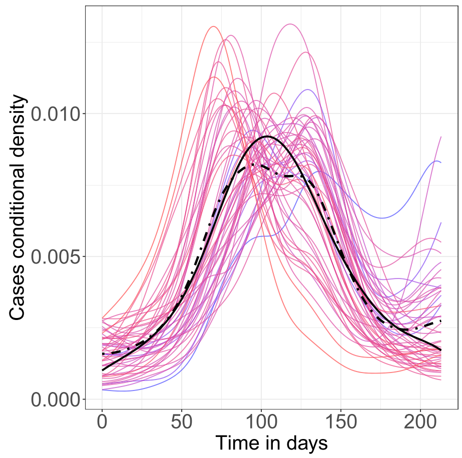

Fig 3. *Density functions of daily confirmed cases of COVID-19 for each state over the time window* $[0, T]$*, where time* $t = 0$ *corresponds to September* 1, 2020 *and* $T = 7$ *months. The color of each density reflects the value of the covariate, which ranges from low values (blue) to high values (red), where the covariate is the total number of cases per capita over the time window or each state. The estimated densities of the Wasserstein-Fréchet Integral (solid black) and classical marginal distributions (dashed black) are also displayed.*

nichd.nih.gov/study/417122) collected longitudinal data on 30,000 pregnancies and 50,000 children from 69 pediatric cohorts to investigate the effects of environmental factors on child health outcomes. The available data to date can be obtained from NIH. Our study focuses on the "AAX03" cohort, which provides 1574 longitudinal measurements from physical exams, including anthropometry for children aged two years and older. We use age as a predictor variable, denoted as X, and analyze the conditional distributions of anthropometric variables Y, such as height and weight, given X.

We first consider the weight of the child as the response variable Y. The age domain $\mathcal{I}$ is divided into 20 bins $S_1 = [a_0, a_1), S_2 = [a_1, a_2), \ldots, S_{20} = [a_{19}, a_{20}]$, where the a_j are chosen such that $\min(\mathcal{I}) = a_0 < a_1 < \ldots < a_{20} = \max(\mathcal{I})$, and the number of samples in each bin is approximately equal. The midpoint of the kth bin is $b_k = (a_{k-1} + a_k)/2$, and an estimate $\hat{f}_{Y|X}(\cdot, b_k)$ of the conditional distribution $f_{Y|X}(\cdot, b_k)$ is obtained based on all observations in the kth bin. The left panel of Figure 4 shows a heat map of the conditional distribution $Y|X$; the distributions of $Y|X = x$ come across as roughly Gaussian, with increasing mean and decreasing variance as age increases. The right panel of Figure 4 compares the Wasserstein-Fréchet Integral with the classical L^2 marginal of Y, where smoothed density functions are obtained by applying a local linear smoother to the histograms, analogous to the previous data illustration in Section 3.2. It is clear from the figure that the Wasserstein-Fréchet Integral is unimodal, while the classical L^2 marginal is flat and uninformative.

In the two-dimensional case, $Y = (Y_1, Y_2)^T$ is the response with Y_1 represent-

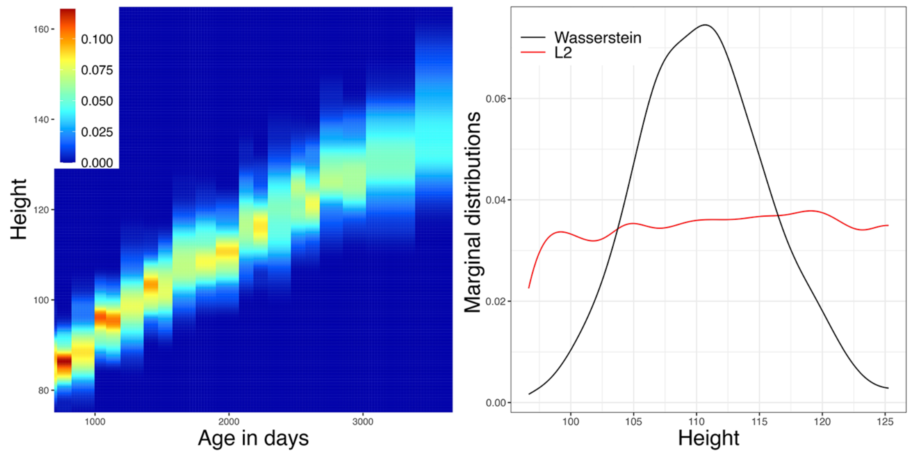

Fig 4. *The left panel displays a heat map of $f_{Y|X}$, where X represents the age of the child in days and Y corresponds to height. The right panel shows the estimated densities of the Wasserstein-Fréchet Integral and classical marginal distributions of height.*

ing the weight and Y_2 representing the height. We utilize the `wasserstein_bary` function from the `WSGeometry` package (Heinemann and Bonneel, 2021) to compute the Wasserstein barycenter in the 2D case. Figure 5 displays heat maps of the conditional distributions of $Y|X = x$ at six randomly selected values of x. Similar to the one-dimensional case, the conditional distribution $Y|X = x$ is unimodal and approximately symmetric around the center. Figure 6 shows that the Wasserstein-Fréchet Integral is also unimodal and resembles a two-dimensional Gaussian shape, while the classical L^2 marginal is ridge-like and can provide misleading information about Y.

7. Discussion

The proposed approach is inspired by the representation of a marginal distributions as an integral of a conditional distribution. We found that this representation of a marginal distribution is more flexible than it seems at first glance once it is interpreted as a Fréchet integral with respect to the L^2 metric. As we demonstrate, replacing the L^2 metric by a different metric in the space of distributions leads to alternative integrated conditional distributions. If the 2-Wasserstein metric is substituted for the L^2 metric one obtains the Wasserstein-Fréchet Integral. The Wasserstein-Fréchet Integral has several attractive properties in terms of invariance for various distributional families which are reproduced under the Wasserstein-Fréchet Integral operation and can reflect the error distribution in regression settings. The latter application of the Wasserstein-Fréchet Integral is of immediate relevance for statistical practice.

The proposed method is not directly comparable with quantile regression where the conditional distribution given a covariate level is the target. In con-

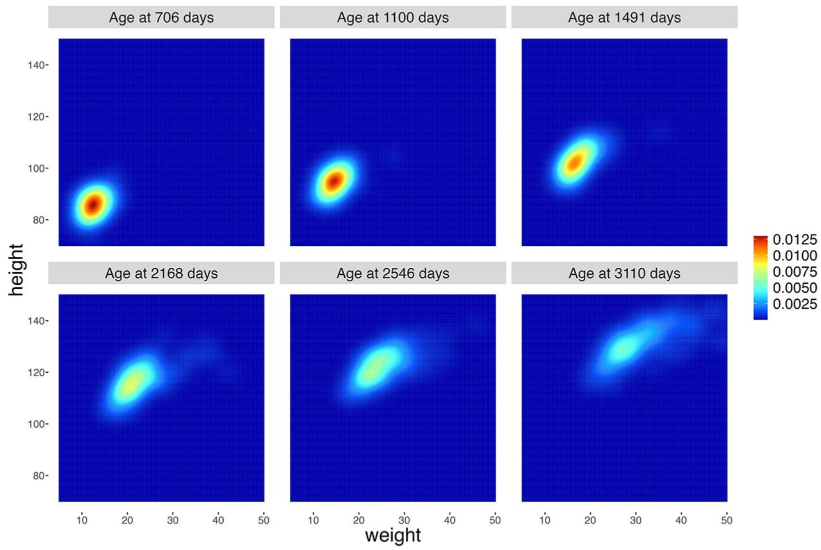

FIG 5. *The heat maps of $f_{Y|X}$, where $Y = (Y_1, Y_2)^T$ represents weight and height. Each figure corresponds to the heat map of $f_{Y|X}$ for a specific value of $X = x$.*

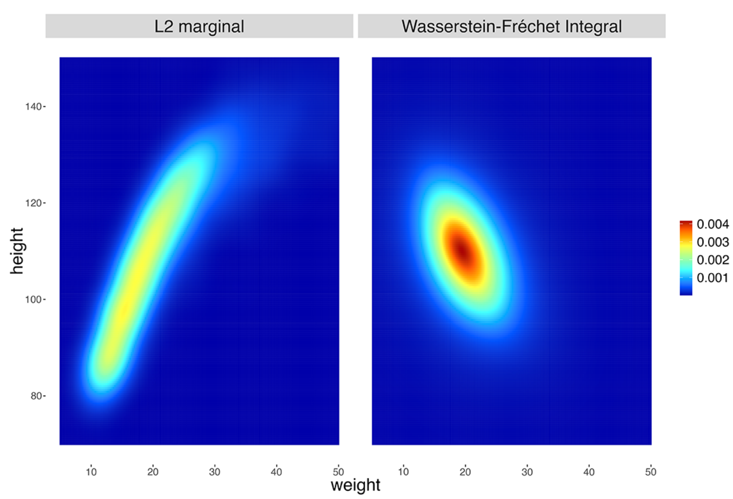

FIG 6. *Heat maps of the densities of the classical L^2 marginal (left panel) and the Wasserstein-Fréchet Integral (right panel) of the joint distribution of weight and height for the ECHO study.*

trast to quantile regression, the proposed method does not require specifying a mean model or imposing any particular structure on the quantiles. Such assumptions are inherent in quantile regression, as this method relies on M-estimation and necessitates parametric specifications (Koenker and Hallock, 2001) and often also further modifications, e.g., to address the crossing problem for quantiles. The Wasserstein-Fréchet Integral and Fréchet integrals based on other metrics provide a general framework for studying the average distribution of a response across all covariates and thus serves as a tool in downstream analyses. For example, a direct application is using the proposed Wasserstein-Fréchet Integral as a reference measure to examine the influence of the covariate X on the distribution of Y. Specifically, one can employ optimal transport from the Wasserstein-Fréchet Integral to each conditional measure $Y|X = x$ as the response and perform Fréchet regression (Petersen and Müller, 2019a) of these responses with predictor x. The optimal transport method serves to center random elements in the Wasserstein space, which is nonlinear and lacks a subtraction operator. In this context, the Wasserstein-Fréchet Integral takes on the role of an intrinsic mean, for which traditional marginal distributions are not suited.

An extension that will be left for future research concerns the case of more general metric objects, including the case of multivariate distributions. Current approaches to obtain distributional results for such data include Tukey's depth (Dai and Lopez-Pintado, 2023) or transport-based definitions of quantiles (del Barrio, González-Sanz and Hallin, 2020; Hallin et al., 2021; del Barrio, Sanz and Hallin, 2024). A specific extension where the Wasserstein-Fréchet Integral may prove useful is slicing in Wasserstein spaces of multivariate distributions (Kolouri, Rohde and Hoffmann, 2018; Chen and Müller, 2023).

Appendix A: Proofs of theorems and ancillary results

A.1. Proof of Propositions 1 and 2

In what follows, write $D(t,x) = \frac{\partial}{\partial t} Q_{Y|X}(t,x)$ and $g_{pq}(y,x) = \frac{\partial^{p+q}}{\partial y^p \partial x^q} F_{Y,X}(y,x)$, $p, q = 0, 1, 2$ with $0 \le p+q \le 2$, and recall that

$$\hat{f}(x) = \sum_{i=1}^{n} \kappa_2 \left(\frac{X_i - x}{h} \right) w(x,h) \Big/ \int_0^1 \kappa_2 \left(\frac{X_i - s}{h} \right) w(s,h) ds.$$

Proof of Proposition 1. It suffices to show that the density of $\nu^*_{L^2}$ is $f_Y = E(f_{Y|X}(\cdot, X))$. By construction, we have

$$\nu^*_{L^2} = \operatorname*{arg\,inf}_{w \in \mathcal{F}_d} E\left(\|f_{Y|X}(\cdot, x) - g_w\|^2_{L^2} \right),$$

where $\|\cdot\|_{L^2}$ denotes the L^2 norm and g_w is the density function corresponding to $w \in \mathcal{F}_d$. By properties of the L^2 inner product, the fact that $f_Y = E(f_{Y|X}(\cdot, X))$ and using similar arguments as in the proof of Proposition 1 in Petersen and Müller (2019a),

$$\nu^*_{L^2} = \operatorname*{arg\,inf}_{w \in \mathcal{F}_d} \|f_Y - g_w\|^2_{L^2},$$

which implies the result. □

Proof of Proposition 2. If $\mathcal{Y} = [0,1]$, the result follows from Theorem 3.1 in Bigot and Klein (2018). Suppose that $\mathcal{Y} = \mathbb{R}$. Since $|Q_{Y|X}(t,X)| \leq g(t)$ with $g \in L^2([0,1]) \subset L^1([0,1])$ and $Q_{Y|X}(\cdot, X)$ is non-decreasing and left-continuous, by the Lebesgue dominated convergence theorem these properties are also shared by $Q^*(t) = E(Q_{Y|X}(t,X))$. Also note that $\nu(x) \in \mathcal{P}_2(\mathcal{Y})(\mathcal{Y})$ since for a random variable R_x with probability distribution $\nu(x)$ it holds that

$$E(R_x^2) = \int_0^1 [Q_{Y|X}(t,x)]^2 dt \leq \int_0^1 g^2(t)dt < \infty.$$

From arguments in the proof of Theorem 3.1 in Bigot and Klein (2018) along with Proposition A.2 in Bobkov and Ledoux (2019) and observing that

$$E(Q^2_{Y|X}(t,X)) \leq g^2(t) < \infty,$$

it follows that there exists a unique probability measure ν^* with cdf F^* such that $Q^* = F^{*-1}$, where F^{*-1} is the left-continuous generalized inverse. For any probability measure $w \in \mathcal{P}_2(\mathcal{Y})$

$$\int_{\mathcal{X}} d^2(\nu(x), w) f(x)dx \geq \int_{\mathcal{X}} d^2(\nu(x), \nu^*) f(x)dx,$$

where ν^* has quantile function Q^* and $\nu^* \in \mathcal{P}_2(\mathcal{Y})$ since

$$\int_0^1 [Q^*(t)]^2 dt \leq \int_0^1 g^2(t)dt < \infty,$$

which is due to $g \in L^2([0,1])$. From Lemma 3.2.1 in Pass (2013) it follows that the application $\int_{\mathcal{X}} d^2(\nu(x), w) f(x)dx$ over $w \in \mathcal{P}_2(\mathcal{Y})$ is strictly convex, which implies that ν^* is the unique solution to (3). □

A.2. Proof of theoretical results when the conditional distribution has compact support

Recall that $h_1, h < 0.5$ and

$$\hat{f}_{Y|X}(y,x) = \frac{\sum_{i=1}^n \kappa_1\left(\frac{Y_i - y}{h_1}\right) w_1(y,h_1) \kappa_2\left(\frac{X_i - x}{h}\right) w(x,h)}{\sum_{i=1}^n \kappa_2\left(\frac{X_i - x}{h}\right) w(x,h) \int_0^1 \kappa_1\left(\frac{Y_i - s}{h_1}\right) w_1(s,h_1) ds},$$

where w_1 is defined analogously as w but replacing the kernel κ_2 by κ_1. We require the following auxiliary lemmas.

Lemma A.1. *Suppose that $\mathcal{Y} = [0,1]$, $\mathcal{X} = [0,1]$, $\kappa_1, \kappa_2 \in \mathcal{K}$ and the regularity conditions (S1) and (S2) hold. Then*

$$E\left[(N(y,x) - f(x)F_{Y|X}(y,x))^2\right] = O(h_1^2 + h^2 + (nh)^{-1}),$$

where

$$N(y,x) = (nh)^{-1} \sum_{i=1}^{n} \kappa_2 \left(\frac{X_i - x}{h} \right) w(x,h) \frac{1}{h_1} \int_0^y \kappa_1 \left(\frac{Y_i - s}{h_1} \right) w_1(s,h_1) ds$$

and the bound is uniform in y and x.

Proof of Lemma A.1. Fubini's theorem implies

$$\begin{aligned} E(N(y,x)) &= \frac{1}{h} \int_{\mathcal{X}} \kappa_2 \left(\frac{r-x}{h} \right) w(x,h) \int_0^y w_1(s,h_1) \\ &\qquad \times \int_{\mathcal{Y}} \frac{1}{h_1} \kappa_1 \left(\frac{u-s}{h_1} \right) f_{Y,X}(u,r) du ds dr \\ &= \frac{1}{h} \int_{\mathcal{X}} \kappa_2 \left(\frac{r-x}{h} \right) w(x,h) \int_0^y w_1(s,h_1) \\ &\qquad \times \int_{-s/h_1}^{(1-s)/h_1} \kappa_1(\nu) f_{Y,X}(s+\nu h_1, r) d\nu ds dr. \end{aligned}$$

By a Taylor expansion, $f_{Y,X}(s+\nu h_1, r) = f_{Y,X}(s,r) + g_{21}(\xi_1, r)\nu h_1$ and

$$f_{Y,X}(s, x+uh) = f_{Y,X}(s,x) + g_{12}(s,\xi_2) uh,$$

where $\xi_1 = \xi_1(s,\nu,h_1,r)$ is between s and $s+\nu h_1$, and $\xi_2 = \xi_2(s,x,u,h)$ is between x and $x+uh$. Since $\int_{-s/h_1}^{(1-s)/h_1} \kappa_1(u) w_1(s,h_1) du = 1$,

$$\int_{-x/h}^{(1-x)/h} \kappa_2(u) w(x,h) du = 1$$

and noting that $f(x)F_{Y|X}(y,x) = \int_0^y f_{Y,X}(s,x)ds$, we obtain

$$\begin{aligned} E(N(y,x)) &= f(x)F_{Y|X}(y,x) + h \int_0^y \int_{-x/h}^{(1-x)/h} u\kappa_2(u) w(x,h) g_{12}(s,\xi_2) du ds \\ &+ \frac{h_1}{h} \int_{\mathcal{X}} \int_0^y \int_{-s/h_1}^{(1-s)/h_1} \kappa_2 \left(\frac{r-x}{h} \right) w(x,h) w_1(s,h_1) \nu \kappa_1(\nu) g_{21}(\xi_1, r) d\nu ds dr. \end{aligned}$$

This along with $\|g_{12}\|_\infty < \infty$ and $\|g_{21}\|_\infty < \infty$, which follows from (S1), (S2), the compactness of $\mathcal{Y}$ and $\mathcal{X}$, $w(x,h) \leq (\int_0^1 \kappa_2(u)du)^{-1}$ and $w_1(s,h_1) \leq (\int_0^1 \kappa_1(u)du)^{-1}$ leads to

$$\begin{aligned} |E(N(y,x)) - f(x)F_{Y|X}(y,x)| &\leq h\|g_{12}\|_\infty \left(\int_0^1 \kappa_2(u)du \right)^{-1} \int_{\mathbb{R}} |u|\kappa_2(u)du \\ &\qquad + h_1 \|g_{21}\|_\infty \left(\int_0^1 \kappa_1(u)du \right)^{-1} \int_{\mathbb{R}} |\nu|\kappa_1(\nu)d\nu. \end{aligned}$$

Thus

$$|E(N(y,x)) - f(x)F_{Y|X}(y,x)| = O(h + h_1), \tag{7}$$

where the bound is uniform in y and x. Similarly, using that

$$\frac{1}{h_1}\int_0^y \kappa_1\left(\frac{Y_i - s}{h_1}\right) w_1(s,h_1)ds \le \left(\int_0^1 \kappa_1(u)du\right)^{-1}$$

for all $y \in \mathcal{Y}$ and $\int_{\mathbb{R}} \kappa_2^2(u)du < \infty$, we obtain

$$\mathrm{Var}(N(y,x)) \le O((nh)^{-1}),$$

where the bound is uniform in y and x. Combining with (7) leads to the result. □

Lemma A.2. *Under the conditions of Lemma A.1 and taking $h = h_n = n^{-1/3}$, it holds that*

$$Z_n := \int_{\mathcal{X}}\int_{\mathcal{Y}} (\hat{F}_{Y|X}(y,x) - F_{Y|X}(y,x))^2 f_{Y|X}(y,x)dydx = O_p(h_1^2 + n^{-2/3}).$$

Proof of Lemma A.2. Define auxiliary quantities

$$N(y,x) = (nh)^{-1}\sum_{i=1}^n \kappa_2\left(\frac{X_i - x}{h}\right) w(x,h)\frac{1}{h_1}\int_0^y \kappa_1\left(\frac{Y_i - s}{h_1}\right) w_1(s,h_1)ds$$

and $f_0(x) = (nh)^{-1}\sum_{i=1}^n \kappa_2\left(\frac{X_i - x}{h}\right) w(x,h)$. Note that $w_1(s,h_1) \ge 1$ for all $s \in [0,1]$. By a change of variables (see for example the proof of Proposition 1 in Petersen and Müller (2016b))

$$\int_0^1 \frac{1}{h_1}\kappa_1\left(\frac{Y_i - s}{h_1}\right) ds = \int_{-Y_i/h_1}^{(1-Y_i)/h_1} \kappa_1(\nu)d\nu \ge \inf_{s\in[0,1]} \int_{-s/h_1}^{(1-s)/h_1} \kappa_1(\nu)d\nu$$
$$\ge \int_0^1 \kappa_1(\nu)d\nu.$$

Thus

$$N(1,x) \ge \left(\int_0^1 \kappa_1(\nu)d\nu\right) f_0(x).$$

From the proof of Proposition 1 in Petersen and Müller (2016b), we have $\|f_0 - f\|_\infty = O_p(n^{-(1/6-\epsilon_0)})$, for any fixed $\epsilon_0 \in (0,1/6)$. Let $\epsilon \in (0,M)$ with $M > 0$ as in (S2). Since $f_0(x) \ge f(x) - \|f_0 - f\|_\infty$, with probability tending to one as $n \to \infty$, it holds that $\inf_{s\in[0,1]} f_0(s) \ge M - \epsilon > 0$ so that $\sup_{s\in[0,1]}[f_0(s)]^{-1} \le (M-\epsilon)^{-1}$. For the remainder of the proof we work conditional on this event. Writing $c_1 = \int_0^1 \kappa_1(\nu)d\nu > 0$,

$$|\hat{F}_{Y|X}(y,x) - F_{Y|X}(y,x)|$$

$$
\begin{aligned}
&= |N(y,x)/N(1,x) - F_{Y|X}(y,x)| \\
&\le c_1^{-1}|N(y,x) - N(1,x)F_{Y|X}(y,x)|/f_0(x) \\
&\le c_1^{-1}(M-\epsilon)^{-1}\left(|N(y,x) - f(x)F_{Y|X}(y,x)| + |f(x) - N(1,x)|F_{Y|X}(y,x)\right).
\end{aligned}
$$

Thus

$$
\begin{aligned}
Z_n \le\ & 2((M-\epsilon)c_1)^{-2}\Big[\int_{\mathcal{X}}\int_{\mathcal{Y}}(N(y,x) - f(x)F_{Y|X}(y,x))^2 f_{Y|X}(y,x)dydx \\
&+ \int_{\mathcal{X}}\int_{\mathcal{Y}}(N(1,x) - f(x))^2 F_{Y|X}^2(y,x)f_{Y|X}(y,x)dydx\Big] \\
=\ & O_p(h_1^2 + h^2 + (nh)^{-1}),
\end{aligned}
$$

where the last equality is due to Lemma A.1 and $0 \le F_{Y|X}(y,x) \le 1$ with $F_{Y|X}(1,x) = 1$. The result follows. □

Proof of Theorem 3.1. Recall that $\hat{Q}^*(t) = \int_{\mathcal{X}} \hat{Q}_{Y|X}(t,x)\hat{f}(x)dx$, $t \in (0,1)$ is the quantile function of the estimated Wasserstein-Fréchet Integral measure $\hat{\nu}^*$ and $Q^*(t) = \int_{\mathcal{X}} Q_{Y|X}(t,x)f(x)dx$ is the true population counterpart. Note that

$$
d_{\mathcal{W}}^2(\hat{\nu}^*, \nu^*) = \int_0^1 (\hat{Q}^*(t) - Q^*(t))^2 dt \tag{8}
$$

$$
= \int_0^1 \left(\int_{\mathcal{X}} \hat{Q}_{Y|X}(t,x)\hat{f}(x) - Q_{Y|X}(t,x)f(x)dx\right)^2 dt, \tag{9}
$$

where

$$
\begin{aligned}
&(\int_{\mathcal{X}} \hat{Q}_{Y|X}(t,x)\hat{f}(x) - Q_{Y|X}(t,x)f(x)dx)^2 \\
\le & 4\|\hat{Q}_{Y|X}(t,\cdot) - Q_{Y|X}(t,\cdot)\|_{L^2(\mathcal{X})}^2 \times (\|\hat{f} - f\|_{L^2(\mathcal{X})}^2 + \|f\|_{L^2(\mathcal{X})}^2) \\
&+ 2\|Q_{Y|X}(t,\cdot)\|_{L^2(\mathcal{X})}^2 \|\hat{f} - f\|_{L^2(\mathcal{X})}^2.
\end{aligned}
$$

Similarly as in the proof of Proposition 1 in Petersen and Müller (2019b), we have from the mean value theorem and Fubini's theorem

$$
\begin{aligned}
&\int_0^1 \|\hat{Q}_{Y|X}(t,\cdot) - Q_{Y|X}(t,\cdot)\|_{L^2(\mathcal{X})}^2 dt \\
&= \int_{\mathcal{X}}\int_0^1 (\hat{Q}_{Y|X}(t,x) - Q_{Y|X}(t,x))^2 dtdx \\
&= \int_{\mathcal{X}}\int_{\hat{Q}_{Y|X}(0,x)}^{\hat{Q}_{Y|X}(1,x)} [D(\xi_{u,x},x)]^2 (\hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x))^2 \hat{f}_{Y|X}(u,x)dudx,
\end{aligned}
$$

where $\xi_{u,x}$ lies between $F_{Y|X}(u,x)$ and $\hat{F}_{Y|X}(u,x)$, and

$$
D(t,x) = 1/f_{Y|X}(Q_{Y|X}(t,x),x)
$$

is the conditional quantile density function. Since $Q_{Y|X}(t,x), \hat{Q}_{Y|X}(t,x) \in \mathcal{Y} = [0,1]$ and $\hat{f}_{Y|X}(u,x) \geq 0$, it follows that

$$\int_0^1 \|\hat{Q}_{Y|X}(t,\cdot) - Q_{Y|X}(t,\cdot)\|^2_{L^2(\mathcal{X})} dt \\ \leq \left(\sup_{(s,\nu)\in[0,1]\times\mathcal{X}} D^2(s,\nu) \right) \int_{\mathcal{X}} \int_0^1 (\hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x))^2 \hat{f}_{Y|X}(u,x) du dx, \tag{10}$$

where $\sup_{s\in[0,1],\nu\in\mathcal{X}} D^2(s,\nu) < \infty$, which is due to the fact that

$$\inf_{(y,x)\in\mathcal{Y}\times\mathcal{X}} f_{Y|X}(y,x) > 0,$$

which in turn follows from (S1), observing that $\mathcal{Y}$ and $\mathcal{X}$ are compact.

Next, using that $\hat{f}_{Y|X}$ is non-negative and $\hat{F}_{Y|X}(\cdot,x)$ is a valid (conditional) cdf with $\hat{F}_{Y|X}(1,x) = 1$ and $\hat{F}_{Y|X}(0,x) = 0$, setting

$$Z_n = \int_{\mathcal{X}} \int_0^1 (\hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x))^2 \hat{f}_{Y|X}(u,x) du dx,$$

one obtains

$$\begin{aligned} Z_n &= \int_{\mathcal{X}} \int_0^1 (\hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x))^2 (\hat{f}_{Y|X}(u,x) - f_{Y|X}(u,x)) du dx \\ &\quad + \int_{\mathcal{X}} \int_0^1 (\hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x))^2 f_{Y|X}(u,x) du dx \\ &= \int_{\mathcal{X}} \int_0^1 (\hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x))^2 f_{Y|X}(u,x) du dx \\ &= O_p(h_1^2 + n^{-2/3}), \end{aligned}$$

where the last equality is due to Lemma A.2. Combining with (9), (10) and $\|\hat{f} - f\|^2_{L^2(\mathcal{X})} = O_p(n^{-2/3})$, which is due to Proposition 1 in Petersen and Müller (2016b), leads to the result. □

A.3. Proof of theoretical results when the conditional distribution has unbounded support

Recall that

$$\hat{f}_{Y|X}(y,x) = \frac{\sum_{i=1}^n \frac{1}{h_1} \kappa_1\left(\frac{Y_i - y}{h_1}\right) \kappa_2\left(\frac{X_i - x}{h}\right) w(x,h)}{\sum_{i=1}^n \kappa_2\left(\frac{X_i - x}{h}\right) w(x,h)},$$

and the conditional cdf $\hat{F}_{Y|X}(y,x) = \int_{-\infty}^y \hat{f}_{Y|X}(s,x) ds$, $y \in \mathbb{R}$, is given by

$$\hat{F}_{Y|X}(y,x) = \sum_{i=1}^n G\left(\frac{y - Y_i}{h_1}\right) \kappa_2\left(\frac{X_i - x}{h}\right) w(x,h) \Big/ \sum_{i=1}^n \kappa_2\left(\frac{X_i - x}{h}\right) w(x,h),$$

where $G(z) = \int_{-\infty}^{z} \kappa_1(s)ds$, $z \in \mathbb{R}$. Denote by $\|g_{pq}\|_\infty = \sup_{(y,x)\in\mathbb{R}\times\mathcal{X}}|g_{pq}(y,x)|$ and $\|f\|_\infty = \sup_{x\in\mathcal{X}}|f(x)|$ the corresponding supremum norms.

Lemma A.3. *Suppose that $\mathcal{Y} = \mathbb{R}$, $\mathcal{X} = [0,1]$, $\kappa_2 \in \mathcal{K}$ and the regularity conditions (S1), (S2), (S2') and (K1') hold. Then*

$$E\left[(N(y,x) - f(x)F_{Y|X}(y,x))^2\right] = O(h_1^2 + h^2 + (nh)^{-1}),$$

where $N(y,x) = (nh)^{-1}\sum_{i=1}^{n} G\left(\frac{y-Y_i}{h_1}\right)\kappa_2\left(\frac{X_i-x}{h}\right)w(x,h)$ and the bound is uniform in y and x.

Proof of Lemma A.3. Integration by parts and (K1') imply $G(z) > 0$ for all $z \in \mathbb{R}$ and

$$\begin{aligned}
E(N(y,x)) &= \frac{1}{h}\int_{\mathcal{X}}\int_{\mathbb{R}} G\left(\frac{y-s}{h_1}\right)\kappa_2\left(\frac{r-x}{h}\right)w(x,h)f_{Y,X}(s,r)dsdr \\
&= h_1\int_{-x/h}^{(1-x)/h}\kappa_2(u)w(x,h)\int_{\mathbb{R}} G(\nu)f_{Y,X}(y-h_1\nu, x+uh)d\nu du \\
&= \int_{-x/h}^{(1-x)/h}\kappa_2(u)w(x,h)\int_{\mathbb{R}}\kappa_1(\nu)g_{01}(y-h_1\nu, x+uh)d\nu du.
\end{aligned}$$

By a Taylor expansion,

$$g_{01}(y-h_1\nu, x+uh) = g_{01}(y,x) + g_{02}(y,\xi_2)uh - h_1 f_{Y,X}(\xi_1, x+uh)\nu,$$

where

$$\xi_2 = \xi_2(x,y,u,\nu,h,h_1)$$

is between x and $x+uh$, and $\xi_1 = \xi_1(x,y,u,\nu,h,h_1)$ is between y and $y-\nu h_1$. With $\|g_{02}\|_\infty < \infty$ and $\|f_{Y,X}\|_\infty < \infty$, as implied by (S2) and (S2'), using that $\int_{-x/h}^{(1-x)/h}\kappa_2(u)w(x,h)du = 1$, one obtains

$$\begin{aligned}
&|E(N(y,x)) - g_{01}(y,x)| \\
&\leq h\int_{-x/h}^{(1-x)/h}|u|\kappa_2(u)w(x,h)|g_{02}(y,\xi_2)|du \\
&\quad + h_1\int_{-x/h}^{(1-x)/h}\kappa_2(u)w(x,h)\int_{\mathbb{R}}|\nu|\kappa_1(\nu)f_{Y,X}(\xi_1, x+uh)d\nu \\
&= O(h+h_1).
\end{aligned}$$

Thus

$$|E(N(y,x)) - g_{01}(y,x)| = O(h+h_1), \tag{11}$$

where the bound is uniform in y and x. Similarly, and using that $G(s) \leq 1$ for any $s \in \mathbb{R}$, we obtain

$$\mathrm{Var}(N(y,x)) \leq \frac{1}{nh^2}E\left(G^2\left(\frac{y-Y_i}{h_1}\right)\kappa_2^2\left(\frac{X_i-x}{h}\right)w^2(x,h)\right) = O((nh)^{-1}),$$

where the bound is uniform in y and x. This along with (11) and noting that $f(x)F_{Y|X}(y,x) = g_{01}(y,x)$ leads to the result. $\square$

Lemma A.4. *Suppose that the conditions of Lemma A.3 hold. If $h = h_n = n^{-1/3}$, then*

$$Z_n := \int_{\mathcal{X}} \int_{\mathbb{R}} (\hat{F}_{Y|X}(y,x) - F_{Y|X}(y,x))^2 f_{Y|X}(y,x) dy dx = O_p(h_1^2 + n^{-2/3}).$$

Proof of Lemma A.4. Let $f_0(x) = (nh)^{-1} \sum_{i=1}^{n} \kappa_2 \left(\frac{X_i - x}{h}\right) w(x,h)$ and $0 < \epsilon < M$ with M as in (S2) and $\epsilon_0 \in (0, 1/3)$. From the proof of Lemma A.2, we have $\|f_0 - f\|_\infty = O_p(n^{-(1/6 - \epsilon_0/2)})$ and also that $\sup_{s \in [0,1]} [f_0(s)]^{-1} \leq (M - \epsilon)^{-1}$ holds with probability tending to 1 as $n \to \infty$. The remainder of the proof is conditional on this event. Defining $N(y,x)$ as in Lemma A.3, we have

$$\begin{aligned}
&|\hat{F}_{Y|X}(y,x) - F_{Y|X}(y,x)| \\
&= |N(y,x)/f_0(x) - F_{Y|X}(y,x)| \\
&\leq |N(y,x) - f(x)F_{Y|X}(y,x)|/f_0(x) + |f(x) - f_0(x)| F_{Y|X}(y,x)/f_0(x) \\
&\leq (M-\epsilon)^{-1} |N(y,x) - f(x)F_{Y|X}(y,x)| + (M-\epsilon)^{-1} |f(x) - f_0(x)| F_{Y|X}(y,x),
\end{aligned}$$

and

$$\begin{aligned}
Z_n &\leq \frac{2}{(M-\epsilon)^2} \int_{\mathcal{X}} \int_{\mathbb{R}} [N(y,x) - f(x)F_{Y|X}(y,x)]^2 f_{Y|X}(y,x) dy dx \\
&\quad + \frac{2}{(M-\epsilon)^2} \int_{\mathcal{X}} \int_{\mathbb{R}} F_{Y|X}^2(y,x) f_{Y|X}(y,x) (f(x) - f_0(x))^2 dy dx \\
&\leq \frac{2}{M(M-\epsilon)^2} \int_{\mathcal{X}} \int_{\mathbb{R}} [N(y,x) - f(x)F_{Y|X}(y,x)]^2 f_{Y,X}(y,x) dy dx \\
&\quad + \frac{2}{(M-\epsilon)^2} \|f_0 - f\|_{L^2(\mathcal{X})}^2. \qquad (12)
\end{aligned}$$

From Lemma A.3 we obtain

$$E\left(\int_{\mathcal{X}} \int_{\mathbb{R}} [N(y,x) - f(x)F_{Y|X}(y,x)]^2 f_{Y,X}(y,x) dy dx \right) = O(h_1^2 + h^2 + (nh)^{-1}), \qquad (13)$$

where the bound is uniform in y and x. Also, from the proof of Theorem 3.1, we have $\|\hat{f} - f\|_{L^2(\mathcal{X})}^2 = O_p(n^{-2/3})$. With (12) and (13) this implies the result. $\square$

Lemma A.5. *Suppose that the conditions of Lemma A.3 hold. Let $C > 0$ and assume $n^{-p} h^{-1}(h_1^{-1} + h^{-1}) = o(1)$ for some $p > 0$. If $nh/\log n \to \infty$ as $n \to \infty$, then*

$$\sup_{(x,y) \in \mathcal{X} \times [-C,C]} |N(y,x) - f(x)F_{Y|X}(y,x)| = o_p(1),$$

where $N(y,x) = (nh)^{-1} \sum_{i=1}^{n} G\left(\frac{y - Y_i}{h_1}\right) \kappa_2 \left(\frac{X_i - x}{h}\right) w(x,h)$.

Proof of Lemma A.5. Let $\mathcal{Y}_n$ and $\mathcal{X}_n$ be equispaced grids on $[-C, C]$ and $[0, 1]$ with spacing $n^{-\gamma}$, where $\gamma > 0$. Then

$$\begin{aligned}
&\sup_{(x,y)\in\mathcal{X}\times[-C,C]} |N(y,x) - f(x)F_{Y|X}(y,x)| \qquad (14)\\
\leq &\sup_{y\in\mathcal{Y}_n, x\in\mathcal{X}_n} |N(y,x) - E[N(y,x)]| + \sup_{|y-y_1|,|x-x_1|\leq n^{-\gamma}} |N(y,x) - N(y_1,x_1)|\\
&+ \sup_{|y-y_1|,|x-x_1|\leq n^{-\gamma}} |E[N(y,x)] - E[N(y_1,x_1)]|\\
&+ \sup_{y\in[-C,C], x\in\mathcal{X}} |E[N(y,x)] - f(x)F_{Y|X}(y,x)|. \qquad (15)
\end{aligned}$$

To control the first term on the right hand side, we follow similar arguments as in the proof of Lemma 2 in Zhang and Wang (2016). Let

$$U_i = G\left(\frac{y - Y_i}{h_1}\right)\kappa_2\left(\frac{X_i - x}{h}\right)w(x,h),$$

$\mu_i = E(U_i)$, $i = 1, \ldots, n$, $M > 0$ and $a_n = (\log n/(nh))^{1/2}$. By independence of the U_i and using Chernoff's bound, we have for large enough n

$$\begin{aligned}
&P\left(N(y,x) - E[N(y,x)] > Ma_n\right)\\
=&P\left(a_n\sum_{i=1}^n (U_i - \mu_i) > M\log n\right)\\
=&n^{-M}\prod_{i=1}^n E\left[\exp(a_n(U_i - \mu_i))\right]\\
\leq&n^{-M}\prod_{i=1}^n [1 + a_n^2 E(U_i^2)]\\
\leq&n^{-M}\prod_{i=1}^n [1 + a_n^2 C_1 h]\\
\leq&n^{-M}\prod_{i=1}^n \exp(C_1 a_n^2 h)\\
=&n^{-M}\exp(C_1 a_n^2 nh),
\end{aligned}$$

where the first inequality is due to $|U_i - \mu_i| \leq C_0$ for some constant $C_0 < \infty$ along with $a_n \to 0$ as $n \to \infty$ and the fact that $e^u \leq 1 + u + u^2$ holds for u in a small enough neighborhood around 0. The second inequality is due to $E(U_i^2) \leq hC_1$ for some constant $C_1 > 0$, and the last inequality is clear. Here C_0 and C_1 can be shown to be uniform in x and y. Thus

$$\begin{aligned}
&P\left(\sup_{y\in\mathcal{Y}_n, x\in\mathcal{X}_n} N(y,x) - E[N(y,x)] > Ma_n\right)\\
\leq&2Cn^{2\gamma}n^{-M}\exp(C_1 a_n^2 nh) = O(n^{C_1+2\gamma-M}),
\end{aligned}$$

where for large enough M we have $n^{C_1+2\gamma-M} = o(1)$ as $n \to \infty$. This shows that

$$\sup_{y\in\mathcal{Y}_n, x\in\mathcal{X}_n} \Big| N(y,x) - E[N(y,x)] \Big| = O_p(a_n). \tag{16}$$

Next, note that

$$\left| G\left(\frac{y-Y_i}{h_1}\right) - G\left(\frac{y_1-Y_i}{h_1}\right) \right| \le \|\kappa_1\|_\infty \frac{|y-y_1|}{h_1}, \tag{17}$$

and let $x, x_1 \in \mathcal{X}$. Observe that if $x, x_1 \in [0,h)$ are such that $|x-x_1| \le n^{-\gamma}$, then

$$|w(x,h) - w(x_1,h)| \le \left(\int_0^1 \kappa_2(u)du\right)^{-2} \|\kappa_2\|_\infty |x-x_1|/h = O(n^{-\gamma}h^{-1}),$$

where the bound is uniform in x and x_1, using $w(s,h) \le \left(\int_0^1 \kappa_2(u)du\right)^{-1}$ for all $s \in \mathcal{X}$. Also, if $x \in [0,h)$ and $x_1 \in [h, 1-h]$ are such that $|x-x_1| \le n^{-\gamma}$, then

$$\begin{aligned}
|w(x,h) - w(x_1,h)| &\le \left(\int_0^1 \kappa_2(u)du\right)^{-1} |1 - \int_{-x/h}^1 \kappa_2(u)du| \\
&\le \left(\int_0^1 \kappa_2(u)du\right)^{-1} \int_{-1}^{-x/h} \kappa_2(u)du \\
&\le \left(\int_0^1 \kappa_2(u)du\right)^{-1} \int_{-1}^{-1+n^{-\gamma}h^{-1}} \kappa_2(u)du \\
&= O(n^{-\gamma}h^{-1}),
\end{aligned}$$

where the bound is uniform in x and x_1, and the last equality is due to $-x/h \le -1 + n^{-\gamma}h^{-1}$. Similarly, if $x \in [h, 1-h]$ and $x_1 \in (1-h, 1]$ are such that $|x-x_1| \le n^{-\gamma}$, then

$$|w(x,h) - w(x_1,h)| = O(n^{-\gamma}h^{-1}),$$

where the bound is uniform in x and x_1. Finally, for n large enough it cannot occur that $x \in [0,h)$ and $x_1 \in (1-h, 1]$ as a necessary condition for this to hold is that $n^{-\gamma} + 2h \ge 1$. Coupling this with the Lipschitz continuity of κ_2 implies that for large enough n

$$\sup_{|x-x_1|\le n^{-\gamma}} \left| \kappa_2\left(\frac{X_i - x}{h}\right) w(x,h) - \kappa_2\left(\frac{X_i - x_1}{h}\right) w(x_1,h) \right| = O(n^{-\gamma}h^{-1}), \tag{18}$$

where the bound is uniform in X_i. Combining (17) and (18) with the uniform bound on $w(s,h)$, $s \in \mathcal{X}$, leads to

$$\sup_{|y-y_1|,|x-x_1|\le n^{-\gamma}} |N(y,x) - N(y_1,x_1)| = O(n^{-\gamma}h^{-1}(h_1^{-1} + h^{-1})),$$

where the bound is non-random. Since $n^{-p}h^{-1}(h_1^{-1}+h^{-1}) = o(1)$ for some $p>0$, for any $\gamma \geq p$

$$\sup_{|y-y_1|,|x-x_1|\leq n^{-\gamma}} |N(y,x) - N(y_1,x_1)| = o_p(1). \tag{19}$$

Observing

$$\begin{aligned}&\sup_{|y-y_1|,|x-x_1|\leq n^{-\gamma}} E|N(y,x) - N(y_1,x_1)|\\ \leq &E\left(\sup_{|y-y_1|,|x-x_1|\leq n^{-\gamma}} |N(y,x) - N(y_1,x_1)|\right)\end{aligned}$$

leads to

$$\sup_{|y-y_1|,|x-x_1|\leq n^{-\gamma}} |E[N(y,x)] - E[N(y_1,x_1)]| = o_p(1). \tag{20}$$

Finally, from the proof of Lemma A.3 we have $|E(N(y,x)) - g_{01}(y,x)| = O(h+h_1)$, where the bound is uniform in y and x. Therefore

$$\sup_{y\in[-C,C],x\in\mathcal{X}} |E[N(y,x)] - f(x)F_{Y|X}(y,x)| = O(h+h_1). \tag{21}$$

Combining (15), (16), (19), (20) and (21) leads to the result. □

Lemma A.6 below shows the uniform convergence in x towards the true conditional quantile $\hat{Q}_{Y|X}(t,x)$ for fixed $t\in(0,1)$. Set

$$N(y,x) = (nh)^{-1}\sum_{i=1}^{n} G\left(\frac{y-Y_i}{h_1}\right)\kappa_2\left(\frac{X_i-x}{h}\right)w(x,h)$$

and $f_0(x) = (nh)^{-1}\sum_{i=1}^{n}\kappa_2\left(\frac{X_i-x}{h}\right)w(x,h)$.

Lemma A.6. *Suppose that the conditions of Lemma A.5 hold. Let $t\in(0,1)$ be fixed and $h=h_n=n^{-1/3}$. Then*

$$\sup_{x\in\mathcal{X}} |\hat{Q}_{Y|X}(t,x) - Q_{Y|X}(t,x)| = o_p(1).$$

Proof of Lemma A.6. Let $\delta>0$ and $\epsilon\in(0,M)$ with M as in (S2). Define auxiliary quantities

$$c_0(t) = \inf_{x\in\mathcal{X}} \inf_{y\in[Q_{Y|X}(t,x),Q_{Y|X}(t,x)+\delta]} f_{Y|X}(y,x),$$

and

$$c_1(t) = \inf_{x\in\mathcal{X}} \inf_{y\in[Q_{Y|X}(t,x)-\delta,Q_{Y|X}(t,x)]} f_{Y|X}(y,x).$$

Note that the regularity conditions (S1) and (S2) along with compactness of $\mathcal{X}$ imply that $Q_{Y|X}(t,\cdot)$ is continuously differentiable and

$$0 < \inf_{x\in\mathcal{X}, y\in[\inf_{s\in\mathcal{X}} Q_{Y|X}(t,s), \sup_{s\in\mathcal{X}} Q_{Y|X}(t,s)+\delta]} f_{Y|X}(y,x) \le c_0(t),$$

as well as

$$0 < \inf_{x\in\mathcal{X}, y\in[\inf_{s\in\mathcal{X}} Q_{Y|X}(t,s)-\delta, \sup_{s\in\mathcal{X}} Q_{Y|X}(t,s)]} f_{Y|X}(y,x) \le c_1(t),$$

which shows that $c_0(t)$ and $c_1(t)$ are strictly positive constants.

By a Taylor expansion, there exists $\xi_1 = \xi_1(t,x)$ between $Q_{Y|X}(t,x)$ and $Q_{Y|X}(t,x)+\delta$ such that $F_{Y|X}(Q_{Y|X}(t,x)+\delta, x) = t + f_{Y|X}(\xi_1, x)\delta \ge t + c_0(t)\delta$. Similarly, $F_{Y|X}(Q_{Y|X}(t,x)-\delta, x) \le t - c_1(t)\delta$. Defining the closed interval $A_\delta(t) = [\inf_{s\in\mathcal{X}} Q_{Y|X}(t,s)-\delta, \sup_{s\in\mathcal{X}} Q_{Y|X}(t,s)+\delta]$, we have that there exists $C_\delta = C_\delta(t) \in (0,\infty)$ such that $A_\delta(t) \subseteq [-C_\delta, C_\delta]$. From the proof of Lemma A.4 and using that $F_{Y|X}(y,x) \le 1$, with probability tending to 1 we have

$$\begin{aligned}|\hat{F}_{Y|X}(y,x) - F_{Y|X}(y,x)| \le (M-\epsilon)^{-1}|N(y,x) - f(x)F_{Y|X}(y,x)| \\ + (M-\epsilon)^{-1}\|f_0 - f\|_\infty,\end{aligned}$$

where $\|f_0 - f\|_\infty = o_p(1)$. Combining this with Lemma A.5 leads to

$$\hat{\Delta}_n := \sup_{x\in\mathcal{X}, y\in[-C_\delta, C_\delta]} |\hat{F}_{Y|X}(y,x) - F_{Y|X}(y,x)| = o_p(1).$$

Let $\varepsilon > 0$ and $\kappa(t) = \min\{c_0(t), c_1(t)\} > 0$. Note that $\hat{\Delta}_n \le \varepsilon\kappa(t)/2$ holds with probability tending to 1. The remainder of the proof is conditional on this event. Observe

$$\begin{aligned}\hat{F}_{Y|X}(Q_{Y|X}(t,x)+\delta, x) &\ge -\hat{\Delta}_n + F_{Y|X}(Q_{Y|X}(t,x)+\delta, x) \\ &\ge -\varepsilon\kappa(t)/2 + t + c_0(t)\delta \\ &\ge t + (\delta - \epsilon/2)\kappa(t).\end{aligned}$$

Choosing $\delta = \varepsilon$ leads to $\hat{F}_{Y|X}(Q_{Y|X}(t,x)+\delta, x) > t$ and then $\hat{Q}_{Y|X}(t,x) \le Q_{Y|X}(t,x)+\delta$. Similarly, $\hat{F}_{Y|X}(Q_{Y|X}(t,x)-\delta, x) \le \hat{\Delta}_n + t - c_1(t)\delta \le t + (\varepsilon/2 - \delta)\kappa(t) < t$ which implies $\hat{Q}_{Y|X}(t,x) \ge Q_{Y|X}(t,x) - \delta$. Therefore $|\hat{Q}_{Y|X}(t,x) - Q_{Y|X}(t,x)| \le \delta = \varepsilon$. Since ε does not depend on x, we obtain with probability tending to 1

$$\sup_{x\in\mathcal{X}} |\hat{Q}_{Y|X}(t,x) - Q_{Y|X}(t,x)| \le \varepsilon,$$

and the result follows. □

Proof of Theorem 3.2. Observe

$$d^2_{\mathcal{W}_\delta}(\hat{\nu}^*, \nu^*) = \int_\delta^{1-\delta} (\hat{Q}^*(t) - Q^*(t))^2 dt$$

$$= \int_{\delta}^{1-\delta} \left(\int_{\mathcal{X}} \hat{Q}_{Y|X}(t,x)\hat{f}(x) - Q_{Y|X}(t,x)f(x)dx \right)^2 dt, \tag{22}$$

where from the proof of Theorem 3.1

$$\begin{aligned} &\left(\int_{\mathcal{X}} \hat{Q}_{Y|X}(t,x)\hat{f}(x) - Q_{Y|X}(t,x)f(x)dx \right)^2 \\ &\leq 4\|\hat{Q}_{Y|X}(t,\cdot) - Q_{Y|X}(t,\cdot)\|^2_{L^2(\mathcal{X})} \left(\|\hat{f} - f\|^2_{L^2(\mathcal{X})} + \|f\|^2_{L^2(\mathcal{X})} \right) \\ &\quad + 2\|Q_{Y|X}(t,\cdot)\|^2_{L^2(\mathcal{X})} \|\hat{f} - f\|^2_{L^2(\mathcal{X})}, \end{aligned}$$

Define $A_\delta = \int_\delta^{1-\delta} \|\hat{Q}_{Y|X}(t,\cdot) - Q_{Y|X}(t,\cdot)\|^2_{L^2(\mathcal{X})} dt$. Similarly as in the proof of Proposition 1 in Petersen and Müller (2019b), we infer from the mean value theorem, the change of variables $u = \hat{Q}_{Y|X}(t,x)$ and Fubini's theorem that

$$\begin{aligned} A_\delta &= \int_{\mathcal{X}} \int_{\delta}^{1-\delta} (\hat{Q}_{Y|X}(t,x) - Q_{Y|X}(t,x))^2 dt dx \\ &= \int_{\mathcal{X}} \int_{\hat{Q}_{Y|X}(\delta,x)}^{\hat{Q}_{Y|X}(1-\delta,x)} [D(\xi_{u,x},x)]^2 (\hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x))^2 \hat{f}_{Y|X}(u,x) du dx, \end{aligned}$$

where $\xi_{u,x}$ lies between $F_{Y|X}(u,x)$ and $\hat{F}_{Y|X}(u,x)$, and

$$D(t,x) = 1/f_{Y|X}(Q_{Y|X}(t,x),x)$$

is the quantile density function. Setting $\hat{\Delta}_n(t) = \sup_s |\hat{Q}_{Y|X}(t,s) - Q_{Y|X}(t,s)|$, $t \in (0,1)$,

$$\hat{Q}_{Y|X}(\delta,x) \geq Q_{Y|X}(\delta,x) - \hat{\Delta}_n(\delta) \geq \inf_{s\in\mathcal{X}} Q_{Y|X}(\delta,s) - \hat{\Delta}_n(\delta), \tag{23}$$

and

$$\hat{Q}_{Y|X}(1-\delta,x) \leq Q_{Y|X}(1-\delta,x) + \hat{\Delta}_n(1-\delta) \tag{24}$$

$$\leq \sup_{s\in\mathcal{X}} Q_{Y|X}(1-\delta,s) + \hat{\Delta}_n(1-\delta). \tag{25}$$

For $\epsilon > 0$ define auxiliary quantities $c_0(\epsilon) = \inf_{s\in\mathcal{X}} Q_{Y|X}(\delta,s) - \epsilon$, $c_1(\epsilon) = \sup_{s\in\mathcal{X}} Q_{Y|X}(1-\delta,s) + \epsilon$, $d_0(\epsilon) = \inf_{s\in\mathcal{X}} F_{Y|X}(c_0(\epsilon),s) - \epsilon$ and

$$d_1(\epsilon) = \sup_{s\in\mathcal{X}} F_{Y|X}(c_1(\epsilon),s) + \epsilon.$$

We note that due to $0 < F_{Y|X}(y,x) < 1$ for all $y \in \mathbb{R}$ and $x \in \mathcal{X}$, the continuity of $F_{Y|X}(y,\cdot)$ and the compactness of X, one has $\inf_{s\in\mathcal{X}} F_{Y|X}(y,s) > 0$ and $\sup_{s\in\mathcal{X}} F_{Y|X}(y,s) < 1$. Similarly, it can be shown that both $\inf_{s\in\mathcal{X}} Q_{Y|X}(t,s)$ and $\sup_{s\in\mathcal{X}} Q_{Y|X}(t,s)$ are achieved for any $t \in (0,1)$. In particular, this shows that the previous quantities are well defined. Next, choose $\epsilon > 0$ such that

$$\epsilon < \min\left(\inf_{\nu\in\mathcal{X}} F_{Y|X}\left(\inf_{s\in\mathcal{X}} Q_{Y|X}(\delta,s) - 1, \nu \right), \right.$$

$$1-\sup_{\nu\in\mathcal{X}} F_{Y|X}\left(\sup_{s\in\mathcal{X}} Q_{Y|X}(1-\delta,s)+1,\nu\right)\Bigg),$$

which implies $\epsilon\in(0,1)$ and $0<d_0(\epsilon),d_1(\epsilon)<1$. From the arguments outlined in the proof of Lemma A.6, it follows that

$$\hat{U}_n(\epsilon):=\sup_{y\in[c_0(\epsilon),c_1(\epsilon)],x\in\mathcal{X}}|\hat{F}_{Y|X}(y,x)-F_{Y|X}(y,x)|=o_p(1),$$

while Lemma A.6 shows that $\hat{\Delta}_n(t)=o_p(1)$, where $t\in\{\delta,1-\delta\}$. Here the conditions of Lemma A.6 are satisfied since $h=h_n=n^{-1/3}$ and the regularity condition $h_1n^{p-1/3}\to\infty$ as $n\to\infty$ for some $p>2/3$ implies $n^{-p}h^{-1}(h_1^{-1}+h^{-1})=o(1)$. Thus, the event where $\max\{\hat{U}_n(\epsilon),\hat{\Delta}_n(\delta),\hat{\Delta}_n(1-\delta)\}\le\epsilon$ occurs with probability tending to 1 as $n\to\infty$, and therefore it suffices to work on this event in what follows.

Next, using (23) and (25) along with $\max\{\hat{\Delta}_n(\delta),\hat{\Delta}_n(1-\delta)\}\le\epsilon$ shows that $c_0(\epsilon)\le\hat{Q}_{Y|X}(\delta,x)$ and $\hat{Q}_{Y|X}(1-\delta,x)\le c_1(\epsilon)$. Hence $u\in[\hat{Q}_{Y|X}(\delta,x),\hat{Q}_{Y|X}(1-\delta,x)]$ implies $u\in[c_0(\epsilon),c_1(\epsilon)]$. Also, by monotonicity

$$F_{Y|X}(c_0(\epsilon),x)-\epsilon\le F_{Y|X}(u,x)-\epsilon\le F_{Y|X}(u,x)-\hat{U}_n(\epsilon)\le\hat{F}_{Y|X}(u,x),$$

and thus $d_0(\epsilon)\le\hat{F}_{Y|X}(u,x)$. Similarly, $\hat{F}_{Y|X}(u,x)\le d_1(\epsilon)$. This shows that $\xi_{u,x}\in[d_0(\epsilon),d_1(\epsilon)]$ and hence

$$D(\xi_{u,x},x)\le\left(\inf_{y\in A(\epsilon),s\in\mathcal{X}} f_{Y|X}(y,s)\right)^{-1}<\infty,$$

where $A(\epsilon)=[\inf_{s\in\mathcal{X}}Q_{Y|X}(d_0(\epsilon),s),\sup_{s\in\mathcal{X}}Q_{Y|X}(d_1(\epsilon),s)]$, and the last inequality is due to the fact that $f_{Y|X}(y,s)$ is pointwise strictly positive and also continuous over the compact set $A(\epsilon)\times\mathcal{X}\subset\mathbb{R}^2$. Setting

$$M_\epsilon=(\inf_{y\in A(\epsilon),s\in\mathcal{X}} f_{Y|X}(y,s))^{-1},$$

we obtain

$$\begin{aligned}A_\delta&\le M_\epsilon^2\int_{\mathcal{X}}\int_{c_0(\epsilon)}^{c_1(\epsilon)}(\hat{F}_{Y|X}(u,x)-F_{Y|X}(u,x))^2\hat{f}_{Y|X}(u,x)dudx\\&\le M_\epsilon^2\int_{\mathcal{X}}\int_{\mathbb{R}}(\hat{F}_{Y|X}(u,x)-F_{Y|X}(u,x))^2\hat{f}_{Y|X}(u,x)dudx\\&=M_\epsilon^2 Z_n,\end{aligned}$$

where the last inequality is due to the fact that $\hat{f}_{Y|X}(u,x)$ is non-negative and Z_n is defined through the last equation. Hence

$$\int_\delta^{1-\delta}\|\hat{Q}_{Y|X}(t,\cdot)-Q_{Y|X}(t,\cdot)\|^2_{L^2(\mathcal{X})}dt\le M_\epsilon^2 Z_n.$$

Since $\hat{f}_{Y|X}$ is non-negative and both $\hat{F}_{Y|X}(\cdot, x)$ and $F_{Y|X}(\cdot, x)$ are valid cdfs, it follows that

$$\begin{aligned}Z_n &= \int_{\mathcal{X}} \int_{\mathbb{R}} (\hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x))^2 (\hat{f}_{Y|X}(u,x) - f_{Y|X}(u,x)) du dx \\ &\quad + \int_{\mathcal{X}} \int_{\mathbb{R}} (\hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x))^2 f_{Y|X}(u,x) du dx \\ &= \int_{\mathcal{X}} \int_{\mathbb{R}} (\hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x))^2 f_{Y|X}(u,x) du dx \\ &= O_p(h_1^2 + n^{-2/3}),\end{aligned}$$

where the second equality is due to the change of variables $\nu = \hat{F}_{Y|X}(u,x) - F_{Y|X}(u,x)$ and the last is due to Lemma A.4. Hence

$$\int_{\delta}^{1-\delta} \|\hat{Q}_{Y|X}(t,\cdot) - Q_{Y|X}(t,\cdot)\|_{L^2(\mathcal{X})}^2 dt = O_p(h_1^2 + n^{-2/3}).$$

Combining with (22) and using that

$$\int_{\delta}^{1-\delta} \|Q_{Y|X}(t,\cdot)\|_{L^2(\mathcal{X})}^2 dt = \int_{\mathcal{X}} \int_{\delta}^{1-\delta} Q_{Y|X}^2(t,x) dt dx < \infty,$$

which is due to (S1') along with Fubini's theorem, and the fact that $\|\hat{f} - f\|_{L^2(\mathcal{X})}^2 = O_p(n^{-2/3})$, which is due to Proposition 1 in Petersen and Müller (2016b), then leads to the result. □

A.4. Proof of theoretical results when the densities are fully observed

Recall that Q_i is the quantile function corresponding to f_i and that when the densities $f_i(\cdot) = f_{Y|X}(\cdot, X_i)$ are fully observed, the Q_i form an i.i.d. sample of random quantile functions with $E(Q_1(t)) = Q^*(t)$. A natural empirical estimate of Q^* is then given by

$$\hat{Q}^*(t) = n^{-1} \sum_{i=1}^{n} Q_i(t), \quad t \in (0,1).$$

Proof of Theorem 3.3. Since the Q_i are i.i.d. random quantile functions, we have $E(\hat{Q}^*(t)) = E(Q_1(t)) = Q^*(t)$, $t \in (0,1)$, and

$$\mathrm{Var}(\hat{Q}^*(t)) = n^{-1}\,\mathrm{Var}(Q_1(t)) = O(n^{-1}),$$

where the bound is uniform in t and the last equality is due to $0 \le Q_1(t) = Q_{Y|X}(t, X_1) \le \sup_{x\in\mathcal{X}} Q_{Y|X}(1,x) < \infty$, which in turn follows from the fact that $f_{Y|X}(\cdot, x)$ has compact support $\mathcal{Y} = [0,1]$. These observations along with Fubini's theorem imply

$$E\left(\int_0^1 (\hat{Q}^*(t) - Q^*(t))^2 dt\right) = O(n^{-1}),$$

and the result follows. □

Consider the optimal transport T_i from the density f_i to the Wasserstein-Fréchet Integral density f^*. From the closed form solution of the 2-Wasserstein distance in the one-dimensional case (Villani, 2003), we have $T_i(y) = Q^*(F_i(y))$, $y \in \mathcal{Y}$. A natural estimate of the optimal transport T_i is then achieved by replacing Q^* with its estimated counterpart, as $F_i(y) = \int_0^y f_i(s)ds$ is assumed to be fully observed.

Proof of Theorem 4.1. Note that for $i \in \{1, \dots, n\}$

$$\sup_{y\in\mathcal{Y}} |\hat{T}_i^*(y) - T_i^*(y)| = \sup_{y\in\mathcal{Y}} |\hat{Q}^*(F_i(y)) - Q^*(F_i(y))| \le \sup_{t\in[0,1]} |\hat{Q}^*(t) - Q^*(t)|,$$

and thus

$$\max_{i=1,\dots,n} \sup_{y\in\mathcal{Y}} |\hat{T}_i^*(y) - T_i^*(y)| \le \sup_{t\in[0,1]} |\hat{Q}^*(t) - Q^*(t)|. \tag{26}$$

Let $\mathcal{T}_n$ be an equispaced partition of $\mathcal{Y} = [0, 1]$ with mesh size $\delta_n = n^{-\gamma}$, where $\gamma > 0$. Using that $\hat{Q}^*(t) = n^{-1} \sum_{i=1}^n Q_i(t)$ and $Q^*(t) = E(Q_1(t))$, we obtain

$$\sup_{t\in[0,1]} |\hat{Q}^*(t) - Q^*(t)| \le \sup_{t\in\mathcal{T}_n} \left| n^{-1} \sum_{i=1}^n [Q_i(t) - E(Q_1(t))] \right| \tag{27}$$

$$+ \sup_{|t-s|\le\delta_n} \left| n^{-1} \sum_{i=1}^n [Q_i(t) - Q_i(s)] \right|$$

$$+ \sup_{|t-s|\le\delta_n} \Big| E(Q_1(t)) - E(Q_1(s)) \Big|. \tag{28}$$

Since $|Q_i(t) - Q_i(s)| \le L|t - s|$, we have

$$\sup_{|t-s|\le\delta_n} \Big| E(Q_1(t)) - E(Q_1(s)) \Big| \le L\delta_n \tag{29}$$

and

$$\sup_{|t-s|\le\delta_n} \left| n^{-1} \sum_{i=1}^n [Q_i(t) - Q_i(s)] \right| \le n^{-1} \sum_{i=1}^n L\delta_n = L\delta_n. \tag{30}$$

It remains to control the first term on the upper bound in (28). For this, we adopt arguments similar to those in the proof of Lemma 2 in Zhang and Wang (2016). Let $M > 0$ and define $a_n = (\log n/n)^{1/2}$. Then, using the independence of the Q_i along with the inequality $\exp(s) \le 1 + s + s^2$ for small enough s, we obtain for large enough n

$$P\left(n^{-1} \sum_{i=1}^n Q_i(t) - E(Q_1(t)) > Ma_n \right)$$

$$
\begin{aligned}
&= P\left(a_n \sum_{i=1}^{n}[Q_i(t) - E(Q_1(t))] > M \log n\right)\\
&\le n^{-M} \prod_{i=1}^{n} E\{\exp(a_n[Q_i(t) - E(Q_1(t))])\}\\
&\le n^{-M} \prod_{i=1}^{n}(1 + a_n^2 E[Q_i^2(t)])\\
&\le n^{-M} \prod_{i=1}^{n}(1 + a_n^2 c_0)\\
&\le n^{-M} \exp(n a_n^2 c_0)\\
&= n^{c_0 - M},
\end{aligned}
$$

where $c_0 = \sup_{y\in\mathcal{Y}} y^2 = 1$. Hence, taking $M > c_0 + \gamma$ implies

$$\sup_{t\in\mathcal{T}_n} \left| n^{-1}\sum_{i=1}^{n}[Q_i(t) - E(Q_1(t))]\right| = O_p(a_n).$$

Combining this with (28), (29), (30) and $\delta_n = o(a_n)$, which remains valid for any fixed $\gamma \ge 1/2$, shows that $\sup_{t\in[0,1]} |\hat{Q}^*(t) - Q^*(t)| = O_p(a_n)$. The result follows from (26). □

Proof of Theorem 4.2. Recall that when the densities f_i are fully observed, we have $\hat{Q}^*(t) = n^{-1}\sum_{i=1}^{n} Q_i(t)$. Under (S1) and by compactness of $\mathcal{X}$ and $\mathcal{Y}$, there exist $M_0 \in (0,\infty)$ and $L_0 \in (0,\infty)$ such that $M_0 \le \inf_{(y,x)\in\mathcal{Y}\times\mathcal{X}} f_{Y|X}(y,x) \le \sup_{(y,x)\in\mathcal{Y}\times\mathcal{X}} f_{Y|X}(y,x) \le L_0$. This implies that $\hat{Q}^*$ is differentiable and strictly increasing over $[0,1]$. Thus, the inverse function $\hat{F}^*(y) = \hat{Q}^{*-1}(y)$, $y \in \mathcal{Y}$, exists and is continuous on $\mathcal{Y} = [0,1]$. Also, $Q^*(t) = \int_{\mathcal{X}} Q_{Y|X}(t,x) f(x)dx$ is continuous and strictly increasing over $[0,1]$ and so is $F^*(y) = Q^{*-1}(y)$, $y \in \mathcal{Y}$. Hence, for $M > 0$ and $a_n = (\log n/n)^{1/2}$, denoting by $\mathbb{Q}$ the set of rational numbers,

$$
\begin{aligned}
P\left(\sup_{y\in\mathcal{Y}} |\hat{F}^*(y) - F^*(y)| > a_n M\right) &= P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} |\hat{F}^*(y) - F^*(y)| > a_n M\right)\\
&\le P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} \hat{F}^*(y) > F^*(y) + a_n M\right)\\
&\quad + P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} \hat{F}^*(y) < F^*(y) - a_n M\right). \quad (31)
\end{aligned}
$$

Note that by the Lebesgue dominated convergence theorem, it holds that for all $t \in (0,1)$

$$\frac{d}{dt}Q^*(t) = \int_{\mathcal{X}} \frac{1}{f_{Y|X}(Q_{Y|X}(t,x),x)} f(x)dx,$$

which under (S1) implies that $L_0^{-1} \leq dQ^*(t)/dt \leq M_0^{-1}$. Writing $\|Q^* - \hat{Q}^*\|_\infty = \sup_{t\in[0,1]}|Q^*(t) - \hat{Q}^*(t)|$, defining $\mathcal{A}_n = \{y \in \mathcal{Y} : F^*(y) + a_n M < 1\}$ and using that $0 \leq \hat{F}^*(y) \leq 1$, $y \in \mathcal{Y}$ and a Taylor expansion, it follows that

$$\begin{aligned}
& P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} \hat{F}^*(y) > F^*(y) + a_n M\right) \\
\leq & P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}\cap\mathcal{A}_n} Q^*(\hat{F}^*(y)) > Q^*(F^*(y) + a_n M)\right) \\
= & P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}\cap\mathcal{A}_n} Q^*(\hat{F}^*(y)) > y + \frac{d}{dt}Q^*(\xi_y) a_n M\right) \\
\leq & P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}\cap\mathcal{A}_n} Q^*(\hat{F}^*(y)) > y + a_n L_0^{-1} M\right) \\
\leq & P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}\cap\mathcal{A}_n} \|Q^* - \hat{Q}^*\|_\infty + \hat{Q}^*(\hat{F}^*(y)) > y + a_n L_0^{-1} M\right) \\
= & P\left(\|Q^* - \hat{Q}^*\|_\infty > a_n L_0^{-1} M\right),
\end{aligned}$$

where $\xi_y \in [0, 1]$. The tthird inequality is due to $Q^*(\hat{F}^*(y)) \leq \|Q^* - \hat{Q}^*\|_\infty + \hat{Q}^*(\hat{F}^*(y))$ and the last equality to $\hat{Q}^*(\hat{F}^*(y)) = y$. Combining this with $\|Q^* - \hat{Q}^*\|_\infty = O_p(a_n)$, which was shown in the proof of Theorem 4.1, leads to

$$P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} \hat{F}^*(y) > F^*(y) + a_n M\right) = o(1), \tag{32}$$

as $n \to \infty$ and for sufficiently large M. Similarly

$$P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} \hat{F}^*(y) < F^*(y) - a_n M\right) = o(1),$$

as $n \to \infty$ and sufficiently large M. Then applying (31) and (32) leads to the result. □

A.5. *Proof of theoretical results when sampling from conditional densities*

Suppose that $Y_{i1}, \ldots, Y_{in_i} \overset{iid}{\sim} f_i$, where $f_i(\cdot) = f_{Y|X}(\cdot, X_i)$, are a sample of observations coming from the conditional distribution $f_{Y|X}(\cdot, X_i)$ at predictor level X_i. We estimate f_i by

$$\hat{f}_i(y) = \sum_{j=1}^{n_i} \kappa_1\left(\frac{Y_{ij}-y}{h_1}\right) w_1(y,h) \Big/ \sum_{j=1}^{n_i} \int_0^1 \kappa_1\left(\frac{Y_{ij}-s}{h_1}\right) w_1(s,h)ds,$$

where $y \in \mathcal{Y} = [0,1]$. Let $\hat{F}_i(y) = \int_0^y \hat{f}_i(\nu)d\nu$ be the estimated cdf corresponding to the density f_i. By the central limit theorem, $n^{-1}\sum_{i=1}^n F_i^{-1}(t)$ converges in probability to $E(F_1^{-1}(t)) = E[E(F_1^{-1}(t)|X_1)] = E[Q_{Y|X}(t,X_1)] = Q^*(t)$ at the $\sqrt{n}$-rate. Thus a natural estimate of $Q^*(t)$, $t \in (0,1)$, is $\hat{Q}^*(t) = n^{-1}\sum_{i=1}^n \hat{F}_i^{-1}(t)$.

Lemma A.7. *Suppose that $\mathcal{Y} = [0,1]$, $\mathcal{X} = [0,1]^p$, $p \geq 1$, $\kappa_1 \in \mathcal{K}$, $f_{Y|X}(\cdot,x)$ is twice-continuously differentiable at each x and the regularity conditions (S1) and (S2) hold. Then*

$$E\left((\hat{F}_i(y) - F_i(y))^2\right) = O(h_1^4 + n_i^{-1}),$$

where the bound is uniform in y and depends on i only through n_i.

Proof of Lemma A.7. Set

$$f_{0i}(y) = (n_i h_1)^{-1} \sum_{j=1}^{n_i} \kappa_1\left(\frac{Y_{ij}-y}{h_1}\right) w_1(y,h_1),$$

and $F_{0i}(y) = \int_0^y f_{0i}(s)ds$. By Fubini's theorem,

$$\begin{aligned}
&\mathrm{Var}(F_{0i}(y)|X_i) \\
=&\frac{1}{n_i h_1^2}\mathrm{Var}\left(\int_0^y \kappa_1\left(\frac{Y_{i1}-s}{h_1}\right) w_1(s,h_1)ds|X_i\right) \\
\leq&\frac{1}{n_i h_1^2}\int_0^y\int_0^y\int_0^1 \kappa_1\left(\frac{r-s}{h_1}\right)\kappa_1\left(\frac{r-u}{h_1}\right) w_1(s,h_1)w_1(u,h_1)f_i(r)drdsdu \\
=&\frac{1}{n_i h_1}\int_0^y\int_0^y\int_{-s/h_1}^{(1-s)/h_1} \kappa_1(\nu)\kappa_1\left(\frac{s-u+\nu h_1}{h_1}\right) \qquad (33)\\
&\qquad \times w_1(s,h_1)w_1(u,h_1)f_i(s+\nu h_1)d\nu dsdu \\
=&\frac{1}{n_i}\int_0^y\int_{-s/h_1}^{(1-s)/h_1}\kappa_1(\nu)w_1(s,h_1)f_i(s+\nu h_1)\int_{-(s+\nu h_1)/h_1}^{(y-s-\nu h_1)/h_1} \qquad (34)\\
&\qquad \times \kappa_1(l)w_1(s+\nu h_1+lh_1,h_1)dld\nu ds \\
=&O(n_i^{-1}), \qquad (35)
\end{aligned}$$

where the bound is uniform in y and X_i. The last equality is due to $f_i(y) \leq \|f_{Y|X}\|_\infty < \infty$, $y \in \mathcal{Y}$, which follows from (S1), the compactness of $\mathcal{Y}$ and $\mathcal{X}$ and

$$1 \leq w_1(s,h_1) \leq \left(\int_0^1 \kappa_1(u)du\right)^{-1},$$

which holds for all $s \in [0,1]$, and $\|\kappa_1\|_\infty < \infty$ since $\kappa_1 \in \mathcal{K}$. Next, by a Taylor expansion, Fubini's theorem and as for all $s \in [0,1]$

$$\int_{-s/h_1}^{(1-s)/h_1} \kappa_1(\nu) w_1(s,h_1) d\nu = 1,$$

we obtain

$$\begin{aligned} E(F_{0i}(y)|X_i) &= \frac{1}{h_1}\int_0^y \int_0^1 \kappa_1\left(\frac{u-s}{h_1}\right) w_1(s,h_1) f_i(u) du ds \\ &= \int_0^y \int_{-s/h_1}^{(1-s)/h_1} \kappa_1(\nu) w_1(s,h_1) f_i(s+\nu h_1) d\nu ds \\ &= F_i(y) + h_1 \int_0^y f_i'(s) \int_{-s/h_1}^{(1-s)/h_1} \nu\kappa_1(\nu) w_1(s,h_1) d\nu ds \\ &\quad + \frac{h_1^2}{2}\int_0^y \int_{-s/h_1}^{(1-s)/h_1} \nu^2 \kappa_1(\nu) w_1(s,h_1) f_i''(\xi) d\nu ds, \end{aligned} \tag{36}$$

where $\xi = \xi(i,s,\nu,h_1)$ is between s and $s+\nu h_1$. Since $\kappa_1 \in \mathcal{K}$, $\|f_i''\|_\infty \le \|f_{Y|X}''\|_\infty < \infty$, due to the compactness of $\mathcal{Y}$ and $\mathcal{X}$, (S1) and the uniform of boundedness of w_1. Then

$$\frac{h_1^2}{2}\Big|\int_0^y \int_{-s/h_1}^{(1-s)/h_1} \nu^2\kappa_1(\nu) w_1(s,h_1) f_i''(\xi) d\nu ds\Big| = O(h_1^2), \tag{37}$$

where the bound is uniform in y and X_i. We next show that

$$h_1\Big|\int_0^y f_i'(s)\int_{-s/h_1}^{(1-s)/h_1} \nu\kappa_1(\nu) w_1(s,h_1) d\nu ds\Big| = O(h_1^2), \tag{38}$$

where the bound again is uniform in y and X_i. Setting $c_1 = \left(\int_0^1 \kappa_1(u)du\right)^{-1}$, it follows that

$$h_1\Big|\int_0^y f_i'(s)\int_{-s/h_1}^{(1-s)/h_1} \nu\kappa_1(\nu) w_1(s,h_1) d\nu ds\Big| \tag{39}$$

$$\le c_1 h_1 \|f_{Y|X}'\|_\infty \Big|\int_0^y \int_{-s/h_1}^{(1-s)/h_1} \nu\kappa_1(\nu) d\nu ds\Big|. \tag{40}$$

Note that if $y \in [0,h_1)$

$$\Big|\int_0^y \int_{-s/h_1}^{(1-s)/h_1} \nu\kappa_1(\nu) d\nu ds\Big| \le \int_0^{h_1} \Big|\int_{-s/h_1}^{1} \nu\kappa_1(\nu) d\nu\Big| ds = O(h_1),$$

which is due to $\kappa_1 \in \mathcal{K}$; this bound is uniform in y. Also, if $y \in [h_1, 1-h_1]$

$$\Big|\int_0^y \int_{-s/h_1}^{(1-s)/h_1} \nu\kappa_1(\nu) d\nu ds\Big| \le \int_0^{h_1} \Big|\int_{-s/h_1}^{1} \nu\kappa_1(\nu) d\nu\Big| ds$$

$$
\begin{aligned}
&+ \int_{h_1}^{y} \Big| \int_{-1}^{1} \nu\kappa_1(\nu) d\nu \Big| ds \\
&= \int_0^{h_1} \Big| \int_{-s/h_1}^{1} \nu\kappa_1(\nu) d\nu \Big| ds \\
&= O(h_1),
\end{aligned}
$$

due to $\int_{-1}^1 \nu\kappa_1(\nu)d\nu = 0$ as $\kappa_1 \in \mathcal{K}$; this bound is also uniform in y. Finally, if $y \in (1-h_1, 1]$ we have

$$
\begin{aligned}
\Big| \int_0^y \int_{-s/h_1}^{(1-s)/h_1} \nu\kappa_1(\nu) d\nu ds \Big| &\le O(h_1) + \int_{1-h_1}^{y} \Big| \int_{-1}^{(1-s)/h_1} \nu\kappa_1(\nu) d\nu \Big| ds \\
&= O(h_1),
\end{aligned}
$$

where the bound is uniform in y. With (40) this shows that (38) holds. Combining (36), (37) and (38) leads to

$$
|E(F_{0i}(y)|X_i) - F_i(y)| = O(h_1^2), \tag{41}
$$

where the bound is uniform in y, i and X_i. From the proof of Proposition 1 in Petersen and Müller (2016b), $F_{0i}(1) = \int_0^1 f_{0i}(s)ds \ge \int_0^1 \kappa_1(u)du = c_1^{-1}$. Since $\hat{F}_i(y) = F_{0i}(y)/F_{0i}(1)$, furthermore

$$
\begin{aligned}
&E\left((\hat{F}_i(y) - F_i(y))^2 \right) \\
\le& 2c_1^2 \Big[E\left((F_{0i}(y) - F_i(y))^2 \right) + E\left((1 - F_{0i}(1))^2 \right) \Big] \\
=& 2c_1^2 E\Big[E\left((F_{0i}(y) - F_i(y))^2 | X_i \right) + E\left((1 - F_{0i}(1))^2 | X_i \right) \Big] \\
=& O(h_1^4 + n_i^{-1}),
\end{aligned}
$$

where the bound is uniform in y and depends on i only through n_i, and the last equality is due to (35), (41) and since $F_i(1) = 1$. The result follows. □

Proof of Theorem 3.4. Note that

$$
\begin{aligned}
d_{\mathcal{W}}^2(\hat{\nu}^*, \nu^*) &= \int_0^1 (\hat{Q}^*(t) - Q^*(t))^2 dt \\
&= \int_0^1 \left(n^{-1} \sum_{i=1}^n \hat{F}_i^{-1}(t) - Q^*(t) \right)^2 dt \\
&= \int_0^1 \left(n^{-1} \sum_{i=1}^n [\hat{F}_i^{-1}(t) - F_i^{-1}(t)] \right)^2 dt \qquad (42) \\
&\quad + \int_0^1 \left(n^{-1} \sum_{i=1}^n F_i^{-1}(t) - Q^*(t) \right)^2 dt + 2 \int_0^1 A_n(t) B_n(t) dt,
\end{aligned}
$$

where $A_n(t) = n^{-1}\sum_{i=1}^n[\hat{F}_i^{-1}(t) - F_i^{-1}(t)]$ and $B_n(t) = n^{-1}\sum_{i=1}^n F_i^{-1}(t) - Q^*(t)$. Since

$$\int_0^1 A_n(t)B_n(t)dt \le \Big[\int_0^1 A_n^2(t)dt\int_0^1 B_n^2(t)dt\Big]^{1/2},$$

it suffices to control the terms $\int_0^1 A_n^2(t)dt$ and $\int_0^1 B_n^2(t)dt$. Since $E(B_n(t)) = 0$,

$$\begin{aligned}E\left(\int_0^1 B_n^2(t)dt\right) &= \int_0^1 E(B_n^2(t))dt = \int_0^1 \mathrm{Var}(B_n(t))dt\\ &= n^{-1}\int_0^1 \mathrm{Var}(F_1^{-1}(t))dt = O(n^{-1}),\end{aligned}$$

where the last equality is due to the compactness of $\mathcal{Y} = [0,1]$ and

$$\mathrm{Var}(F_1^{-1}(t)) = \mathrm{Var}(Q_{Y|X}(t, X_1)) \le [\sup_{y\in\mathcal{Y}} y - \inf_{y\in\mathcal{Y}} y]^2/4,$$

whence

$$\int_0^1 B_n^2(t)dt = O_p(n^{-1}). \tag{43}$$

Next, similar arguments as in the proof of Theorem 3.1 and (S1) show that $M = \sup_{(t,x)\in[0,1]\times\mathcal{X}} D(t,x) < \infty$, where $D(t,x) = 1/f_{Y|X}(Q_{Y|X}(t,x),x)$ is the conditional quantile density, and

$$\begin{aligned}\int_0^1 (\hat{F}_i^{-1}(t) - F_i^{-1}(t))^2 dt &\le M\int_0^1 (\hat{F}_i(\nu) - F_i(\nu))^2 \hat{f}_i(\nu)d\nu\\ &= M\int_0^1 (\hat{F}_i(\nu) - F_i(\nu))^2 f_i(\nu)d\nu\\ &\le M\left(\sup_{y\in\mathcal{Y},x\in\mathcal{X}} f_{Y|X}(y,x)\right)\int_0^1 (\hat{F}_i(\nu) - F_i(\nu))^2 d\nu,\end{aligned}$$

where the first equality is due to $\hat{\Delta}_n(s) := \hat{F}_i(s) - F_i(s)$, $s \in \mathcal{Y}$, with $\hat{\Delta}_n(1) = 0$ and $\hat{\Delta}_n(0) = 0$. Lemma A.7 leads to

$$E\left(\int_0^1 (\hat{F}_i^{-1}(t) - F_i^{-1}(t))^2 dt\right) = O(h_1^4 + n_i^{-1}), \tag{44}$$

where the bound depends on i only through n_i. Note that

$$\begin{aligned}\int_0^1 A_n^2(t)dt = {}& n^{-2}\sum_{i=1}^n \int_0^1 (\hat{F}_i^{-1}(t) - F_i^{-1}(t))^2 dt\\ &+ n^{-2}\sum_{i=1}^n\sum_{j\ne i}\int_0^1 (\hat{F}_i^{-1}(t) - F_i^{-1}(t))(\hat{F}_j^{-1}(t) - F_j^{-1}(t))dt.\end{aligned} \tag{45}$$

Using (44) along with a conditioning argument and the fact that $n_i \geq m(n)$,

$$E\left(n^{-2}\sum_{i=1}^{n}\int_0^1(\hat{F}_i^{-1}(t)-F_i^{-1}(t))^2dt\right)=O\left(\frac{h_1^4+m(n)^{-1}}{n}\right).$$

Similarly, by independence,

$$E\left(n^{-2}\sum_{i=1}^{n}\sum_{j\neq i}\int_0^1(\hat{F}_i^{-1}(t)-F_i^{-1}(t))(\hat{F}_j^{-1}(t)-F_j^{-1}(t))dt\right)$$
$$=O\left(h_1^4+1/m(n)\right).$$

Combining this with (45) leads to

$$\int_0^1 A_n^2(t)dt=O_p\left(h_1^4+1/m(n)\right). \tag{46}$$

The result then follows from (42), (43), (46) and the condition $m(n)h_1^4=O(1)$. □

In what follows, recall that F_i is the cdf correponding to $f_i = f_{Y|X}(\cdot, X_i)$, $i = 1,\dots,n$.

Lemma A.8. *Suppose that $\mathcal{Y}=[0,1]$, $\mathcal{X}=[0,1]^p$, $p\geq 1$, $\kappa_1\in\mathcal{K}$, $n_i\geq m(n)$ with $m(n)\to\infty$ as $n\to\infty$, the regularity condition (S1) holds and $f_{Y|X}(\cdot,x)$ is twice-continuously differentiable at each x. If $m(n)h_1^4=O(1)$, $m(n)^{2\gamma_0-1}h_1^2\log m(n)\to\infty$ and $n/m(n)^{\rho_1}=o(1)$ as $n\to\infty$ for some $\gamma_0>1/2$ and $\rho_1>0$, then*

$$\sup_{t\in[0,1]}|\hat{Q}^*(t)-Q^*(t)|=O_p((\log n/n)^{1/2}+a_n),$$

where $a_n=\sqrt{\log m(n)/m(n)}$.

Proof of Lemma A.8. Observe

$$\begin{aligned}\sup_{t\in[0,1]}|\hat{Q}^*(t)-Q^*(t)| &\leq \sup_{t\in[0,1]}|n^{-1}\sum_{i=1}^{n}\hat{Q}_i(t)-Q_i(t)| &(47)\\ &\quad+\sup_{t\in[0,1]}|n^{-1}\sum_{i=1}^{n}Q_i(t)-E(F_1^{-1}(t))| \\ &\leq \sup_{t\in[0,1]}|n^{-1}\sum_{i=1}^{n}\hat{Q}_i(t)-Q_i(t)|+O_p((\log n/n)^{1/2}), &(48)\end{aligned}$$

where the last inequality follows from the proof of Theorem 4.1 and using that $Q_{Y|X}(\cdot,x)$ is $(1/\kappa_0)$-Lipschitz with $\kappa_0=\inf_{(y,x)\in\mathcal{Y}\times\mathcal{X}}f_{Y|X}(y,x)>0$, which is due to (S1). For $M>0$,

$$P\left(\sup_{t\in[0,1]}\left|n^{-1}\sum_{i=1}^{n}\hat{Q}_i(t)-Q_i(t)\right|>Ma_n\right) \tag{49}$$

$$\leq P\left(n^{-1}\sum_{i=1}^{n}\sup_{t\in[0,1]}|\hat{Q}_i(t)-Q_i(t)|>Ma_n\right)$$
$$\leq \sum_{i=1}^{n} P\left(\sup_{t\in[0,1]}|\hat{Q}_i(t)-Q_i(t)|>Ma_n\right), \quad (50)$$

so that it suffices to control the upper bound term in (50). For this, let $\chi_n(\gamma)$ be an equidistant grid in $[0,1]$ with spacing $m^{-\gamma_0}$. Then

$$P\left(\sup_{t\in[0,1]}|\hat{Q}_i(t)-Q_i(t)|>Ma_n\right) \quad (51)$$
$$\leq P\left(\sup_{t\in\chi_n(\gamma)}|\hat{Q}_i(t)-Q_i(t)|>Ma_n/2\right)$$
$$+ P\left(\sup_{t,s\in[0,1],|t-s|\leq m^{-\gamma_0}}|\hat{Q}_i(t)-\hat{Q}_i(s)|>Ma_n/4\right)$$
$$+ P\left(\sup_{t,s\in[0,1],|t-s|\leq m^{-\gamma_0}}|Q_i(t)-Q_i(s)|>Ma_n/4\right). \quad (52)$$

From the $(1/\kappa_0)$-Lipschitz continuity of the Q_i, we have for large enough n

$$P\left(\sup_{t,s\in[0,1],|t-s|\leq m^{-\gamma_0}}|Q_i(t)-Q_i(s)|>Ma_n/4\right)\leq 1_{\{\kappa_0^{-1}m^{-\gamma_0}\geq Ma_n/4\}}=0, \quad (53)$$

where the last equality is due to $m(n)^{2\gamma_0-1}\log m(n)\to\infty$ as $n\to\infty$.

Next, let $\Delta_i=\sup_{y\in\mathcal{Y}}|\hat{F}_i(y)-F_i(y)|$ and suppose that we can show

$$\max_{i=1,\ldots,n}\Delta_i=o_p(1). \quad (54)$$

Then it suffices to work on the event where $\max_{i=1,\ldots,n}\Delta_i\leq 1/4$, which holds with probability tending to 1 as $n\to\infty$. Also, note that for large enough n it holds that $m^{-\gamma_0}<1/4$.

Analogously to the proof of Corollary 1 in Bonnéry, Breidt and Coquet (2012), observing the inclusion

$$\{y\in\mathcal{Y}:F_i(y)\geq t+\Delta_i\}\subseteq\{y\in\mathcal{Y}:\hat{F}_i(y)\geq t\}\subseteq\{y\in\mathcal{Y}:F_i(y)\geq t-\Delta_i\},$$

where $t\in[0,1]$, it then follows that

$$Q_i(t-\Delta_i)\leq\hat{Q}_i(t)\leq Q_i(t+\Delta_i),$$

where we define $Q_i(r)=0$ for $r<0$ and $Q_i(r)=1$ for $r>1$. Thus

$$(\hat{Q}_i(t)-\hat{Q}_i(s))1_{\{s\leq t\}}\leq Q_i(t+\Delta_i)-Q_i(s-\Delta_i). \quad (55)$$

Note that if $t+\Delta_i \geq 1$ and $|t-s| \leq m^{-\gamma_0}$, then for large enough n we have

$$s - \Delta_i \geq 1 - 2\Delta_i - m^{-\gamma_0} \geq 1/4,$$

and thus $s - \Delta_i \in [1/4, 1]$. Hence, by a Taylor expansion and large enough n, we obtain

$$\begin{aligned}
&(Q_i(t+\Delta_i) - Q_i(s-\Delta_i))1_{\{t \geq 1-\Delta_i, t \geq s, |t-s| \leq m^{-\gamma_0}\}} \\
\leq & 1 - (Q_i(1) + Q_i'(\xi_i)(s - \Delta_i - 1)) \\
\leq & \kappa_0^{-1}(1 - s + \Delta_i) \\
\leq & \kappa_0^{-1}(2\Delta_i + m^{-\gamma_0}),
\end{aligned}$$

where ξ_i lies between $s - \Delta_i$ and 1. Also, if $t + \Delta_i \leq 1$, $|t-s| \leq m^{-\gamma_0}$ and $s \geq \Delta_i$, then by the κ_0^{-1}-Lipschitz continuity of Q_i and large enough n

$$(Q_i(t+\Delta_i) - Q_i(s-\Delta_i))1_{\{t \leq 1-\Delta_i, t \geq s, s \geq \Delta_i, |t-s| \leq m^{-\gamma_0}\}} \leq \kappa_0^{-1}(m^{-\gamma_0} + 2\Delta_i).$$

Similarly, if $t + \Delta_i \leq 1$, $|t-s| \leq m^{-\gamma_0}$ and $s \leq \Delta_i$, then $Q_i(s - \Delta_i) = 0$ and $t \leq \Delta_i + m^{-\gamma_0}$. By a Taylor expansion and large enough n, it then follows that

$$\begin{aligned}
(Q_i(t+\Delta_i) - Q_i(s-\Delta_i))1_{\{t \leq 1-\Delta_i, t \geq s, s \leq \Delta_i, |t-s| \leq m^{-\gamma_0}\}} \leq & \kappa_0^{-1}(t + \Delta_i) \\
\leq & \kappa_0^{-1}(m^{-\gamma_0} + 2\Delta_i).
\end{aligned}$$

Combining these observations leads to

$$(Q_i(t+\Delta_i) - Q_i(s-\Delta_i))1_{\{t \geq s, |t-s| \leq m^{-\gamma_0}\}} \leq \kappa_0^{-1}(m^{-\gamma_0} + 2\Delta_i),$$

which holds for large enough n. This along with (55) and as $\hat{Q}_i$ is non-decreasing, implies

$$|\hat{Q}_i(t) - \hat{Q}_i(s)|1_{\{|t-s| \leq m^{-\gamma_0}\}} \leq \kappa_0^{-1}(m^{-\gamma_0} + 2\Delta_i)$$

for large enough n, whence

$$\begin{aligned}
& P\left(\sup_{t,s \in [0,1], |t-s| \leq m^{-\gamma_0}} |\hat{Q}_i(t) - \hat{Q}_i(s)| > Ma_n/4\right) && (56) \\
\leq & P\left(\kappa_0^{-1} m^{-\gamma_0} + 2\kappa_0^{-1}\Delta_i > Ma_n/4\right) \\
\leq & P\left(\Delta_i > M\kappa_0 a_n/16\right). && (57)
\end{aligned}$$

Here the last inequality follows from the fact that $a_n - 4m^{-\gamma_0}\kappa_0^{-1}M^{-1} > a_n/2$, which holds for large enough n and is due to $\gamma_0 > 1/2$.

Let $M' > 0$. Suppose we can show that there exists a constant $C_0 > 0$ depending only on γ_0 such that for any $M' > 16C_0$

$$P\left(\Delta_i > M'a_n\right) = O(m(n)^{C_0 - M'/16}), \tag{58}$$

as $n \to \infty$, where the bound is uniform in i. Then, combining this with (57) and taking $M > 256C_0/\kappa_0$ leads to

$$P\left(\sup_{t,s\in[0,1],|t-s|\le m^{-\gamma_0}} |\hat{Q}_i(t) - \hat{Q}_i(s)| > Ma_n/4\right) = O\left(m(n)^{C_0 - M\kappa_0/256}\right), \tag{59}$$

as $n \to \infty$, where the bound is uniform in i. We next note that (58) implies (54) since for any $M_0 > 0$

$$P\left(\max_{i=1,\dots,n} \Delta_i > M_0\right) \le \sum_{i=1}^{n} P(\Delta_i > M_0) = O(n/m(n)^{M_0/16 - C_0}) = o(1),$$

where the last equality is due to the condition $n/m(n)^{\rho_1} = o(1)$ for some $\rho_1 > 0$, taking $M_0 > 16(C_0 + \rho_1)$.

Next, by a Taylor expansion and for $M > 32C_0/\kappa_0$, we have

$$\begin{aligned}
&P\left(\hat{Q}_i(t) - Q_i(t) > Ma_n/2\right)\\
\le&P\left(\hat{F}_i(Q_i(t) + Ma_n/2) < t \wedge Q_i(t) + Ma_n/2 \le 1\right)\\
\le&P\left(-\Delta_i + F_i(Q_i(t) + Ma_n/2) < t \wedge Q_i(t) + Ma_n/2 \le 1\right)\\
\le&P\left(\Delta_i > \kappa_0 Ma_n/2\right)\\
=&O(m(n)^{C_0 - \kappa_0 M/32}),
\end{aligned}$$

as $n \to \infty$, where the bound is uniform in i. Similar arguments lead to

$$P\left(\hat{Q}_i(t) - Q_i(t) < -Ma_n/2\right) \le P\left(\Delta_i > \kappa_0 Ma_n/2\right).$$

Hence, for $M > 32C_0/\kappa_0$

$$P\left(|\hat{Q}_i(t) - Q_i(t)| > Ma_n/2\right) = O(m(n)^{C_0 - \kappa_0 M/32}),$$

as $n \to \infty$ and therefore

$$P\left(\sup_{t\in\chi_n(\gamma)} |\hat{Q}_i(t) - Q_i(t)| > Ma_n/2\right) = O(m(n)^{C_0 + \gamma_0 - \kappa_0 M/32}), \tag{60}$$

as $n \to \infty$, where the bound is uniform in i. Combining (52), (53), (59) and (60) leads to

$$P\left(\sup_{t\in[0,1]} |\hat{Q}_i(t) - Q_i(t)| > Ma_n\right) = O(m(n)^{C_0 + \gamma_0 - \kappa_0 M/32} + m(n)^{C_0 - M\kappa_0/256}),$$

as $n \to \infty$, where the bound is uniform in i and we take $M > 256C_0/\kappa_0$. Therefore, for large enough M

$$\sum_{i=1}^{n} P\left(\sup_{t\in[0,1]} |\hat{Q}_i(t) - Q_i(t)| > Ma_n\right) = o(1),$$

as $n \to \infty$, where the last equality follows from the condition $n/m(n)^{\rho_1} = o(1)$ for some $\rho_1 > 0$. The result then follows from (48) and (50).

It remains to prove (58). Let

$$f_{0i}(y) = (n_i h_1)^{-1} \sum_{j=1}^{n_i} \kappa_1 \left(\frac{Y_{ij} - y}{h_1} \right) w_1(y, h_1),$$

with corresponding cdf $F_{0i}(y) = \int_0^y f_{0i}(s)ds$ and $a_{in} = \sqrt{\log n_i / n_i}$. Let

$$b_{ij} = h_1^{-1} \int_0^y \kappa_1((s - Y_{ij})/h_1) w_1(s, h_1) ds,$$

and $U_{ij} = b_{ij} - E(b_{ij}|X_i)$, $j = 1, \ldots, n_i$, $i = 1, \ldots, n$. Note that $E(U_{ij}|X_i) = 0$ and that the U_{ij} are (conditionally) independent in j given X_i. By similar arguments as in the proof of Lemma A.5 and using that $\kappa_1 \in \mathcal{K}$, we have $b_{ij} \le (\int_0^1 \kappa_1(s)ds)^{-1} = 2$ and for large enough n and $M' > 4$

$$\begin{aligned} P\left(F_{0i}(y) - E(F_{0i}(y)|X_i) > M' a_{in} | X_i\right) &= P\left(\frac{1}{n_i} \sum_{j=1}^{n_i} U_{ij} > M' a_{in} | X_i \right) \\ &= P\left(a_{in} \sum_{j=1}^{n_i} U_{ij} > M' \log n_i | X_i \right) \\ &\le n_i^{-M'} \prod_{j=1}^{n_i} E(\exp(a_{in} U_{ij}) | X_i) \\ &\le n_i^{-M'} \prod_{j=1}^{n_i} (1 + 4a_{in}^2) \\ &\le n_i^{-M'} \exp(4 n_i a_{in}^2) \\ &\le m(n)^{4-M'}. \end{aligned}$$

Therefore, for large enough n and $M' > 4$

$$P\left(F_{0i}(y) - E(F_{0i}(y)|X_i) > M' a_{in}\right) \le m(n)^{4-M'},$$

and similarly

$$P\left(F_{0i}(y) - E(F_{0i}(y)|X_i) < -M' a_{in}\right) \le m(n)^{4-M'}.$$

Let $\mathcal{Y}_n$ be an equidistant partition of $\mathcal{Y} = [0, 1]$ with spacing $m(n)^{-\gamma_0}$. Thus, for large enough n and $M' > 4 + \gamma_0$

$$P\left(\sup_{y \in \mathcal{Y}_n} |F_{0i}(y) - E(F_{0i}(y)|X_i)| > M' a_{in} \right) \le 2m(n)^{\gamma_0 + 4 - M'}. \tag{61}$$

Now, from the proof of Lemma A.7 we have $|E(F_{0i}(y)|X_i) - F_i(y)| = O(h_1^2)$, where the bound is uniform in X_i, i and y. This shows that

$$\sup_{y\in\mathcal{Y}} |E(F_{0i}(y)|X_i) - F_i(y)| = O(h_1^2), \tag{62}$$

where the bound is uniform in X_i and i. From the uniform bound on w_1 and κ_1,

$$\sup_{|y-s|\le m(n)^{-\gamma_0}} |F_{0i}(y) - F_{0i}(s)| = O(m(n)^{-\gamma_0} h_1^{-1}), \tag{63}$$

where the bound is uniform in X_i, Y_{ij}, $j = 1, \dots, n_i$, and i. Similarly

$$\sup_{|y-s|\le m(n)^{-\gamma_0}} |E(F_{0i}(y)|X_i) - E(F_{0i}(s)|X_i)| = O(m(n)^{-\gamma_0} h_1^{-1}), \tag{64}$$

where the bound is uniform in X_i and i. Hence, for large enough n and taking $M' > 16+4\gamma_0$, observing that $a_{in} \le a_n$ for n large enough so that $m(n) > \exp(1)$ is satisfied, we obtain

$$\begin{aligned}
&P\left(\sup_{y\in\mathcal{Y}} |F_{0i}(y) - F_i(y)| > M'a_n\right)\\
\le &P\left(\sup_{y\in\mathcal{Y}} |F_{0i}(y) - E(F_{0i}(y)|X_i)| > M'a_n/2\right)\\
&+ P\left(\sup_{y\in\mathcal{Y}} |E(F_{0i}(y)|X_i) - F_i(y)| > M'a_n/2\right)\\
&\le P\left(\sup_{y\in\mathcal{Y}} |F_{0i}(y) - E(F_{0i}(y)|X_i)| > M'a_n/2\right)\\
&+ 1_{\{O(h_1^2)\ge M'a_n/2\}}\\
&\le P\left(\sup_{y\in\mathcal{Y}_n} |F_{0i}(y) - E(F_{0i}(y)|X_i)| > M'a_n/4\right)\\
&+ P\left(\sup_{|y-s|\le m(n)^{-\gamma_0}} |F_{0i}(y) - F_{0i}(s)| > M'a_n/8\right)\\
&+ P\left(\sup_{|y-s|\le m(n)^{-\gamma_0}} |E(F_{0i}(y) - F_{0i}(s)|X_i)| > M'a_n/8\right)\\
&\le 2m(n)^{\gamma_0+4-M'/4} + 1_{\{O(m(n)^{-\gamma_0}h_1^{-1})\ge M'a_n/8\}}\\
&\le 2m(n)^{\gamma_0+4-M'/4},
\end{aligned}$$

where the second inequality is due to (62), the third follows from (62) and $m(n)h_1^4 = O(1)$ as $n \to \infty$. The fourth inequality is due to (61), (63) and (64) and $M' > 16+4\gamma_0$. The last inequality follows from the condition $m(n)^{2\gamma_0-1}h_1^2 \times \log m(n) \to \infty$ as $n \to \infty$. This shows that

$$P\left(\sup_{y\in\mathcal{Y}} |F_{0i}(y) - F_i(y)| > M'a_n\right) \le 2m(n)^{\gamma_0+4-M'/4}, \tag{65}$$

for large enough n and $M' > 16 + 4\gamma_0$.

Next, set $\tau_i = F_{0i}(1)$ so that $\hat{F}_i(y) = F_{0i}(y)/\tau_i$ and consider a fixed $M' > 16(\gamma_0 + 4)$. Then (65) and $F_i(1) = 1$ imply

$$P(|\tau_i - 1| > M'a_n) \leq 2m(n)^{\gamma_0+4-M'/4},$$

and thus for large enough n

$$\begin{aligned} P\left(\sup_{y\in\mathcal{Y}} |F_{0i}(y) - F_i(y)| > M'a_n\tau_i/2\right) &\leq P\left(\sup_{y\in\mathcal{Y}} |F_{0i}(y) - F_i(y)| > M'a_n/4\right) \\ &\quad + 2m(n)^{\gamma_0+4-M'/4} \\ &\leq 2m(n)^{\gamma_0+4-M'/16} + 2m(n)^{\gamma_0+4-M'/4} \\ &= O(m(n)^{\gamma_0+4-M'/16}). \end{aligned}$$

Here the first inequality follows by using that $M'a_n < 1/2$ holds for large enough n, the second inequality is due to (65) and the last equality holds since $M' > 16(\gamma_0 + 4)$. Similarly

$$P\left(|\tau_i - 1| > M'a_n\tau_i/2\right) = O(m(n)^{\gamma_0+4-M'/16}),$$

as $n \to \infty$. Therefore

$$\begin{aligned} P\left(\sup_{y\in\mathcal{Y}} |\hat{F}_i(y) - F_i(y)| > M'a_n\right) &\leq P\left(\sup_{y\in\mathcal{Y}} |F_{0i}(y) - F_i(y)| > M'a_n\tau_i/2\right) \\ &\quad + P\left(|\tau_i - 1| > M'a_n\tau_i/2\right) \\ &= O(m(n)^{\gamma_0+4-M'/16}), \end{aligned}$$

as $n \to \infty$. The result in (58) then follows by taking $C_0 = \gamma_0 + 4$. □

Proof of Theorem 4.3. Observe that

$$\begin{aligned} \sup_{y\in\mathcal{Y}} |\hat{T}_i^*(y) - T_i^*(y)| &= \sup_{y\in\mathcal{Y}} |\hat{Q}^*(\hat{F}_i(y)) - Q^*(F_i(y))| \\ &\leq \sup_{y\in\mathcal{Y}} |\hat{Q}^*(\hat{F}_i(y)) - Q^*(\hat{F}_i(y))| \\ &\quad + \sup_{y\in\mathcal{Y}} |Q^*(\hat{F}_i(y)) - Q^*(F_i(y))| \\ &\leq \sup_{y\in\mathcal{Y}} |\hat{Q}^*(\hat{F}_i(y)) - Q^*(\hat{F}_i(y))| \\ &\quad + (1/\kappa_0) \sup_{y\in\mathcal{Y}} |\hat{F}_i(y) - F_i(y)|, \end{aligned}$$

where $\kappa_0 = \inf_{(y,x)\in\mathcal{Y}\times\mathcal{X}} f_{Y|X}(y,x) > 0$ and the last inequality is due to the $(1/\kappa_0)$-Lipschitz continuity of Q^*, which follows from (S1). Hence

$$\max_{i=1,\ldots,n} \sup_{y\in\mathcal{Y}} |\hat{T}_i^*(y) - T_i^*(y)| \leq \max_{i=1,\ldots,n} \sup_{y\in\mathcal{Y}} |\hat{Q}^*(\hat{F}_i(y)) - Q^*(\hat{F}_i(y))|$$

$$+ (1/\kappa_0) \max_{i=1,\ldots,n} \sup_{y\in\mathcal{Y}} |\hat{F}_i(y) - F_i(y)|. \tag{66}$$

Let $a_n = \sqrt{\log m(n)/m(n)}$. From the proof of Lemma A.8, we have that there exist $C > 0$ and $C_0 > 0$ such that for $M > 16C_0$ and large enough n

$$P\left(\sup_{y\in\mathcal{Y}} |\hat{F}_i(y) - F_i(y)| > Ma_n\right) \le \frac{C}{m(n)^{M/16-C_0}}.$$

This along with the independence across i of the quantities $\hat{F}_i(y)$ and $F_i(y)$ implies that for large enough n and $M > 16C_0$

$$\begin{aligned}
&P\left(\max_{i=1,\ldots,n} \sup_{y\in\mathcal{Y}} |\hat{F}_i(y) - F_i(y)| > Ma_n\right)\\
=&1 - \prod_{i=1}^{n} P\left(\sup_{y\in\mathcal{Y}} |\hat{F}_i(y) - F_i(y)| \le Ma_n\right)\\
=&1 - \prod_{i=1}^{n}\left[1 - P\left(\sup_{y\in\mathcal{Y}} |\hat{F}_i(y) - F_i(y)| > Ma_n\right)\right]\\
\le&1 - \left(1 - \frac{C}{m(n)^{M/16-C_0}}\right)^n.
\end{aligned}$$

Since $n/m(n)^{\rho_1} = o(1)$ as $n \to \infty$ for some $\rho_1 > 0$, it is easy to show that for large enough M

$$1 - \left(1 - \frac{C}{m(n)^{M/16-C_0}}\right)^n = o(1),$$

as $n \to \infty$. Therefore

$$\max_{i=1,\ldots,n} \sup_{y\in\mathcal{Y}} |\hat{F}_i(y) - F_i(y)| = O_p(a_n). \tag{67}$$

From Lemma A.8, we have

$$\begin{aligned}
\max_{i=1,\ldots,n} \sup_{y\in\mathcal{Y}} |\hat{Q}^*(\hat{F}_i(y)) - Q^*(\hat{F}_i(y))| \le& \sup_{t\in[0,1]} |\hat{Q}^*(t) - Q^*(t)|\\
=&O_p((\log n/n)^{1/2} + a_n),
\end{aligned}$$

which together with (66) and (67) leads to the result. □

Proof of Theorem 4.4. Recall that $\hat{F}^*(y) = \inf_{t\in[0,1]}\{\hat{Q}^*(t) > y\}$ and $\hat{Q}^*(t) = n^{-1}\sum_{i=1}^n \hat{Q}_i(t)$, where $\hat{Q}_i(t) = \inf_{y\in\mathcal{Y}}\{\hat{F}_i(y) \ge t\}$, $t \in [0,1]$. Also denote by $\|Q^* - \hat{Q}^*\|_\infty = \sup_{t\in[0,1]}|Q^*(t) - \hat{Q}^*(t)|$. Since $\hat{F}_i$ is continuous over $\mathcal{Y}$ with $\hat{F}^*(0) = 0$, $\hat{F}^*(1) = 1$, it holds that the range $\mathrm{ran}(\hat{F}^*) = [0,1]$ and thus similar arguments as in Embrechts and Hofert (2013) show that $\hat{Q}_i$ is strictly increasing on $[0,1]$ and therefore so is $\hat{Q}^*$. Since $\hat{Q}^*$ is non-decreasing, arguments in Feng et al. (2012) show that $\hat{F}^*$ is right-continuous. To show that $\hat{F}^*$ is continuous, we argue by contradiction. Suppose that $\hat{F}^*$ is not left-continuous so that there exists $y_0 \in \mathcal{Y} = [0,1]$ such that $\hat{F}^*(y_0^-) < \hat{F}^*(y_0)$. Then there exists $\phi_1 < \phi_2$ such that $\hat{F}^*(y_0^-) < \phi_1 < \phi_2 < \hat{F}^*(y_0)$. Then for small enough $\delta > 0$,

$$\hat{F}^*(y_0 - \delta) \le \hat{F}^*(y_0^-) < \phi_1 < \phi_2 < \hat{F}^*(y_0).$$

This implies $y_0 - \delta < \hat{Q}^*(\phi_1) \le y_0$ and taking $\delta \downarrow 0$ shows that $\hat{Q}^*(\phi_1) = y_0$ and similarly $\hat{Q}^*(\phi_2) = y_0$. Thus $\hat{Q}^*(\phi_1) = \hat{Q}^*(\phi_2)$ with $\phi_1 < \phi_2$, which contradicts the fact that $\hat{Q}^*$ is strictly increasing. Hence $\hat{F}^*$ is continuous over $\mathcal{Y} = [0,1]$.

Next, setting $b_n = (\log n/n)^{1/2} + (\log m(n)/m(n))^{1/2}$ and using the continuity of $\hat{F}^*$ and the arguments in the proof of Theorem 4.2 leads to

$$P\left(\sup_{y\in\mathcal{Y}} |\hat{F}^*(y) - F^*(y)| > b_n M\right) \le P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} \hat{F}^*(y) > F^*(y) + b_n M\right) + P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} \hat{F}^*(y) < F^*(y) - b_n M\right), \quad (68)$$

and $L_0^{-1} \le dQ^*(t)/dt \le M_0^{-1}$, $t \in [0,1]$, for some positive constants $L_0, M_0 > 0$. Setting $\mathcal{A}_n = \{y \in \mathcal{Y} : F^*(y) + b_n M < 1\}$ and using that $0 \le \hat{F}^*(y) \le 1$, $y \in \mathcal{Y}$, we obtain

$$\begin{aligned}
&P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} \hat{F}^*(y) > F^*(y) + b_n M\right)\\
\le &P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}\cap\mathcal{A}_n} Q^*(\hat{F}^*(y)) > Q^*(F^*(y) + b_n M)\right)\\
\le &P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}\cap\mathcal{A}_n} \|Q^* - \hat{Q}^*\|_\infty + \hat{Q}^*(\hat{F}^*(y)) > y + b_n L_0^{-1} M\right)\\
\le &P\left(\|Q^* - \hat{Q}^*\|_\infty > b_n L_0^{-1} M\right),
\end{aligned}$$

where the last inequality follows from the fact that $\hat{Q}^*(\hat{F}^*(y)) \le y$, which we show next. Indeed, suppose that there exists $y \in \mathcal{Y}$ such that $\hat{Q}^*(\hat{F}^*(y)) > y$. Then, there exists $\upsilon \in \mathcal{Y}$ such that $y < \upsilon < \hat{Q}^*(\hat{F}^*(y))$. By left-continuity of $\hat{Q}^*$, we have that for any $\epsilon > 0$ there exists $\delta = \delta(\epsilon) > 0$ such that $\hat{Q}^*(\hat{F}^*(y) - \delta) \ge \hat{Q}^*(\hat{F}^*(y)) - \epsilon > \nu - \epsilon$. Taking $\epsilon = (\upsilon - y)/2 > 0$ implies

$$\hat{Q}^*(\hat{F}^*(y) - \delta) > (\upsilon + y)/2 > y,$$

and thus $\hat{F}^*(y) \le \hat{F}^*(y) - \delta$, which is a contradiction as $\delta > 0$. Thus $\hat{Q}^*(\hat{F}^*(y)) \le y$.

Combining this with $\|Q^* - \hat{Q}^*\|_\infty = O_p(b_n)$, which is due to Lemma A.8, leads to

$$P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} \hat{F}^*(y) > F^*(y) + b_n M\right) = o(1), \tag{69}$$

as $n \to \infty$ for sufficiently large M. Similarly, setting $\mathcal{B}_n = \{y \in \mathcal{Y} : F^*(y) - b_n M > 0\}$,

$$\begin{aligned}
&P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}} \hat{F}^*(y) < F^*(y) - b_n M\right) \leq P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}\cap\mathcal{B}_n} \hat{Q}^*(F^*(y) - b_n M) > y\right)\\
\leq &P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}\cap\mathcal{B}_n} \|Q^* - \hat{Q}^*\|_\infty + Q^*(F^*(y) - b_n M) > y\right)\\
\leq &P\left(\bigcup_{y\in\mathcal{Y}\cap\mathbb{Q}\cap\mathcal{B}_n} \|Q^* - \hat{Q}^*\|_\infty + y - L_0^{-1} b_n M > y\right)\\
\leq &P\left(\|Q^* - \hat{Q}^*\|_\infty > L_0^{-1} b_n M\right) = o(1),
\end{aligned}$$

as $n \to \infty$ and sufficiently large M. Combining this with (68) and (69) leads to the result. □

Appendix B: Additional simulation results

S.6 Additional simulation results

We investigate the situation when only scatterplot data (X_i, Y_i), $i = 1, \ldots, n$, are available but the conditional densities f_i remain unobserved. For this, we first generate predictors $X_1, \ldots, X_n \overset{iid}{\sim} f$, where f corresponds to the density of a $U(0,1)$ random variate. Then, conditional on X_i, we generate Y_i from a mixture Gaussian distribution as follows: We first draw a uniform variate $p \sim U(0,1)$ independently of all other random quantities. If $p \geq 0.5$, we sample Y_i from a normal distribution $N(\mu_1, \sigma_1)$ with $\mu_1 = X_i$ and $\sigma_1 = 0.1$. Otherwise, if $p < 0.5$, we sample Y_i from a $N(\mu_2, \sigma_2)$ with $\mu_2 = 1 - X_i$ and $\sigma_2 = 0.1$. Thus, the underlying conditional density function $f_{Y|X}$ is given by the Gaussian mixture

$$f_{Y|X}(y, x) = \pi_1 \frac{1}{\sigma_1(x)} \phi\left(\frac{y - \mu_1(x)}{\sigma_1(x)}\right) + \pi_2 \frac{1}{\sigma_2(x)} \phi\left(\frac{y - \mu_2(x)}{\sigma_2(x)}\right), \tag{70}$$

where ϕ corresponds to the density of a standard normal random variable, $\sigma_j(x) = 0.1$, $\pi_j = 0.5$, $j = 1, 2$, $\mu_1(x) = x$ and $\mu_2(x) = 1 - x$, $x \in \mathcal{X} = [0, 1]$. Figure 7 shows the conditional Gaussian mixture densities over a dense grid of x values along with the density of the Wasserstein and classical marginal measures.

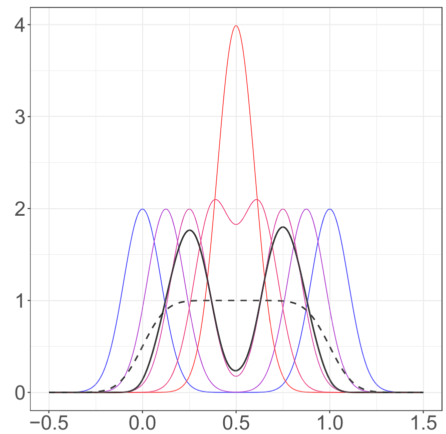

FIG 7. *Conditional Gaussian mixture densities $f_{Y|X}(\cdot, x)$ as in (70) over a dense grid of increasing x values in $[0, 0.5]$ (from blue to red) as outlined in the simulation setting for Gaussian mixtures with $X \sim U(0,1)$. Here $\mu_1(x) = x$, $\mu_2(x) = 1 - x$, $\sigma_j(x) = 0.1$ and $\pi_j = 0.5$, $j = 1, 2$. By symmetry, if $x \in (1/2, 1]$ then $f_{Y|X}(\cdot, x) = f_{Y|X}(\cdot, 1 - x)$ with $1 - x \in [0, 1/2)$. The density of the Wasserstein-Fréchet Integral is shown in solid black while the classical marginal is dashed.*

We take κ_1 and κ_2 to be Gaussian and Epanechnikov kernel functions, respectively, bandwidth sequences $h_1 = h = n^{-1/3}$ and compute $\hat{Q}^*$ over a dense grid of values in $[0.05, 0.95]$. Since the quantile function of a Gaussian mixture has no closed form expression, we obtain the true quantile function of the Wasserstein-Fréchet Integral measure by using that $Q^*(t) = \int_{\mathcal{X}} Q_{Y|X}(t, x) f(x) dx$ and numerically approximating this term. To assess the finite sample performance of the empirical estimates, we utilize the squared Wasserstein error measure

$$\mathcal{D}_{n,\alpha} = \int_{\alpha}^{1-\alpha} (\hat{Q}^*(t) - Q^*(t))^2 dt, \quad \alpha \in (0, 1/2),$$

which is obtained numerically over a dense grid in $[\alpha, 1 - \alpha]$. We set $\alpha = 0.05$. Figure 8 shows the boxplots of $\mathcal{D}_{n,\alpha}$ across 1000 simulations, and $\mathcal{D}_{n,\alpha}$ is seen to rapidly converge to zero.

For the previous Gaussian mixture scatterplot setting, we further explore the situation when X has a beta instead of a uniform distribution, where we expect the shape of the density of the Wasserstein-Fréchet Integral to be closer to the conditional densities $f_{Y|X}(\cdot, x)$ especially for larger x. For this, we slightly adjust the previous Gaussian mixture setting where we now set $\mu_1(x) = 0$ and $f \sim \text{Beta}(2, 2)$ while all other population variables remain the same as before. Figure 9 shows the conditional Gaussian mixture densities over a dense grid of x values along with the density of the Wasserstein and classical marginal measures. For sufficiently large values of x, the conditional density $f_{Y|X}(\cdot, x)$ is bimodal with peaks at 0 and $1 - x$ while for smaller values of x the conditional density converges to a zero-mean Gaussian density. Since $X \sim \text{Beta}(2, 2)$, higher weight

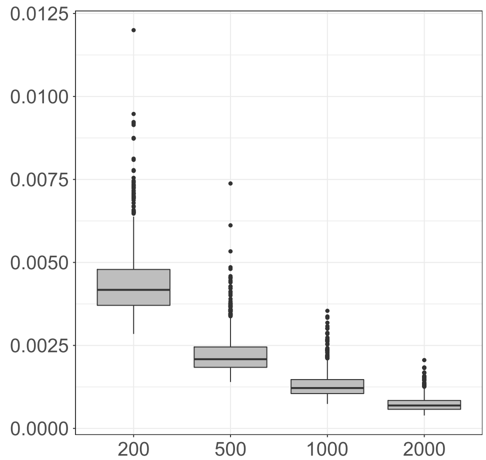

FIG 8. *Boxplot of the squared Wasserstein error measure* $\mathcal{D}_{n,\alpha} = \int_{\alpha}^{1-\alpha}(\hat{Q}^*(t) - Q^*(t))^2 dt$ *with* $\alpha = 0.05$ *for* 1000 *simulations and increasing sample sizes in a simulation setting with Gaussian mixtures when* $X \sim U(0,1)$.

values $f(x)$ are concentrated around $x = 0.5$ and thus the Wasserstein-Fréchet Integral should be closer to a bimodal Gaussian variate which is clearly seen to be the case in Figure 9. In contrast, the classical marginal density does not represent the vertical variation around the means. Figure 10 shows the boxplots for $\mathcal{D}_{n,\alpha}$, $\alpha = 0.05$, which again converge rapidly to zero with increasing sample size, indicating consistency of the Wasserstein-Fréchet Integral. The associated transports are shown in Figure 11.

Acknowledgments

One of the data sets used for illustration of the proposed methodology is from the first public release of the NIH program on Environmental Influences on Child Health Outcomes (ECHO). In this context, we note that the content of this paper is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the ECHO Cohort investigators. We also acknowledge NICHD DASH for providing these publicly available data and are indebted to the Associate Editor and two referees for providing insightful comments that led to numerous improvements.

Funding

Research supported in part by NSF Grants DMS-2014626 and DMS-2310450.

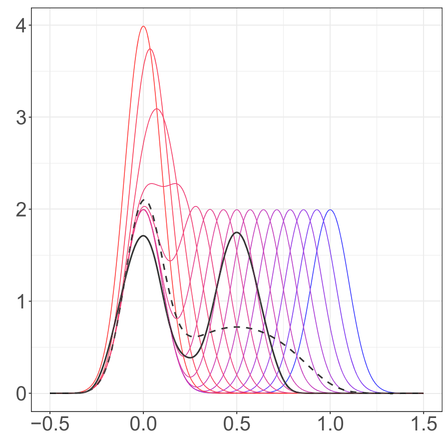

FIG 9. *Conditional Gaussian mixture densities $f_{Y|X}(\cdot, x)$ as in (70) over a dense grid of increasing x values in $[0,1]$ (from blue to red) as outlined in the simulation setting for Gaussian mixtures with $X \sim Beta(2,2)$. Here $\mu_1(x) = 0$, $\mu_2(x) = 1-x$, $\sigma_j(x) = 0.1$ and $\pi_j = 0.5$, $j = 1,2$. The density of the Wasserstein-Fréchet Integral is shown in solid black while the classical marginal is dashed.*

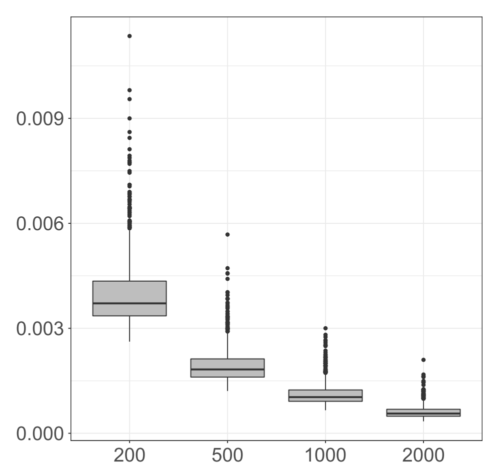

FIG 10. *Boxplot of the squared Wasserstein error measure $\mathcal{D}_{n,\alpha} = \int_{\alpha}^{1-\alpha}(\hat{Q}^*(t) - Q^*(t))^2 dt$ with $\alpha = 0.05$ for 1000 simulations and increasing sample sizes in a simulation setting with Gaussian mixtures when $X \sim Beta(2,2)$.*

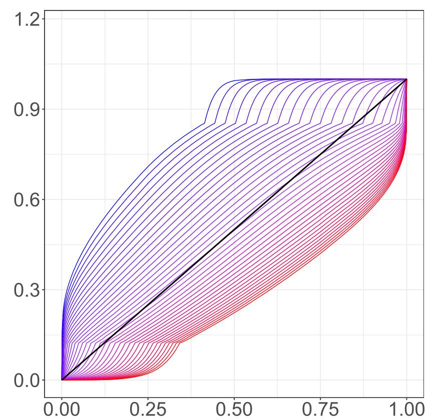

Fig 11. *Individual optimal transports $T_i = Q^* \circ F_i$ from the conditional distributions ν_i to the Wasserstein measure ν^* in the simulation setting of Figure 1. Low values of the predictor x are displayed in blue and higher values in red. The identity map is displayed in solid black.*

References

Agueh, M. and Carlier, G. (2011). Barycenters in the Wasserstein space. *SIAM Journal on Mathematical Analysis* **43** 904–924. MR2801182

Álvarez Esteban, P. C., del Barrio, E., Cuesta-Albertos, J. A. and Matrán, C. (2016). A fixed-point approach to barycenters in Wasserstein space. *Journal of Mathematical Analysis and Applications* **441** 744–762. MR3491556

Ambrosio, L., Gigli, N. and Savaré, G. (2005). *Gradient Flows: in Metric Spaces and in the Space of Probability Measures*. Springer Science & Business Media. MR2129498

Bhattacharya, P. K. and Gangopadhyay, A. K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *The Annals of Statistics* **18** 1400–1415. MR1062716

Bhattacharya, R. and Patrangenaru, V. (2005). Large sample theory of intrinsic and extrinsic sample means on manifolds - II. *The Annals of Statistics* **33** 1225–1259. MR2195634

Bigot, J. (2013). Fréchet means of curves for signal averaging and application to ECG data analysis. *The Annals of Applied Statistics* **7** 2384–2401. MR3161727

Bigot, J. and Charlier, B. (2011). On the consistency of Fréchet means in deformable models for curve and image analysis. *Electronic Journal of Statistics* **5** 1054–1089. MR2836769

Bigot, J. and Klein, T. (2018). Characterization of barycenters in the Wasserstein space by averaging optimal transport maps. *ESAIM: PS* **22** 35–57. MR3872127

Bigot, J., Gouet, R., Klein, T. and López, A. (2018). Upper and lower risk bounds for estimating the Wasserstein barycenter of random measures on the real line. *Electronic Journal of Statistics* **12** 2253–2289. MR3830834

Bobkov, S. and Ledoux, M. (2019). *One-Dimensional Empirical Measures, Order Statistics, and Kantorovich Transport Distances* **261**. American Mathematical Society. MR4028181

Bolstad, B. M., Irizarry, R. A., Åstrand, M. and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19** 185–193. MR2702454

Bonnéry, D., Breidt, F. J. and Coquet, F. (2012). Uniform convergence of the empirical cumulative distribution function under informative selection from a finite population. *Bernoulli* **18** 1361–1385. MR2995800

Carroll, C., Bhattacharjee, S., Chen, Y., Dubey, P., Fan, J., Gajardo, A., Zhou, X., Müller, H.-G. and Wang, J.-L. (2020). Time Dynamics of COVID-19. *Scientific Reports* **10** 1–14.

Chen, Y., Dawson, M. and Müller, H.-G. (2020). Rank dynamics for functional data. *Computational Statistics & Data Analysis* **149** 106963. MR4091224

Chen, H. and Müller, H.-G. (2023). Sliced Wasserstein Regression. *arXiv preprint arXiv:2306.10601*.

Chen, Y., Álvaro Gajardo, Fan, J., Zhong, Q., Dubey, P., Han, K., Bhattacharjee, S. and Müller, H.-G. (2020). Fréchet: statistical analysis for random objects and non-Euclidean data R package version 0.2.0.

Chernozhukov, V., Fernández-Val, I. and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica* **78** 1093–1125. MR2667913

Dai, X. and Lopez-Pintado, S. (2023). Tukey's depth for object data. *Journal of the American Statistical Association* **118** 1760–1772. MR4646604

del Barrio, E., González-Sanz, A. and Hallin, M. (2020). A note on the regularity of optimal-transport-based center-outward distribution and quantile functions. *Journal of Multivariate Analysis* **180** 104671. MR4147635

del Barrio, E., González-Sanz, A. and Loubes, J.-M. (2024). Central limit theorems for general transportation costs. *Annales de l'Institut Henri Poincaré (B) Probabilites et statistiques* **60** 847–873. MR4757510

del Barrio, E., Sanz, A. G. and Hallin, M. (2024). Nonparametric multiple-output center-outward quantile regression. *Journal of the American Statistical Association* 1–15.

Dubey, P. and Müller, H.-G. (2020). Functional models for time-varying random objects. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **82** 275–327. MR4084166

Embrechts, P. and Hofert, M. (2013). A note on generalized inverses. *Mathematical Methods of Operations Research* **77** 423–432. MR3072795

Falk, M. (1984). Relative deficiency of kernel type estimators of quantiles. *The Annals of Statistics* **12** 261–268. MR0733512

Falk, M. (1985). Asymptotic normality of the kernel quantile estimator. *The Annals of Statistics* **13** 428–433. MR0773180

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applica-*

tions. Chapman & Hall, London. MR1383587

Fan, J., Hu, T.-C. and Truong, Y. K. (1994). Robust non-parametric function estimation. *Scandinavian Journal of Statistics* **21** 433–446. MR1310087

Feng, C., Wang, H., Tu, X. M. and Kowalski, J. (2012). A note on generalized inverses of distribution function and quantile transformation. *Applied Mathematics* **3** 2098–2100.

Fréchet, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. In *Annales de l'Institut Henri Poincaré* **10** 215–310. MR0027464

Gajardo, Á. and Müller, H.-G. (2021). Cox point pocess regression. *IEEE Transactions on Information Theory* **68** 1133–1156. MR4395462

Gajardo, A. and Müller, H. G. (2022). Point process models for COVID-19 cases and deaths. *Journal of Applied Statistics* **50** 8–19. MR4620435

Gangbo, W. and McCann, R. J. (1996). The geometry of optimal transportation. *Acta Mathematica* **177** 113–161. MR1440931

Gervini, D. and Khanal, M. (2019). Exploring patterns of demand in bike sharing systems via replicated point process models. *Journal of the Royal Statistical Society Series C: Applied Statistics* **68** 585–602. MR3937464

Hall, P., Wolff, R. C. L. and Yao, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* **94** 154–163. MR1689221

Hallin, M., del Barrio, E., Cuesta-Albertos, J. and Matrán, C. (2021). Distribution and quantile functions, ranks and signs in dimension d: A measure transportation approach. *The Annals of Statistics* **49** 1139–1165. MR4255122

Heinemann, F. and Bonneel, N. (2021). WSGeometry: Geometric tools based on balanced/unbalanced optimal transport R package version 1.2.1.

Jones, M. C. and Hall, P. (1990). Mean squared error properties of kernel estimates or regression quantiles. *Statistics & Probability Letters* **10** 283–289. MR1069903

Kneip, A. and Gasser, T. (1992). Statistical tools to analyze data representing a sample of curves. *The Annals of Statistics* **20** 1266–1305. MR1186250

Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica* **46** 33–50. MR0474644

Koenker, R. and Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives* **15** 143–156. MR2268657

Kolouri, S., Rohde, G. K. and Hoffmann, H. (2018). Sliced Wasserstein distance for learning gaussian mixture models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3427–3436.

Le Gouic, T. and Loubes, J.-M. (2017). Existence and consistency of Wasserstein barycenters. *Probability Theory and Related Fields* **168** 901–917. MR3663634

Li, Q. and Racine, J. S. (2008). Nonparametric estimation of conditional CDF and quantile functions with mixed categorical and continuous data. *Journal of Business & Economic Statistics* **26** 423–434. MR2459343

Lin, Z. and Müller, H.-G. (2021). Total variation regularized Fréchet regression for metric-space valued data. *The Annals of Statistics* **49** 3510–3533. MR4352539

Liu, X. and Müller, H.-G. (2004). Functional convex averaging and synchronization for time-warped random curves. *Journal of the American Statistical Association* **99** 687–699. MR2090903

Panaretos, V. M. and Zemel, Y. (2019). Statistical aspects of Wasserstein distances. *Annual Review of Statistics and Its Application* **6** 405–431. MR3939527

Parzen, E. (1979). Nonparametric statistical data modeling. *Journal of the American Statistical Association* **74** 105–121. MR0529528

Pass, B. (2013). Optimal transportation with infinitely many marginals. *Journal of Functional Analysis* **264** 947–963. MR3004954

Petersen, A. and Müller, H.-G. (2016a). Fréchet integration and adaptive metric selection for interpretable covariances of multivariate functional data. *Biometrika* **103** 103–120. MR3465824

Petersen, A. and Müller, H.-G. (2016b). Functional data analysis for density functions by transformation to a Hilbert space. *The Annals of Statistics* **44** 183-218. MR3449766

Petersen, A. and Müller, H.-G. (2019a). Fréchet regression for random objects with Euclidean predictors. *The Annals of Statistics* **47** 691–719. MR3909947

Petersen, A. and Müller, H.-G. (2019b). Wasserstein covariance for multiple random densities. *Biometrika* **106** 339–351. MR3949307

Sakoe, H. and Chiba, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* **26** 43–49.

Samanta, M. (1989). Non-parametric estimation of conditional quantiles. *Statistics & Probability Letters* **7** 407–412. MR1001144

Segers, J. (2022). Graphical and uniform consistency of estimated optimal transport plans. *arXiv preprint arXiv:2208.02508*.

Stone, C. J. (1977). Consistent nonparametric regression. *The Annals of Statistics* **5** 595–620. MR0443204

Stute, W. (1986). Conditional empirical processes. *The Annals of Statistics* **14** 638–647. MR0840519

Takatsu, A. (2011). Wasserstein geometry of Gaussian measures. *Osaka Journal of Mathematics* **48** 1005–1026. MR2871291

Villani, C. (2003). *Topics in Optimal Transportation*. American Mathematical Society. MR1964483

Villani, C. (2009). *Optimal Transport: Old and New*. Springer. MR2914136

Wang, K. and Gasser, T. (1997). Alignment of curves by dynamic time warping. *The Annals of Statistics* **25** 1251–1276. MR1447750

Yu, K. and Jones, M. C. (1998). Local linear quantile regression. *Journal of the American Statistical Association* **93** 228–237. MR1614628

Zhang, X. and Wang, J.-L. (2016). From sparse to dense functional data and beyond. *The Annals of Statistics* **44** 2281–2321. MR3546451