WASSERSTEIN F-TESTS AND CONFIDENCE BANDS FOR THE FRÉCHET REGRESSION OF DENSITY RESPONSE CURVES

BY ALEXANDER PETERSEN[1,*], XI LIU[1,†] AND AFSHIN A. DIVANI[2]

[1]*Department of Statistics and Applied Probability, University of California, Santa Barbara, *petersen@pstat.ucsb.edu;
†xliu@pstat.ucsb.edu*

[2]*Department of Neurology, University of New Mexico, adivani@salud.unm.edu*

Data consisting of samples of probability density functions are increasingly prevalent, necessitating the development of methodologies for their analysis that respect the inherent nonlinearities associated with densities. In many applications, density curves appear as functional response objects in a regression model with vector predictors. For such models, inference is key to understand the importance of density-predictor relationships, and the uncertainty associated with the estimated conditional mean densities, defined as conditional Fréchet means under a suitable metric. Using the Wasserstein geometry of optimal transport, we consider the Fréchet regression of density curve responses and develop tests for global and partial effects, as well as simultaneous confidence bands for estimated conditional mean densities. The asymptotic behavior of these objects is based on underlying functional central limit theorems within Wasserstein space, and we demonstrate that they are asymptotically of the correct size and coverage, with uniformly strong consistency of the proposed tests under sequences of contiguous alternatives. The accuracy of these methods, including nominal size, power and coverage, is assessed through simulations, and their utility is illustrated through a regression analysis of post-intracerebral hemorrhage hematoma densities and their associations with a set of clinical and radiological covariates.

1. Introduction. Samples of probability density functions arise naturally in many modern data analysis settings, including population age and mortality distributions across different countries or regions (Bigot et al. (2017), Hron et al. (2016), Petersen and Müller (2019)), and distributions of functional connectivity patterns in the brain (Petersen and Müller (2016)). Methods for the analysis of density data began with the work of Kneip and Utikal (2001), who applied standard functional principal components analysis (FPCA) in order to quantify a mean density and prominent modes of variability about that mean. However, due to inherent constraints of density functions, which must be nonnegative and integrate to one, nonlinear methods are steadily replacing such standard procedures. For example, Hron et al. (2016) and Petersen and Müller (2016) both proposed to apply a preliminary transformation to the densities, mapping them into a Hilbert space, after which linear methods such as FPCA can be suitably applied. The transformation of Hron et al. (2016) was specifically motivated by the extension of the Aitchison geometry (Aitchison (1986)) to infinite-dimensional compositional data due to the work of Egozcue, Díaz-Barrero and Pawlowsky-Glahn (2006), while those in Petersen and Müller (2016) were generic and not motivated by any particular geometry.

Parallel developments in the analysis of nonlinear data have been made in the broader field of object-oriented data analysis (Marron and Alonso (2014)), where a complex data space is

endowed with a chosen metric that, in turn, defines the parameters of the model. Particular attention has been paid to manifold-valued data (e.g., Fletcher et al. (2004), Panaretos, Pham and Yao (2014)), with the main objects of interest being the Fréchet mean and variance, as well as dimension reduction tools designed to optimally retain variability in the data (Patrangenaru and Ellingson (2016)). Within this context, Srivastava, Jermyn and Joshi (2007) and Bigot et al. (2017) developed manifold-based dimension reduction techniques specifically designed for samples of density functions, the former utilizing the Fisher–Rao geometry and the latter the Wasserstein geometry of optimal transport. Of these two, the Wasserstein metric has proved in recent years to have greater appeal both theoretically, given its clear interpretation as an optimal transport cost (Ambrosio, Gigli and Savaré (2008), Villani (2003)), as well as in applied settings (Bolstad et al. (2003), Broadhurst et al. (2006), Panaretos and Zemel (2016), Zhang and Müller (2011)).

In this paper, we study a regression model with density functions as response variables under the Wasserstein geometry, with predictors being Euclidean vectors. Such data are frequently encountered in practice (e.g., Nerini and Ghattas (2007), Talská et al. (2018)). The goal of the model is to perform inference, specifically tests for covariate effects and confidence bands for the fitted conditional mean densities. For this reason, we assume a global regression model that does not require any smoothing or other tuning parameter to fit. Standard functional response models, such as the linear model (Faraway (1997)), are not suitable for the nonlinear space of density functions unless the densities are first transformed into a linear space, as demonstrated recently in Talská et al. (2018) using the compositional approach. However, there is no such transformation that is suitable for the Wasserstein geometry. Global regression models for Riemannian manifolds have been developed (Cornea et al. (2017), Fletcher (2013), Niethammer, Huang and Vialard (2011)), but are also not directly applicable to the Wasserstein geometry. Instead, we will develop our inferential techniques under the Fréchet regression model proposed in Petersen and Müller (2019), which defines a global regression function between response data in an arbitrary metric space and vector predictors. Although the theory of estimation under this model was well studied in the general metric space framework, this is not the case for other forms of inference.

In Section 2, we briefly describe the necessary components of the Wasserstein geometry and its implications when applied to the Fréchet regression model. An additional component not considered in Petersen and Müller (2019) that is available through this particular formulation is the random optimal transport map between the conditional Wasserstein mean density and the observed one. These maps serve the purpose of the error term in the regression model, although they do not act additively or even linearly. In Section 3, we develop intuitive test statistics for covariate effects and derive their asymptotic distributions, leading to root-n consistent testing procedures. We also describe different methods for implementing these tests, including a bootstrap procedure involving residual optimal transport maps obtained from the fitted model. Section 4 demonstrates how to compute asymptotic confidence bands about the estimated conditional mean distributions. Finally, these methods are illustrated through extensive simulations (Section 5) and the analysis of distributions of hematoma density for stroke patients (Section 6), as captured by computed tomography scans, and their dependency on patient-specific covariates (Figure 1).

2. Preliminaries. We begin with a brief definition of the Wasserstein metric in the language of optimal transport. This Wasserstein metric has been known under other names, including "Mallows" and "earth-mover's" (Levina and Bickel (2001)), and its use in statistics is rapidly expanding (Panaretos and Zemel (2019)). Let $\mathcal{D}$ be the class of univariate probability density functions f that satisfy $\int_{\mathbb{R}} u^2 f(u) \, du < \infty$, that is, absolutely continuous distributions on with a finite second moment. In fact, the Wasserstein metric is well defined
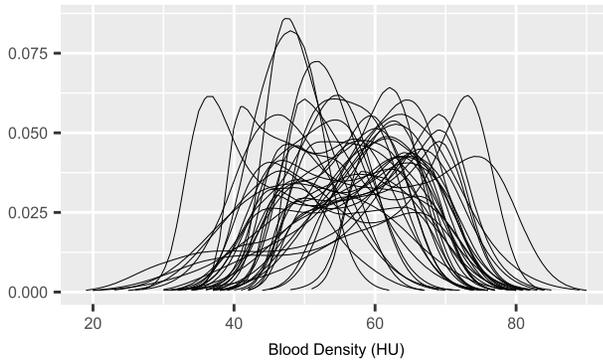
FIG. 1. *Observed hematoma density distributions for a randomly selected subset of* 40 *post-intracerebral hemmorhage patients.*

for such distributions without requiring a density, but for simplicity of presentation we deal specifically with this subclass. For a comprehensive treatment in the more general case, see Villani (2003) or Ambrosio, Gigli and Savaré (2008). For $f, g \in \mathcal{D}$, consider the collection of maps $\mathcal{M}_{f,g}$ such that, if $U \sim f$ and $M \in \mathcal{M}_{f,g}$, then $M(U) \sim g$. The squared Wasserstein distance between these two distributions is

$$d_W^2(f, g) = \inf_{M \in \mathcal{M}_{f,g}} \int_{\mathbb{R}} (M(u) - u)^2 f(u) \, du.$$

It is well known that the infimum above is attained by the optimal transport map $M_{f,g}^{\mathrm{opt}} = G^{-1} \circ F$, where F and G are the cumulative distribution functions of f and g, respectively, leading to the closed forms

$$(2.1) \qquad d_W^2(f, g) = \int_{\mathbb{R}} (M_{f,g}^{\mathrm{opt}}(u) - u)^2 f(u) \, du = \int_0^1 (F^{-1}(t) - G^{-1}(t))^2 \, dt,$$

where the last equality follows by the change of variables $t = F(u)$. A more proper term for this metric is the Wasserstein-2 distance, since it is just one among an entire class of Wasserstein metrics.

Within this larger class of metrics is also the Wasserstein-∞ metric, which will be useful in the formation of confidence bands. For two densities $f, g \in \mathcal{D}$, their Wasserstein-∞ distance is

$$(2.2) \qquad d_\infty(f, g) = f\text{-}\mathrm{ess\,sup}_{u \in \mathcal{I}} |M_{f,g}^{\mathrm{opt}}(u) - u|,$$

where f-$\mathrm{ess\,sup}$ refers to the essential supremum with respect to f.

2.1. *Random densities.* A random density $\mathfrak{F}$ is a random variable taking its values almost surely in $\mathcal{D}$. It will also be useful to refer to the corresponding random cdf F, quantile function $Q = F^{-1}$ and quantile density $q = Q' = 1/(\mathfrak{F} \circ Q)$ (Parzen (1979)). For clarity, $u, v \in \mathbb{R}$ will consistently be used as density and cdf arguments throughout, whereas $s, t \in [0, 1]$ will be used as arguments for the quantile and quantile density functions. Following the ideas of Fréchet (1948), the Wasserstein–Fréchet (or simply Wasserstein) mean and variance of $\mathfrak{F}$ are

$$(2.3) \qquad f_\oplus^* := \underset{f \in \mathcal{D}}{\mathrm{argmin}} \, E(d_W^2(\mathfrak{F}, f)), \qquad \mathrm{Var}_\oplus(\mathfrak{F}) := E(d_W^2(\mathfrak{F}, f_\oplus^*)).$$

In the regression setting, we model the distribution of $\mathfrak{F}$ conditional on a vector $X \in \mathbb{R}^p$ of predictors, where the pair $(X, \mathfrak{F})$ is distributed according to a probability measure $\mathcal{G}$ on the

product space $\mathbb{R}^p \times \mathcal{D}$. In this sense, the objects in (2.3) are the marginal Fréchet mean and variance of $\mathfrak{F}$. Let $\mathfrak{S}_X$ denote the support of the marginal distribution of X. Our interest is in the Fréchet regression function, or function of conditional Fréchet means,

$$
(2.4) \qquad f_\oplus(x) := \underset{f \in \mathcal{D}}{\arg\min} \, E\big[d_W^2(\mathfrak{F}, f)|X = x\big], \quad x \in \mathfrak{S}_X.
$$

Let $F_\oplus(x)$, $Q_\oplus(x)$, and $q_\oplus(x)$ denote, respectively, the cdf, quantile and quantile density functions corresponding to $f_\oplus(x)$. We will use the notation $f_\oplus(x, u)$ to denote the value of the conditional mean density $f_\oplus(x)$ at argument $u \in \mathbb{R}$, and similarly for $F_\oplus(x, u)$, $Q_\oplus(x, t)$, and $q_\oplus(x, t)$, $t \in [0, 1]$. For a pair $(X, \mathfrak{F})$, define $T = Q \circ F_\oplus(X)$ to be the optimal transport map from the conditional mean $f_\oplus(X)$ to the random density $\mathfrak{F}$. By (2.1), it must be that $E(Q(t)|X = x) = Q_\oplus(x, t)$, so that $E(T(u)|X = x) = u$ for u such that $f_\oplus(x, u) > 0$. Then the conditional Fréchet variance is

$$
(2.5) \qquad \begin{aligned}
\mathrm{Var}_\oplus(\mathfrak{F}|X = x) &= E\big[d_W^2(\mathfrak{F}, f_\oplus(x))|X = x\big] \\
&= \int_\mathbb{R} E\big[(T(u) - u)^2|X = x\big] f_\oplus(x, u) \, du \\
&= \int_\mathbb{R} \mathrm{Var}(T(u)|X = x) f_\oplus(x, u) \, du.
\end{aligned}
$$

In these developments, we have assumed that the marginal and conditional Wasserstein mean densities exist and are unique. However, this is not automatic and some conditions are needed. Previous work on existence, uniqueness and regularity of Wasserstein means, or barycenters, for a finite collection of probability meaures on $\mathbb{R}^d$ was done by Agueh and Carlier (2011), and extended to continuously-indexed measures by Pass (2013). For random probability measures with support on [0, 1], Panaretos and Zemel (2016) (see Proposition 2 therein) gave sufficient conditions for the existence and uniqueness of a Wasserstein mean measure, although it was not guaranteed to have a density. However, none of these are sufficiently strong for the purposes of this paper, where existence, uniqueness and regularity of both marginal (2.3) and conditional (2.4) Wasserstein means are needed. To this end, consider the following assumptions on the joint distribution $\mathcal{G}$.

(A1) $\mathfrak{F}(u) \in (0, \infty)$ for $u \in (Q(0), Q(1))$ almost surely, $\mathrm{Var}(Q(t)) < \infty$ for all $t \in (0, 1)$, and $\int_0^1 \mathrm{Var}(Q(t)) \, dt < \infty$.

(A2) For any $t \in (0, 1)$, there exists δ, $0 < \delta < \min\{t, 1-t\}$ such that $E(\sup_{|s-t|<\delta} q(s)) < \infty$.

(A3) For all $x \in \mathfrak{S}_X$, $P(\sup_{u \in (Q(0), Q(1))} \mathfrak{F}(u) < \infty|X = x) > 0$.

Assumption (A1) is essentially the same as that made in Proposition 2 of Panaretos and Zemel (2016), with additional moment assumptions on Q since $\mathfrak{F}$ is not assumed to be supported on any bounded interval, and implies existence and uniqueness of the Wasserstein mean measures. Assumption (A2) is a regularity condition to ensure these Wasserstein means have densities in $\mathcal{D}$, and (A3) implies that these mean densities are bounded; see Pass (2013) for a similar assumption. The proof of the following and all other theoretical results can be found in Petersen, Liu and Divani (2020).

PROPOSITION 1. *Suppose conditions* (A1)–(A3) *hold. Then the marginal Wasserstein mean* $f_\oplus^*$ *in* (2.3) *exists, is a unique element of* $\mathcal{D}$, *and satisfies* $\sup_u f_\oplus^*(u) < \infty$ *and* $\mathrm{Var}_\oplus(\mathfrak{F}) < \infty$. *For all* $x \in \mathfrak{S}_X$, *the conditional Wasserstein mean* $f_\oplus(x)$ *in* (2.4) *exists, is a unique element of* $\mathcal{D}$, *and satisfies* $\sup_u f_\oplus(x, u) < \infty$ *and* $\mathrm{Var}_\oplus(\mathfrak{F}|X = x) < \infty$.

2.2. *Global Wasserstein–Fréchet regression.* In order to facilitate inference, specifically tests for no or partial effects of the covariates X and confidence bands for the conditional Wasserstein means, we consider a particular global regression model for the conditional Wassersteins $f_\oplus(x)$ defined in (2.4). This model, proposed by Petersen and Müller (2019), is termed Fréchet regression, and takes the form of a weighted Fréchet mean

$$(2.6) \qquad f_\oplus(x) = \underset{f \in \mathcal{D}}{\operatorname{argmin}}\, E\big[s(X, x)d_W^2(\mathfrak{F}, f)\big],$$

where the weight function is

$$s(X, x) = 1 + (X - \mu)^\top \Sigma^{-1}(x - \mu), \quad \mu = E(X), \Sigma = \operatorname{Var}(X),$$

and Σ is assumed to be positive definite. The model is motivated by multiple linear regression, and is its direct generalization to the case of a response variable in a metric space. Specifically, if a scalar response Y is jointly distributed with X and $E(Y|X = x)$ is linear in x, Petersen and Müller (2019) showed that an alternative characterization of linear regression is

$$E(Y|X = x) = \underset{y \in \mathbb{R}}{\operatorname{argmin}}\, E\big[s(X, x)(Y - y)^2\big].$$

Thus, model (2.6) generalizes linear regression to the case of density response by substituting $\mathfrak{F}$ for Y and $(\mathcal{D}, d_W)$ in place of the usual metric space $(\mathbb{R}, |\cdot|)$ implicitly used in multiple linear regression. Although $\mathcal{D}$ is not a linear space, (2.6) provides a sensible regression model for the conditional Wasserstein means that retains some properties of linear regression. For example, since $s(z, \mu) \equiv 1$, we have $f_\oplus(\mu) = f_\oplus^*$, so that the regression function passes through the point $(\mu, f_\oplus^*)$.

For the remainder of the paper, and implicit in the statement of all theoretical results, we assume that the distribution $\mathcal{G}$ satisfies model (2.6), with $f_\oplus^*$, $f_\oplus(x)$ being unique elements of $\mathcal{D}$. We now give a very basic example of this model, motivated by the well-known connection between the Wasserstein metric and location-scale families (e.g., Bigot et al. (2017)).

EXAMPLE 1. Suppose $X \in \mathbb{R}$ and fix $f_0 \in \mathcal{D}$. Letting $a_j, b_j \in \mathbb{R}$, define

$$v(x) = a_0 + a_1 x, \qquad \tau(x) = b_0 + b_1 x,$$

and suppose $\tau(X) > 0$ almost surely. Let V_1, V_2 satisfy $E(V_1|X) = 0$, $E(V_2|X) = 1$, and $V_2 > 0$ almost surely. Then the location-scale model

$$\mathfrak{F}(u) = \frac{1}{V_2\tau(X)} f_0\left(\frac{u - V_1 - V_2 v(X)}{V_2\tau(X)}\right)$$

corresponds to (2.6) with

$$f_\oplus(x, u) = \frac{1}{\tau(x)} f_0\left(\frac{u - v(x)}{\tau(x)}\right).$$

To show the above, it is sufficient to show that $E(Q(t)|X) = Q_\oplus(X, t)$ almost surely. Let F_0 be the cdf corresponding to f_0, so that $Q_\oplus(x, t) = v(x) + \tau(x)Q_0(t)$. Since

$$Q(t) = V_1 + V_2 v(X) + V_2\tau(X)Q_0(t),$$

it is easily verified that $E(Q(t)|X) = Q_\oplus(X, t)$ by the properties of V_1, V_2. Moreover, note that the optimal transport map from $f_\oplus(X)$ to $\mathfrak{F}$ is

$$T(u) = Q \circ F_\oplus(X, u) = V_1 + V_2 u,$$

and satisfies $E(T(u)|X) = u$ almost surely.

Thus far, the regression model (2.6) provides us with a formula for the conditional Wasserstein mean of $\mathfrak{F}$, whereas one also needs information on the conditional variance in order to conduct inference. To this end, observe that $Q = T \circ Q_\oplus(X)$, so that T acts as a residual transport, although it acts on the quantile function, and not the density, and does so through composition and not additively. As observed in Section 2.1, the first-order behavior of the residual transport T is completely specified by the model as $E(T(u)|X) = u$ almost surely. We also impose the following assumption on the covariance.

(T1) The covariance function $C_T(u, v) = \text{Cov}(T(u), T(v))$ is continuous, $\sup_{u \in \mathbb{R}} \text{Var}(T(u)) < \infty$, and $\text{Cov}(T(u), T(v)|X) = C_T(u, v)$ almost surely.

This corresponds to the classical constant variance or exogeneity assumption requiring that the second-order behavior of the residual transport be independent of the predictors. Define $S = Q_\oplus(X) \circ F_\oplus^*$, which is the optimal transport map from the marginal Wasserstein mean to the conditional one. Again, it is easily verified that $E(T \circ S(u)) = u$ for u such that $f_\oplus^*(u) > 0$. These observations lead to the following decomposition of the Wasserstein variance.

PROPOSITION 2. *Suppose that assumption (T1) is satisfied. Then*

$$\text{Var}_\oplus(\mathfrak{F}) = E[\text{Var}_\oplus(\mathfrak{F}|X)] + \text{Var}_\oplus(f_\oplus(X))$$

(2.7)
$$= \int_{\mathbb{R}} E[C_T(S(u), S(u))] f_\oplus^*(u) \, du$$

$$+ \int_{\mathbb{R}} (S(u) - u)^2 f_\oplus^*(u) \, du.$$

While a standard result in Euclidean spaces, the above variance decomposition does not generally hold in metric spaces such as $\mathcal{D}$. However, Proposition 2 demonstrates that this decomposition does indeed hold for random densities under the Wasserstein metric whenever model (2.6) holds. This finding motivates the specific choices of test statistics developed in Section 3.

2.3. *Estimation.* In order to estimate the regression function $f_\oplus(x)$, we utilize an empirical version of the least-squares Wasserstein criterion in (2.6). First, set $\bar{X} = n^{-1} \sum_{i=1}^n X_i$, $\hat{\Sigma} = n^{-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^\top$, and compute empirical weights $s_{in}(x) = 1 + (X_i - \bar{X})^\top \hat{\Sigma}^{-1}(x - \bar{X})$, Let $\mathfrak{Q}$ be the set of quantile functions in $L^2[0, 1]$. With $\|\cdot\|_{L^2}$ denoting the standard Hilbert norm on $L^2[0, 1]$, an estimator of $Q_\oplus(x)$ is

(2.8)
$$\hat{Q}_\oplus(x) = \underset{Q \in \mathfrak{Q}}{\text{argmin}} \sum_{i=1}^n s_{in}(x) \|Q - Q_i\|_{L^2}^2.$$

Implementation of this estimator is given in Algorithm 1 in Petersen, Liu and Divani (2020). In finite samples, $\hat{Q}_\oplus(x)$ will not necessarily admit a density. For almost all procedures described, this quantile estimate is sufficient, since it can be used to compute Wasserstein distances as in (2.1). Nevertheless, we will establish (see Lemma 2 in Petersen, Liu and Divani (2020)) that $\hat{Q}_\oplus(x)$ admits a density $\hat{f}_\oplus(x) \in \mathcal{D}$ for large samples with high probability. When this holds, the estimate

(2.9)
$$\hat{f}_\oplus(x) = \underset{f \in \mathcal{D}}{\text{argmin}} \frac{1}{n} \sum_{i=1}^n s_{in}(x) d_W^2(\mathfrak{F}_i, f)$$

is well defined, and $\hat{f}_\oplus(x)$ is the density corresponding to the quantile estimate $\hat{Q}_\oplus(x)$ above. It can be computed in practice using Algorithm 2 in Petersen, Liu and Divani (2020).

Since we will also consider hypothesis tests of partial effects, write $x \in \mathbb{R}^p$ as $x = (y, z)$, where y corresponds to the first q entries of x, $q < p$. Similarly, take $X = (Y, Z)$, $\mu = (\mu_Y, \mu_Z)$, and

$$\Sigma = \begin{pmatrix} \Sigma_{YY} & \Sigma_{YZ} \\ \Sigma_{ZY} & \Sigma_{ZZ} \end{pmatrix},$$

with corresponding notations for the partitions of $\bar{X}$ and $\hat{\Sigma}$. Consider the null hypothesis $\mathcal{H}_0^P : f_\oplus(x) = f_{0,\oplus}(y)$,

$$f_{0,\oplus}(y) := \underset{f \in \mathcal{D}}{\mathrm{argmin}}\, E[s_0(Y, y) d_W^2(\mathfrak{F}_i, f)], \quad s_0(Y, y) = 1 + (Y - \mu_Y) \Sigma_{YY}^{-1}(y - \mu_Y).$$

Then the restricted estimators $\hat{Q}_{0,\oplus}(y)$ and $\hat{f}_{0,\oplus}(y)$ are defined analogously to (2.8) and (2.9), only using submodel weights

$$s_{in,0}(y) = 1 + (Y - \bar{Y})^\top \hat{\Sigma}_{YY}^{-1}(y - \bar{Y}).$$

3. Hypothesis testing. Once estimation is under control, the first goal of any global regression model is to test for effects of the predictors. In the more abstract setting of a response variable in an arbitrary metric space, Petersen and Müller (2019) suggested a permutation approach based on a Fréchet generalization of the coefficient of determination R^2 in multiple linear regression, though the theoretical properties of this test were not investigated. In a recent preprint, Dubey and Müller (2019) developed a test statistic and its asymptotic distribution for the case of a random object response and categorical predictors, giving a Fréchet extension of analysis of variance. Given that we are considering the more specific case of density-valued response variables under the Wasserstein geometry, we are able to expand on these results in order to develop asymptotically justified tests for both global and partial null hypotheses, where predictors can be of any type.

3.1. *Test of no effects.* We begin with the global null hypothesis of no effects, $\mathcal{H}_0^G : f_\oplus(x) \equiv f_\oplus^*$. Given the Wasserstein variance decomposition in (2.7), under $\mathcal{H}_0^G$ we have $\mathrm{Var}_\oplus(f_\oplus(X)) = E(d_W^2(f_\oplus(X), f_\oplus^*)) = 0$. This motivates

$$(3.1) \qquad\qquad F_G^* = \sum_{i=1}^{n} d_W^2(\hat{\mathfrak{F}}_i, \hat{f}_\oplus^*)$$

as a test statistic, where $\hat{\mathfrak{F}}_i = \hat{f}_\oplus(X_i)$ are the fitted densities and $\hat{f}_\oplus^* = \hat{f}_\oplus(\bar{X})$ is the sample Wasserstein mean. This can be viewed as a generalization of the numerator of the global F-test in multiple linear regression, and we thus refer to F_G^* in (3.1) as the (global) Wasserstein F-statistic.

In order to establish the asymptotic null distribution of F_G^*, we require the following assumptions. Define $C_T^{(l,m)} = \frac{\partial^{l+m}}{\partial u^l \partial v^m} C_T$ and, for any $x \in \mathbb{R}^p$, set $\tilde{x} = (1 x^\top)^\top$. Let $R = T \circ S = Q \circ F_\oplus^*$ and $\mathcal{I}_x = (Q_\oplus(x, 0), Q_\oplus(x, 1))$.

(T2) T is differentiable almost surely, with T' having random Lipschitz constant L. The covariance function C_T has continuous partial derivatives $C_T^{(l,m)}$, for $l, m = 0, 1$, with $\sup_{u \in \mathbb{R}} C_T^{(1,1)}(u, u) < \infty$.

(T3) $E(\|X\|_E^4)$, $E(\|\tilde{X}\|_E^4 \sup_{u \in \mathbb{R}} |T'(u)|^2)$, and $E(\|\tilde{X}\|_E^6 L^2)$ are all finite.

(T4) $\mathcal{I}^* = \mathcal{I}_\mu$ is a bounded interval, and $\gamma(u) = \mathrm{Cov}(X, R(u))$ has bounded second derivative on $\mathcal{I}^*$. Also, $\inf_{u \in \mathcal{I}^*} f_\oplus^*(u) > 0$ and, for some $M > 0$, $\sup_{u \in \mathcal{I}_x} f_\oplus(x, u) < M$ for almost all x.

Assumptions (T2) and (T3) impose conditions on the joint distribution of (X, T). Condition (T2) is a smoothness condition on the optimal transport process T, while the moment requirements in (T3) are tightness conditions that allow for the necessary asymptotic Gaussianity to hold; see Lemma 1 in Petersen, Liu and Divani (2020). Assumption (T4) involves the regression relationship between X and $\mathfrak{F}$; importantly, it ensures that the conditional mean densities are sufficiently and uniformly separated from the boundary of $\mathcal{D}$ within the space of distributions with finite second moments. In the spirit of regression, we will consider the asymptotic behavior of F_G^* conditional on the observed predictors. Define the covariance kernel

$$(3.2) \qquad K(u, v) = E\big[C_T\big(S(u), S(v)\big)\big] = \sum_{j=1}^{\infty} \lambda_j \phi_j(u) \phi_j(v), \quad u, v \in \bar{\mathcal{I}}^*,$$

where $\bar{\mathcal{I}}^*$ is the closure of $\mathcal{I}^*$. The right-hand side is the Mercer decomposition of K (Hsing and Eubank (2015)), so that $\{\phi_j\}_{j=1}^{\infty}$ forms an orthonormal set in $L^2(\bar{\mathcal{I}}^*, f_\oplus^*)$ and λ_j are positive, nonincreasing in j, and satisfy $\sum_{j=1}^{\infty} \lambda_j < \infty$. Since K can be associated with a linear integral operator on $L^2(\bar{\mathcal{I}}^*, f_\oplus^*)$, we will refer to λ_j as the eigenvalues of K.

THEOREM 1. *Suppose assumptions* (T1)–(T4) *hold. Then*

$$F_G^* | X_1, \ldots, X_n \xrightarrow{D} \sum_{j=1}^{\infty} \lambda_j \omega_j \quad \textit{almost surely,}$$

where ω_j are i.i.d. χ_p^2 random variables and λ_j are the eigenvalues in (3.2).

While this limiting distribution may appear surprising at first sight given the non-Euclidean setting, they are a result of the fact that central limit theorems can still be derived for data on manifolds (e.g., Barden, Le and Owen (2013)), including the space of distributions under the Wasserstein metric (Panaretos and Zemel (2016)).

The limiting distribution obtained in Theorem 1 depends on unknown parameters, namely the eigenvalues λ_j, that must be approximated to formulate a rejection region. A natural approach would be to estimate the kernel K in (3.2) directly, followed by applying a modified Mercer decomposition incorporating the estimated marginal Wasserstein mean. Thankfully, this complicated approach is not necessary. Define the quantile conditional covariance kernel $C_Q(s, t) = E\{\text{Cov}(Q(s), Q(t)|X)\}$. A simple calculation reveals that $C_Q(s, t) = K(Q_\oplus^*(s), Q_\oplus^*(t))$, so that λ_j are also the eigenvalues of C_Q, which can be estimated by

$$(3.3) \qquad \hat{C}_Q(s, t) = \frac{1}{n} \sum_{i=1}^{n} (Q_i(s) - \hat{Q}_i(s))(Q_i(t) - \hat{Q}_i(t)),$$

where $\hat{Q}_i$ are the fitted quantile functions corresponding to densities $\hat{\mathfrak{F}}_i$. Let $\hat{\lambda}_j$ be the corresponding eigenvalues of $\hat{C}_Q$.

One can consistently approximate the conditional null distribution of F_G^* as follows. For $\alpha \in (0, 1)$ and eigenvalue estimates $\hat{\lambda}_j$, let $\hat{b}_\alpha^G$ be the $1 - \alpha$ quantile of $\sum_{j=1}^{\infty} \hat{\lambda}_j \omega_j$, where ω_j are as in Theorem 1. Computation of this critical value is outlined in Algorithm 3 of Petersen, Liu and Divani (2020). The following result on the conditional power $\beta_n^G = P_{\mathbb{X}}(F_G^* > \hat{b}_\alpha^G)$ follows from Theorem 1. Let $\mathfrak{G}$ denote the collection of distributions on $\mathbb{R}^p \times \mathcal{D}$ such that model (2.6) holds.

COROLLARY 1. *If $\mathcal{G} \in \mathfrak{G}$ satisfies $\mathcal{H}_0^G$ and* (T1)–(T4) *hold, then $\beta_n^G \to \alpha$ almost surely.*

In addition to having the correct asymptotic size for any null model, we demonstrate the power performance of the above test under a sequence of contiguous alternatives. To do so, consider a subclass $\mathfrak{G}^* \subset \mathfrak{G}$ of Wasserstein regression models, with the marginal distribution of X being fixed and satisfying $E(\|X\|_E^4) < \infty$, for which the criteria below are satisfied. Recall that $\gamma(u) = \text{Cov}(X, R(u))$, where $R = T \circ S$.

(G1) For all $\mathcal{G} \in \mathfrak{G}^*$, (T1) and (T2) hold, with $\sup_{u \in \mathbb{R}} C_T(u, u)$ and $\sup_{u \in \mathbb{R}} C_T^{(1,1)}(u, u)$ uniformly bounded in $\mathfrak{G}^*$. In addition, $\sup_{u \in \mathcal{I}^*} \|\gamma'(u)\|_E$ and $\sup_{u \in \mathcal{I}^*} \|\gamma''(u)\|_E$ are both uniformly bounded in $\mathfrak{G}^*$.

(G2) With L being the random Lipschitz constant in (T2), the following uniform moment conditions are satisfied:

$$\limsup_{M \to \infty} \sup_{\mathcal{G} \in \mathfrak{G}^*} E\Big[\|\tilde{X}\|_E^2 \sup_{u \in \mathbb{R}} |T'(u)| \mathbf{1}\Big(\|X\|_E^2 \sup_{u \in \mathbb{R}} |T'(u)| > M\Big)\Big] = 0,$$

$$\limsup_{M \to \infty} \sup_{\mathcal{G} \in \mathfrak{G}^*} E\big[\|\tilde{X}\|_E^3 L \mathbf{1}(\|\tilde{X}\|_E^3 L > M)\big] = 0.$$

(G3) There exists $M_1 > 1$ such that $\inf_{\mathcal{G} \in \mathfrak{G}^*} \inf_{u \in \mathcal{I}^*} f_{\oplus}^*(u) \geq M_1^{-1}$.

(G4) There exists $M_2 < \infty$ such that $\sup_{\mathcal{G} \in \mathfrak{G}^*} \sup_{x \in \mathfrak{S}_X} \sup_{u \in \mathcal{I}_x} f_{\oplus}(x, u) \leq M_2$.

THEOREM 2. *Let $\mathfrak{G}^*$ satisfy (G1)–(G4), and a_n be a sequence such that $a_n \to 0$ and $\sqrt{n} a_n \to \infty$. Consider a sequence of alternative global hypotheses*

$$\mathcal{H}_{A,n}^G : \mathcal{G} \in \mathfrak{G}_{A,n}, \quad \mathfrak{G}_{A,n} = \{\mathcal{G} \in \mathfrak{G}^* : E[d_W^2(f_{\oplus}(X), f_{\oplus}^*)] \geq a_n^2\}.$$

Then the worst case power converges strongly and uniformly to 1, that is, for any $\epsilon > 0$,

$$\inf_{\mathcal{G} \in \mathfrak{G}_{A,n}} P\Big(\inf_{m \geq n} \beta_m^G \geq 1 - \epsilon\Big) \to 1.$$

3.2. *Test of partial effects.* Beyond a test for no effect, it is often necessary to test the effect for just a single predictor or a subset of them. With $X = (Y, Z)$ as in Section 2.3, under the partial null hypothesis $\mathcal{H}_0^P : f_{\oplus}(x) = f_{0,\oplus}(y)$,

$$E(d_W^2(f_{\oplus}(X), f_{\oplus}^*)) = E(d_W^2(f_{0,\oplus}(Y), f_{\oplus}^*)),$$

motivating the partial Wasserstein F-statistic

(3.4) $$F_P^* = \sum_{i=1}^n [d_W^2(\hat{\mathfrak{F}}_i, \hat{f}_{\oplus}^*) - d_W^2(\hat{\mathfrak{F}}_{0,i}, \hat{f}_{\oplus}^*)], \quad \hat{\mathfrak{F}}_{0,i} = \hat{f}_{0,\oplus}(Y_i),$$

corresponding to the numerator of the partial F-statistic in the multiple linear regression setting.

Setting $J^\top = [-\Sigma_{ZY}\Sigma_{YY}^{-1} I_{p-1}]$ and $\Sigma_{Z|Y} = \Sigma_{ZZ} - \Sigma_{ZY}\Sigma_{YY}^{-1}\Sigma_{YZ}$, define the covariance matrix kernel

(3.5)
$$K^*(u, v) = \Sigma_{Z|Y}^{-1/2} J^\top E[(X - \mu)(X - \mu)^\top C_T(S(u), S(v))] J \Sigma_{Z|Y}^{-1/2}$$
$$= \sum_{l=1}^\infty \tau_l \varphi_l(u) \varphi_l^\top(v).$$

Here, the functions φ_l form an orthonormal set in $(L^2(\bar{\mathcal{I}}^*, f_{\oplus}^*))^{p-q}$ and the positive, nonincreasing eigenvalues τ_l of K^* satisfy $\sum_{l=1}^\infty \tau_l < \infty$.

THEOREM 3. *Under (T1)–(T4),*

$$F_P^* | X_1, \ldots, X_n \xrightarrow{D} \sum_{l=1}^{\infty} \tau_l \xi_l^2 \quad \text{almost surely,}$$

where ξ_l are independent standard normal random variables. If, in addition, $E(Z|Y)$ is linear in Y and $\text{Var}(Z|Y) = \Sigma_{Z|Y}$ almost surely, then

$$\{\tau_l\}_{l=1}^{\infty} = \underbrace{\lambda_1, \ldots, \lambda_1}_{p-q}, \underbrace{\lambda_2, \ldots, \lambda_2}_{p-q}, \ldots$$

so that $\sum_{l=1}^{\infty} \tau_l \xi_l^2 \stackrel{D}{=} \sum_{j=1}^{\infty} \lambda_j \omega_j'$, where ω_j' are i.i.d. χ_{p-q}^2 random variables.

Similar to the global case, the τ_l in (3.5) also correspond to the eigenvalues of the kernel $C_Q^*(s, t) = K^*(Q_{\oplus}^*(s), Q_{\oplus}^*(t))$. Setting $X_{i,c} = X_i - \bar{X}$, a natural estimator is

$$(3.6) \quad \hat{C}_Q^*(s, t) = \hat{\Sigma}_{Z|Y}^{-1/2} \hat{J}^{\top} \left\{ \frac{1}{n} \sum_{i=1}^{n} X_{i,c} X_{i,c}^{\top} (Q_i(s) - \hat{Q}_i(s))(Q_i(t) - \hat{Q}_i(t)) \right\} \hat{J} \hat{\Sigma}_{Z|Y}^{-1/2},$$

where $\hat{\Sigma}_{Z|Y}$ and $\hat{J}$ are plug-in estimates. Let $\hat{\tau}_j$ be the corresponding eigenvalue estimates from $\hat{C}_Q^*(s, t)$, and $\hat{b}_\alpha^P$ be the $1 - \alpha$ percentile of $\sum_{l=1}^{\infty} \hat{\tau}_l \xi_l^2$, with ξ_l as in Theorem 3. If the additional assumptions in the second part of Theorem 3 hold, then let $\hat{b}_\alpha^P$ be the $1 - \alpha$ quantile of $\sum_{j=1}^{\infty} \hat{\lambda}_j \omega_j'$. Computation of these critical values can be done in the same was as the global case. We have the following size and power results for the partial Wasserstein F-test, with $\beta_n^P = P_{\mathbb{X}}(F_p^* > \hat{b}_\alpha^P)$ denoting the conditional power as a function of the underlying model $\mathcal{G}$.

COROLLARY 2. *If $\mathcal{G} \in \mathfrak{G}$ satisfies $\mathcal{H}_0^P$ and (T1)–(T4) hold, then $\beta_n^P \to \alpha$ almost surely.*

THEOREM 4. *Let $\mathfrak{G}^*$ satisfy (G1)–(G4), and a_n be as in Theorem 2. Consider a sequence of alternative partial hypotheses*

$$\mathcal{H}_{A,n}^P : \mathcal{G} \in \mathfrak{G}_{A,n}', \quad \mathfrak{G}_{A,n}' = \{\mathcal{G} \in \mathfrak{G}^* : E[d_W^2(f_{\oplus}(X), f_{0,\oplus}(Y))] \geq a_n^2\}.$$

Then the worst case power converges strongly and uniformly to 1, that is, for any $\epsilon > 0$,

$$\inf_{\mathcal{G} \in \mathfrak{G}_{A,n}'} P\left(\inf_{m \geq n} \beta_m^P \geq 1 - \epsilon\right) \to 1.$$

3.3. *Alternative testing approximations.* As an alternative to estimating the eigenvalues in the limiting distributions of F_G^* and F_P^*, Satterthwaite's approximation (Satterthwaite (1941), Shen and Faraway (2004)) can also be employed. Using the global test as an example, we approximate the null distribution of the test statistic by $a \chi_m^2$, where $a, m > 0$ are scalars chosen to satisfy the moment matching conditions $am = p \sum_{j=1}^{\infty} \lambda_j$, $a^2 m = p \sum_{j=1}^{\infty} \lambda_j^2$. Using this approximation, one does not need to estimate the individual eigenvalues λ_j, given the equalities (Hsing and Eubank (2015))

$$\sum_{j=1}^{\infty} \lambda_j = \int_0^1 C_Q(t, t) \, dt, \qquad \sum_{j=1}^{\infty} \lambda_j^2 = \int_0^1 \int_0^1 C_Q^2(s, t) \, ds \, dt.$$

Hence, one can compute the approximate values a and m using the corresponding norms of the estimate $\hat{C}_Q$. This alternative approach is outlined in Algorithm 4 of Petersen, Liu and Divani (2020).

Finally, as the limiting distributions in Theorems 1 and 3 are the result of underlying central limit theorems, one may employ a bootstrap approach to testing these hypotheses. Since the inference is conditional on the observed predictors X_i, a natural approach is to perform a residual transport bootstrap. Using the global test as an example, let $\hat{T}_{i,0} = Q_i \circ \hat{F}_\oplus^*$ be the approximate versions of the residual transports T_i under $\mathcal{H}_0^G$. Obtain B independent bootstrap samples $\{\tilde{T}_i^b\}_{i=1}^n$, $b = 1, \ldots B$, by sampling with replacement from the $\hat{T}_{i,0}$, and form the bootstrapped quantile functions $\tilde{Q}_i^b = \tilde{T}_i^b \circ \hat{Q}_\oplus^*$. Then compute bootstrap estimates $\hat{f}_\oplus^b(x)$ and $\hat{f}_\oplus^{*,b}$ using the data $(X_i, \tilde{Q}_i^b)$, $i = 1, \ldots, n$. Finally, compute the bootstrap statistics $\tilde{F}_G^{*,b} = \sum_{i=1}^n d_W^2(\hat{f}_\oplus^b(X_i), \hat{f}_\oplus^{*,b})$. The bootstrap p-value then becomes

$$\frac{\#\{\tilde{F}_G^{*,b} > F_G^*\} + 1}{B + 1}.$$

This global residual bootstrap test is outlined in Algorithm 5 of Petersen, Liu and Divani (2020). For the partial test, this residual bootstrap can only be employed if the support of the densities is fixed, since otherwise the null residual transports $\hat{T}_{i,0} = Q_i \circ \hat{F}_{0,\oplus}(Y_i)$ will have different supports.

4. Confidence bands.
We now develop methodology for producing a confidence set for $f_\oplus(x)$, where x is considered to be fixed. In similar settings where one desires to make a confidence statement for a functional parameter, such as nonparametric regression (Claeskens and Van Keilegom (2003), Eubank and Speckman (1993)), mean and covariance estimation in functional data analysis (Cao, Yang and Todem (2012), Degras (2011), Wang and Yang (2009)), and the varying coefficient model (Fan and Zhang (2008)), one can either build pointwise or simultaneous bands. In the current setting of Wasserstein regression for density response data, the constraints inherent to the density targets $f_\oplus(x)$ render pointwise confidence bands of little use.

Hence, the approach we take will be motivated by simultaneous confidence bands. Here, the descriptor "simultaneous" refers to the argument u of the functional parameter $f_\oplus(x, u)$, and not to the specific regressor value x under consideration. Let g be a generic functional parameter of interest, where we assume that g is bounded. Given an estimator $\hat{g}$, a typical approach to formulating a simultaneous confidence band is to show that $b_n(\hat{g}(u) - g(u))/a(u)$ converges weakly to a limiting process (usually Gaussian) in the space of bounded functions under the uniform metric (van der Vaart and Wellner (1996)), where $a > 0$ is a scaling function and b_n^{-1} is the rate of convergence. By an application of the continuous mapping theorem, one can then obtain a confidence band for g of the form

$$\{g^* : \hat{g}(u) - ca(u)b_n^{-1} \le g^*(u) \le \hat{g}(u) + ca(u)b_n^{-1} \text{ for all } u\}.$$

This band corresponds to all functions that are almost everywhere between the lower and upper bounds, and is the closest one can get to a confidence interval in function space. This partial ordering is induced explicitly by the uniform metric, and indicates why simultaneous confidence bands are so useful for functional parameters, in that one can visualize the entire set graphically. We will explore two different approaches to formulating simultaneous confidence bands. The first arises naturally from the Wasserstein geometry, and provides a distributional band for either $Q_\oplus(x)$ or $F_\oplus(x)$, but not the density parameter. In the second approach, we strengthen the convergence results and utilize the delta method to construct a simultaneous confidence band for $f_\oplus(x)$.

4.1. *Intrinsic Wasserstein-∞ bands.* The first method is directly related to the geometry imposed by the Wasserstein-∞ metric; see (2.2). Specifically, if $\hat{V}_x = \hat{Q}_\oplus(x) \circ F_\oplus(x)$ is the optimal transport from the target to the estimate, then

$$d_\infty(f_\oplus(x), \hat{f}_\oplus(x)) = f_\oplus(x)\text{-}\operatorname*{ess\,sup}_u |\hat{V}_x(u) - u| = \sup_{u \in \mathcal{I}_x} |\hat{V}_x(u) - u|.$$

It is then natural to establish weak convergence (denoted by $\rightsquigarrow$) of the process $\hat{V}_x(u)$ within the space of bounded functions on $\mathcal{I}_x$, denoted $L^\infty(\mathcal{I}_x)$. Define $\Lambda = E(\tilde{X}\tilde{X}^\top)$ and the covariance kernel

$$(4.1) \qquad \tilde{K}(u, v) = \Lambda^{-1} E\big[\tilde{X}\tilde{X}^\top C_T\big(S(u), S(v)\big)\big]\Lambda^{-1}, \quad u, v \in \bar{\mathcal{I}}^*.$$

THEOREM 5. *Suppose that* (T1)–(T4) *hold. Then there exists a zero-mean Gaussian process* $\mathcal{M}_x$ *on* $L^\infty(\mathcal{I}_x)$ *such that*

$$\sqrt{n}(\hat{V}_x - \mathrm{id})|X_1, \ldots, X_n \rightsquigarrow \mathcal{M}_x \quad \text{almost surely.}$$

With $u_x = Q_\oplus^* \circ F_\oplus(x, u)$ *for any* $u \in \mathcal{I}_x$, *the covariance of* $\mathcal{M}_x$ *is*

$$(4.2) \qquad C_x(u, v) = \tilde{x}^\top \tilde{K}(u_x, v_x)\tilde{x}.$$

As in the testing procedures, estimation of the covariance kernel $\tilde{K}$ is simplified by moving to quantile functions. Set $D_Q(s, t) = \tilde{K}(Q_\oplus^*(s), Q_\oplus^*(t))$, with estimate

$$(4.3) \qquad \hat{D}_Q(s, t) = \hat{\Lambda}^{-1}\left\{\frac{1}{n}\sum_{i=1}^n \tilde{X}_i\tilde{X}_i^\top(Q_i(s) - \hat{Q}_i(s))(Q_i(t) - \hat{Q}_i(t))\right\}\hat{\Lambda}^{-1},$$

where $\hat{\Lambda} = n^{-1}\sum_{i=1}^n \tilde{X}_i\tilde{X}_i^\top$. This leads to

$$(4.4) \qquad \hat{C}_x(u, v) = \tilde{x}^\top \hat{D}_Q\big(\hat{F}_\oplus(x, u), \hat{F}_\oplus(x, v)\big)\tilde{x}, \quad u, v \in \big[\hat{Q}_\oplus(x, 0), \hat{Q}_\oplus(x, 1)\big].$$

Let m_α be the $1 - \alpha$ quantile of the distribution of

$$\zeta_x := \sup_{u \in \mathcal{I}_x} C_x(u, u)^{-1/2}|\mathcal{M}_x(u)|.$$

Since m_α is unknown, we estimate it as follows. Observe that $\mathcal{N}_x = \mathcal{M}_x \circ Q_\oplus(x)$ is a Gaussian process on $L^\infty[0, 1]$ with covariance $\tilde{x}^\top D_Q(s, t)\tilde{x}$. Conditional on the data, let $\hat{\mathcal{N}}_x$ be a zero-mean Gaussian process with covariance $\tilde{x}^\top \hat{D}_Q(s, t)\tilde{x}$. Define

$$\hat{\zeta}_x = \sup_{t \in [0, 1]} \big[\tilde{x}^\top \hat{D}_Q(t, t)\tilde{x}\big]^{-1/2}|\hat{\mathcal{N}}_x(t)|,$$

and set $\hat{m}_\alpha$ as the $1 - \alpha$ quantile of $\hat{\zeta}_x$. Then the Wasserstein-∞ confidence band for $f_\oplus(x)$ is

$$(4.5) \qquad C_{\alpha,n}(x) = \left\{g \in \mathcal{D} : \sup_{u \in \hat{\mathcal{I}}_x} \frac{|T_{g,x}(u) - u|}{\hat{C}_x(u, u)^{1/2}} \leq \frac{\hat{m}_\alpha}{\sqrt{n}}\right\}, \quad T_{g,x} = G^{-1} \circ \hat{F}_\oplus(x).$$

We have the following corollary of Theorem 5.

COROLLARY 3. *Suppose* (T1)–(T4) *hold. If* $\inf_{u \in \mathcal{I}^*} C_T(u, u) > 0$, *then*

$$P_{\mathbb{X}}\big(f_\oplus(x) \in C_{\alpha,n}(x)\big) \to 1 - \alpha \quad \text{almost surely.}$$

An important case arising in practice that is ruled out by the requirement of strictly positive covariance is when the support of $\mathfrak{F}$ is some fixed interval $\mathcal{I}$, so that the random transport T is necessarily fixed at the boundaries. In this case, a slight adjustment can be made as outlined in Corollary 5 of Petersen, Liu and Divani (2020).

Next, we demonstrate a connection between these simultaneous Wasserstein-∞ bands and the usual partial stochastic ordering of distributions. Recall that a cdf F is said to be stochastically greater than another G, written $F \succ G$, if $F(u) \leq G(u)$ for all u. For $F_2 \succ F_1$, define the bracket

$$[F_1, F_2] = \{g \in \mathcal{D} : F_2 \succ G \succ F_1\}$$

consists of all distributions in $\mathcal{D}$ which lie between F_1 and F_2 in the stochastic ordering. From (4.5), we deduce that the simultaneous confidence band consists of all densities $g \in \mathcal{D}$ such that

$$M_L(u) \leq G^{-1} \circ \hat{F}_\oplus(x, u) \leq M_U(u), \qquad (M_L(u), M_U(u)) = u \pm n^{-1/2} \hat{m}_\alpha \hat{C}_x(u, u)^{1/2}.$$

Define T_L to be the unique projection (in $L^2[0, 1]$) of M_L onto the closed and convex set of nondecreasing functions M for which $M(u) \geq M_L(u)$. Similarly, T_U is unique largest nondecreasing function below M_U. Note that $T_L(u) \leq T_{g,x}(u) \leq T_U(u)$ for all $g \in C_{\alpha,n}(x)$. Then the cdf bounds $F_L = \hat{F}_\oplus(x) \circ T_L^{-1}$ and $F_U = \hat{F}_\oplus(x) \circ T_U^{-1}$ represent the Wasserstein-∞ band, that is, $C_{\alpha,n}(x) = [F_L, F_U]$. Algorithm 6 in Petersen, Liu and Divani (2020) outlines the steps for computing $C_{\alpha,n}(x)$ in practice.

4.2. *Wasserstein density bands.* One drawback of the above simultaneous confidence bands is that they do not readily translate to the space of densities, since, if $F_1 \prec F_2$ in the stochastic ordering, their derivatives need not satisfy $f_1 \geq f_2$. Thus, we take a second approach to form a confidence band in density space, based on the direct difference $\hat{f}_\oplus(x) - f_\oplus(x)$ rather than the optimal transport map between these distributions.

An interesting challenge associated with this approach is that the supports of the target $f_\oplus(x)$ and its estimate $\hat{f}_\oplus(x)$ may differ, so that convergence may be ill-behaved near the boundaries. To resolve this, for any $\delta \in (0, 1/2)$, define $\mathcal{I}_x^\delta = [Q_\oplus(x, \delta), Q_\oplus(x, 1 - \delta)]$. We consider conditional weak convergence of the process $\sqrt{n}(\hat{f}_\oplus(x) - f_\oplus(x))$ in the space $L^\infty(\mathcal{I}_x^\delta)$. For $l, m \in \{0, 1\}$ and $\tilde{K}$ as in (4.1), define $\tilde{K}^{(l,m)} = \frac{\partial^{l+m}}{\partial u^l \partial v^m} \tilde{K}$. Lastly, set

$$\mathbb{K}(u, v) = \begin{pmatrix} \tilde{K}(u, v) & [f_\oplus^*(v)]^{-1} \tilde{K}^{(0,1)}(u, v) \\ [f_\oplus^*(u)]^{-1} \tilde{K}^{(1,0)}(u, v) & [f_\oplus^*(u) f_\oplus^*(v)]^{-1} \tilde{K}^{(1,1)}(u, v) \end{pmatrix}, \quad u, v \in \mathcal{I}^*.$$

THEOREM 6. *Suppose (T1)–(T4) hold, and that $f_\oplus^*$ is continuously differentiable. Then, for almost all x, there exists a zero-mean Gaussian process $\mathcal{F}_x$ on $L^\infty(\mathcal{I}_x^\delta)$ such that*

$$\sqrt{n}(\hat{f}_\oplus(x) - f_\oplus(x)) | X_1, \ldots, X_n \rightsquigarrow \mathcal{F}_x \quad \text{almost surely.}$$

With $u_x = Q_\oplus^* \circ F_\oplus(x, u)$ and $c_u^\top = (\partial/\partial u f_\oplus(x, u), -f_\oplus^2(x, u))$ for any $u \in \mathcal{I}_x^\delta$, and $\otimes$ denoting the Kronecker product, the covariance of $\mathcal{F}_x$ is

$$(4.6) \qquad \mathcal{R}_x(u, v) = c_u^\top (\tilde{x}^\top \otimes \mathbb{K}(u_x, v_x) \otimes \tilde{x}) c_v.$$

We now describe the method for estimating $\mathcal{R}_x$. With $\hat{D}_Q$ as in (4.3), let $\hat{D}_Q^{(l,m)} = \frac{\partial^{l+m}}{\partial s^l \partial t^m} \hat{D}_Q$. Then define $\mathbb{D}(s, t) = \mathbb{K}(Q_\oplus^*(s), Q_\oplus^*(t))$ and its estimate

$$(4.7) \qquad \hat{\mathbb{D}}(s, t) = \begin{pmatrix} \hat{D}_Q(s, t) & \hat{D}_Q^{(0,1)}(s, t) \\ \hat{D}_Q^{(1,0)}(s, t) & \hat{D}_Q^{(1,1)}(s, t), \end{pmatrix} \quad s, t \in [\delta, 1 - \delta],$$

and set

$$\hat{\mathbb{K}}(u, v) = \hat{\mathbb{D}}(\hat{F}_{\oplus}(x, u), \hat{F}_{\oplus}(x, v)), \quad u, v \in \hat{\mathcal{I}}_x^{\delta} = [\hat{Q}_{\oplus}(x, \delta), \hat{Q}_{\oplus}(x, 1 - \delta)].$$

Next, the chain rule implies that $\frac{\partial}{\partial u} f_{\oplus}(x, u) = -f_{\oplus}^3(x, u) \frac{\partial}{\partial t} q_{\oplus}(x, F_{\oplus}(x, u))$. Hence, define $\frac{\partial}{\partial t} \hat{q}_{\oplus}(x, t) = n^{-1} \sum_{i=1}^{n} s_{in}(x) q_i'(t)$ for $t \in [\delta, 1 - \delta]$, yielding

$$\frac{\partial}{\partial u} \hat{f}_{\oplus}(x, u) = -\hat{f}_{\oplus}^3(x, u) \frac{\partial}{\partial t} \hat{q}_{\oplus}(x, \hat{F}_{\oplus}(x, u)), \quad u \in \hat{\mathcal{I}}_x^{\delta}.$$

Finally, for $u, v \in \hat{\mathcal{I}}_x^{\delta}$, set

$$(4.8) \qquad \hat{\mathcal{R}}_x(u, v) = \hat{c}_u^{\top}(\tilde{x}^{\top} \otimes \hat{\mathbb{K}}(u, v) \otimes \tilde{x})\hat{c}_v, \quad \hat{c}_u^{\top} = \left(\frac{\partial}{\partial u} \hat{f}_{\oplus}(x, u), -\hat{f}_{\oplus}^2(x, u) \right).$$

Let l_{α} be the unknown $1 - \alpha$ quantile of

$$\xi_x = \sup_{u \in \mathcal{I}_x^{\delta}} \mathcal{R}_x(u, u)^{-1/2} |\mathcal{F}_x(u)|.$$

Similar to the case of the Wasserstein-∞ band, conditional on the data, let $\hat{\mathcal{J}}_x(t)$ be a zero-mean Gaussian process on $(0, 1)$, with covariance $\hat{\mathcal{R}}_x(\hat{Q}_{\oplus}(x, s), \hat{Q}_{\oplus}(x, t))$. Define $\hat{l}_{\alpha}$ as the the $1 - \alpha$ quantile of

$$\hat{\xi}_x = \sup_{t \in [\delta, 1-\delta]} [\hat{\mathcal{R}}_x(\hat{Q}_{\oplus}(x, t), \hat{Q}_{\oplus}(x, t))]^{-1/2} |\hat{\mathcal{J}}_x(t)|.$$

Then the nearly simultaneous Wasserstein density confidence band is

$$(4.9) \qquad B_{\alpha,n}(x) = \left\{ g \in \mathcal{D} : \sup_{u \in \hat{\mathcal{I}}_x^{\delta}} \frac{|g(u) - \hat{f}_{\oplus}(x, u)|}{\hat{\mathcal{R}}_x(u, u)^{1/2}} \leq \frac{\hat{l}_{\alpha}}{\sqrt{n}} \right\}.$$

See Algorithm 7 in Petersen, Liu and Divani (2020) for an implementation of this confidence band.

The final corollary demonstrates almost sure convergence of the coverage rate of the Wasserstein density confidence band. Since the estimation of the covariance $\mathcal{R}_x$ is considerably more complex than for the Wasserstein-∞ band, additional smoothness assumptions are required on the random transport map T.

COROLLARY 4. *Suppose the assumptions of Theorem 6 hold, and that* $\inf_{u \in \mathcal{I}_x^{\delta}} \mathcal{R}_x(u, u) > 0$. *Furthermore, assume that T is twice differentiable almost surely, with T'' having Lipschitz constant $\tilde{L}$ and satisfying*

$$E \left(\|\tilde{X}\|_E^6 \sup_{u \in \mathbb{R}} |T''(u)|^2 \right) < \infty, \quad E(\tilde{L}) < \infty.$$

Then

$$P_{\mathbb{X}}(f_{\oplus}(x) \in B_{\alpha,n}) \to 1 - \alpha \quad \text{almost surely}.$$

5. Simulations. Extensive simulations were conducted to investigate the empirical performance of the F-tests and confidence bands developed in Sections 3 and 4. Data were generated according to model (2.6) for increasing sample sizes and different random optimal transport processes T. Furthermore, to assess the robustness of our procedures when densities are only indirectly observed through samples, simulations were also conducted using only raw data generated from the random densities. For brevity, results for the indirectly observed case are reported in Petersen, Liu and Divani (2020).

We first describe the simulation settings for the Wasserstein F-tests. Following Example 1, set f_0 as the standard normal distribution, truncated to the interval $[-2.5, 2.5]$ and renormalized to have integral one. Bivariate predictors $X_i = (X_{i1}, X_{i2})$ were generated as independent uniform random variables on $[-0.5, 0.5]$. Let $v(x) = \alpha_1 x_1 + \alpha_2 x_2$ and $\tau(x) = 2 + \beta_1 x_1 + \beta_2 x_2$, where the parameters α_j, β_j were varied to increase signal strength for assessing power performance. The conditional Wasserstein mean density was

$$(5.1) \qquad f_\oplus(x, u) = \frac{1}{\tau(x)} f_0 \left(\frac{u - v(x)}{\tau(x)} \right).$$

To produce the random densities $\mathfrak{F}_i$, random optimal transport maps T_i were generated in two different ways. In the first setting, $T_i(u) = V_{i1} + V_{i2}u$, where $V_{i1} \sim \mathcal{U}(-0.5, 0.5)$ and $V_{i2} \sim \mathcal{U}(0.5, 1.5)$ were generated independently, yielding linear optimal transport maps. In the second setting, following the method described in Section 8.1 of Panaretos and Zemel (2016), nonlinear optimal transport maps were generated. Specifically, define template transport maps $M_0(u) = u$ and

$$M_k(u) = u - \frac{\sin(ku)}{|k|}, \quad k \neq 0.$$

For $J = 10$, $(W_{i1}, \ldots, W_{iJ})$ were sampled from a order J Dirichlet distribution, so that $W_j > 0$ and $\sum_{j=1}^{J} W_j = 1$. Then $K_{ij}, j = 1, \ldots, J$, were sampled independently and uniformly from $(-0.250, -0.125, 0, 0.125, 0.250)$. The resulting optimal transport map in the second setting was $T_i(u) = \sum_{j=1}^{J} W_{ij} M_{K_{ij}}(u)$. Finally, the random density was generated as

$$\mathfrak{F}_i(u) = f_\oplus(X_i, T_i^{-1}(u))(T_i^{-1})'(u).$$

In the case when the densities were not directly observed, secondary samples of size 300 from each $\mathfrak{F}_i$ were generated, and local linear smoothing of the empirical quantile function was used to estimate the quantile functions $\hat{Q}_i$ for use in the testing algorithms. The Matlab function "smooth" was used with default choice of the smoothing parameter.

In both the global and partial F tests simulations, the empirical size approached the nominal size $\alpha = 0.05$ as the sample size grew, and the empirical power function increased with the increasing signal strength. To create models satisfying global null and alternative hypotheses, $\alpha_1 = \alpha_2 = \beta_1 = \beta_2$ were set to be equal, with the common value running through $(0, 0.5)$. In the partial F-test, α_1 and β_1 were fixed as $\alpha_1 = 2$ and $\beta_1 = 1$, respectively, while $\alpha_2 = \beta_2$ varied across $(0, 0.5)$. Therefore, the null hypothesis in the partial Wasserstein F-test was $\mathcal{H}_0^P : f_\oplus(x) = f_\oplus^0(x_1)$.

For each of three sample sizes $n = 100, 200, 500$, five hundred simulations were used to compute empirical power curves shown in Figure 2. For the global tests, performance was similar among the χ^2 mixture and alternative Satterthwaite and bootstrap tests outlined in Section 3.3. The nominal sizes converged to the right level, and power converged to one with increasing values of the common parameter value. Power was generally higher for the linear transport map setting, as expected. The same pattern is observed for the partial test, although the Satterthwaite alternative more accurately maintained the nominal level for low sample sizes. The bootstrap test could not be implemented for the partial test due to the fact that the support of the density varies with x. Similar results for indirectly observed densities are given in Figure 5 in Petersen, Liu and Divani (2020). For both global and partial tests, the difference in performance is more prominent in low sample sizes (in n, the number of densities), with the power converging at a slower rate and slightly larger deviations from the nominal level compared to the case of observed densities.

For simplicity, in the simulations for confidence bands, a single predictor $X_i \sim \mathcal{U}(-0.5, 0.5)$ was used. With $v(x) = 2x$ and $\tau(x) = 2 + x$, the mean was again given by (5.1),
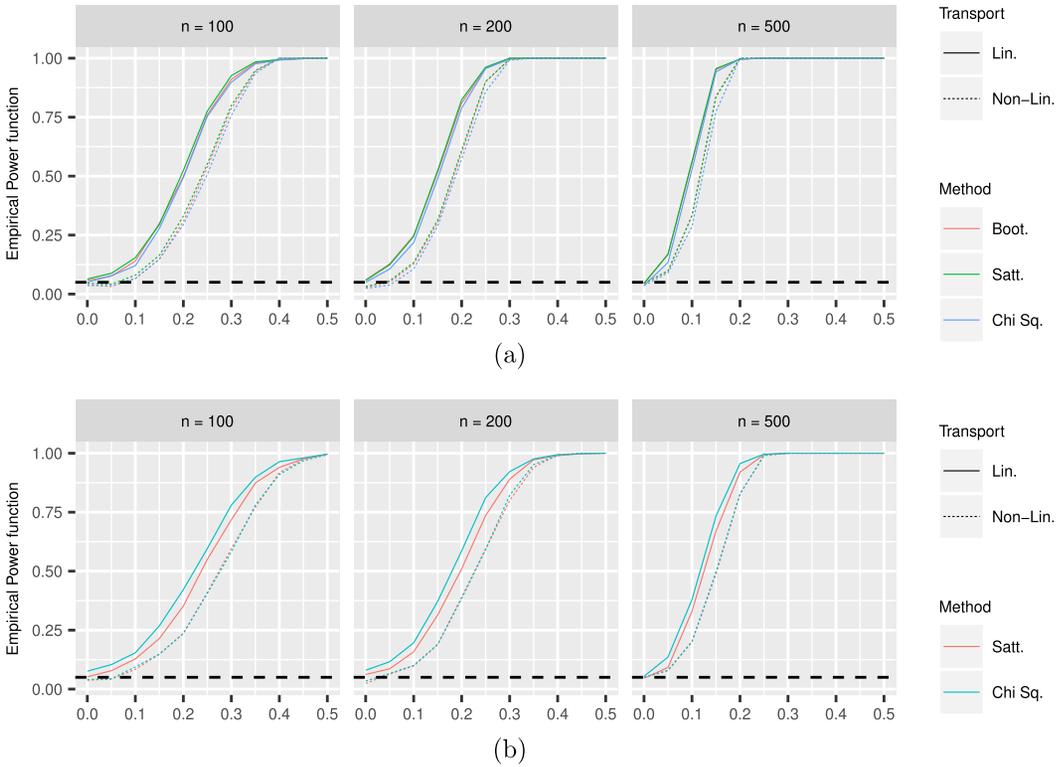
FIG. 2. *Power curves for global (top) and partial (bottom) Wasserstein F-tests, for sample sizes* $n = 100, 200, 500$. *The dotted horizontal line represents the nominal level* $\alpha = 0.05$.

where f_0 was the same as in the testing simulations. The random optimal transports and case of indirectly observed densities were handled the same as in the testing simulations as well. Both types of confidence intervals were computed for predictor values $x \in (-0.30, -0.24, \ldots, 0.24, 0.30)$, with $\delta = 0.1$ being used for the density bands. In addition, for the case of indirectly observed densities, coverage of Wasserstein-∞ bands was also measured using $\delta = 0.1$ due to boundary effects associated with the preliminary smoothing, as the random quantile functions are very steep near the boundary. The error rates in Table 1 were computed using 500 runs for each setting. Overall, error rates improved as sample size increased, with error rates near $\alpha = 0.05$ at all x values when $n = 500$.

Table 3 in Petersen, Liu and Divani (2020) contains the corresponding results when for indirectly observed densities. Note that Algorithm 7 for computing the Wasserstein density band also requires estimation of q_i and q_i' in addition to the quantile function Q_i. In our experiments, $\hat{q}_i$ and $\hat{q}_i'$ were computed by numerical differentiation of $\hat{Q}_i$. Clearly, there are alternative ways in which one might estimate these functions, but numerical differentiation was chosen because it represents a worst-case scenario in order to reveal sensitivities of the density bands to errors induced by this preprocessing step. Convergence to the nominal coverage rate was slower for indirectly observed densities. In particular, the Wasserstein-∞ band suffered from the aforementioned boundary issues associated with local linear estimation of quantile functions that are steep near the boundary. However, with $n = 500$, all but one of the error rates were below 0.1, compared to the nominal rate $\alpha = 0.05$.

6. Application to stroke data. Intracerebral hemorrhage (ICH), caused by small blood vessel ruptures inside the brain, is the second most common stroke subtype (Morgenstern et al. (2017)). Computed tomography (CT) is the most utilized imaging modality to diagnose

TABLE 1
Confidence band error rates for fully observed densities over 500 *simulation runs*

	Linear transport map			Nonlinear transport map		
	$n = 100$	$n = 200$	$n = 500$	$n = 100$	$n = 200$	$n = 500$
	Wasserstein-∞ band					
$x = -0.30$	0.050	0.050	0.046	0.044	0.074	0.048
$x = -0.24$	0.056	0.040	0.058	0.044	0.068	0.044
$x = -0.18$	0.058	0.042	0.058	0.052	0.064	0.046
$x = -0.12$	0.056	0.036	0.062	0.058	0.064	0.054
$x = -0.06$	0.056	0.040	0.048	0.058	0.070	0.058
$x = 0$	0.056	0.042	0.042	0.052	0.074	0.056
$x = 0.06$	0.062	0.052	0.036	0.052	0.072	0.056
$x = 0.12$	0.060	0.050	0.036	0.054	0.072	0.058
$x = 0.18$	0.062	0.052	0.040	0.054	0.066	0.052
$x = 0.24$	0.068	0.052	0.042	0.058	0.064	0.058
$x = 0.30$	0.076	0.044	0.050	0.058	0.062	0.056
	Wasserstein density band*					
$x = -0.30$	0.062	0.056	0.054	0.066	0.074	0.064
$x = -0.24$	0.056	0.056	0.052	0.062	0.072	0.058
$x = -0.18$	0.052	0.052	0.054	0.054	0.058	0.050
$x = -0.12$	0.034	0.042	0.052	0.048	0.052	0.046
$x = -0.06$	0.038	0.038	0.038	0.046	0.050	0.044
$x = 0$	0.030	0.026	0.040	0.058	0.050	0.052
$x = 0.06$	0.028	0.030	0.036	0.066	0.054	0.058
$x = 0.12$	0.048	0.046	0.032	0.068	0.046	0.062
$x = 0.18$	0.052	0.048	0.034	0.074	0.060	0.062
$x = 0.24$	0.062	0.046	0.036	0.074	0.066	0.062
$x = 0.30$	0.064	0.048	0.036	0.086	0.076	0.064

**$\delta = 0.1$ used to avoid boundary issue.

and study ICH in clinical settings. Various studies of ICH have revealed the importance of the density of the hematoma, in addition to important factors such as the size and location of the hematoma inside brain parenchyma. (Barras et al. (2009), Delcourt et al. (2016), Salazar et al. (2020)). The density of the hematoma is not homogenous, and common practice is to summarize this important feature by some set of subjective scalar measures that are often obtained by visual inspection and are thus user-dependent. For example, Boulouis et al. (2016) demonstrated the importance of the CT hypodensity, a binary variable indicating the presence of low-density regions within the hematoma. Instead of using any particular summary, one can study the distribution of hematoma density (i.e., a probability density function of hematoma density) throughout the entire hematoma as a functional object. Specifically, the hematoma density is measured on the Hounsfield scale (0–100 HU) for each voxel within the hematoma, and these can be used to produce a probability density function in a variety of ways. In this application, the hematoma densities were first combined into a histogram, followed by smoothing to obtain a probability density function; see Figure 1. Other methods, such as local linear smoothing to obtain quantile function estimates as illustrated in the simulations, could also be used.

To illustrate the utility of the proposed inferential methods for Wasserstein regression, we considered a study of $n = 393$ ICH anonymized subjects and analyzed the associations between 5 clinical and 4 radiological variables as predictors of the head CT hematoma densities as distributional responses. The clinical predictors were age, weight, history of diabetes and two variables indicating history of coagulopathy (Warfarin and AntiPt). Radiological predic-

TABLE 2
LQD transformation method and Wasserstein regression p-values

Category	Predictor	Wasserstein	LQD
Clinical	Age	0.862	0.761
	Weight	<0.001	0.479
	DM	0.034	0.742
	Warfarin	0.298	0.640
	AntiPt	0.078	0.900
Radiological	log(Volume)	<0.001	<0.001
	Hematoma Shape	<0.001	<0.001
	Midline Shift	<0.001	0.039
	TimetoCT	0.902	0.657

tors included the logarithm of hematoma volume, a continuous index of hematoma shape (Shape), presence of a shift in the midline of the brain, and length of the interval between stroke event and the CT scan (TimetoCT). These predictors were selected based on the natural history and published clinical studies (Al-Mufti et al. (2018), Salman, Labovitz and Stapf (2009)). A complete description of the data source can be found in Hevesi et al. (2018). As seen in Figure 1, important features that vary from subject to subject are the location and mode of the hematoma density distribution, as well as the spread, where some hematoma densities are homogeneous and concentrated around the single mode, while others are more heterogeneous or even bimodal. Thus, while the mean is an important aspect of the hematoma density, this natural summary fails to capture other forms of variability that are automatically taken into account by the Wasserstein regression model.

To assess the goodness of fit, the Wasserstein coefficient of determination (Petersen and Müller (2019)) was computed as

$$\hat{R}_\oplus^2 = 1 - \frac{\sum_{i=1}^n d_W^2(\mathfrak{F}_i, \hat{\hat{\mathfrak{F}}}_i)}{\sum_{i=1}^n d_W^2(\mathfrak{F}_i, \hat{f}_\oplus^*)} = 0.2242,$$

representing the fraction of Wasserstein variability explained by the model. As a comparison, a standard linear functional response model (Faraway (1997)) was also fit, after transforming the densities into a linear function space using the log quantile density (LQD) transformation of Petersen and Müller (2016). Because the hematoma densities fail to have common supports, as required by the LQD transformation, the densities were regularized to be strictly positive on (0, 100). This was done by appropriately mixing each density with the uniform distribution on (0, 100), with the mixture coefficient chosen so that the densities were all bounded below by 0.0005. This alternative LQD model yielded fitted densities by means of the inverse LQD transformation. The LQD method suffered from quite poor recovery of the quantile functions near the boundaries $t \in \{0, 1\}$, due to the fact that the observed densities decay to zero at the boundary of their supports. The Wasserstein distance in (2.1) was sensitive to these errors, and resulted in a much lower Wasserstein coefficient of determination of $\hat{R}_\oplus^2 = 0.0389$, so that the Wasserstein regression model provided a substantial improvement on the LQD approach.

The p-value for the global Wasserstein F-test was approximately zero using the mixture χ^2 test, giving strong evidence of a significant regression relationship between the hematoma densities and candidate predictors. The alternative testing procedures of Section 3.3 gave similar results. Significance of the effects of individual predictors are found in Table 2, where these were computed via partial Wasserstein F-tests. As a baseline for comparison, the F-tests of Shen and Faraway (2004) were performed on the LQD functional linear model. The
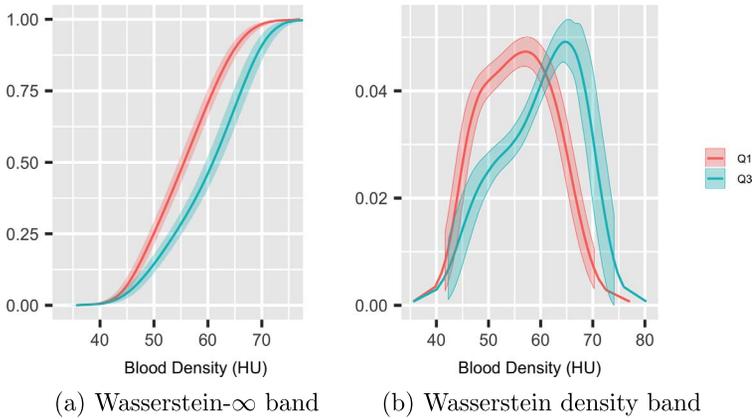
FIG. 3. *95% confidence bands for conditional Wasserstein mean CDF (left) and density (right) when hematoma volume is equal to the first (Q1) and third (Q3) quartile in the sample, with other variables set to their mean value for continuous variables and mode for binary variables.*

global F-test on the LQD model was also approximately zero, with partial p-values given in the last column of Table 2. Both models identified hematoma size and shape, as well as presenece of midline shift, as important factors in determining the distribution of density within the hematoma. However, the Wasserstein regression model identifies two clinical variables related to weight and diabetes as also affecting the hematoma density distribution.

Finally, we demonstrate the use of the two types of Wasserstein confidence bands developed in Section 4. Figure 3(a) shows two fitted Wasserstein mean cdfs. The first corresponds to a hematoma volume equivalent to the first quartile (Q1) of the observed values, with all other predictors set at their mean (for continuous variables) or mode (for binary variables). The second is similar, but for the third quartile (Q3) of hematoma volume. Each fitted cdf is also accompanied by a Wasserstein-∞ band, represented by the stochastic ordering bracket $[F_L, F_U]$. These bands may be interpreted as bounds on the "horizontal" sampling variability of the fitted distributions, that is, the sampling variability of the fitted quantile function. Figure 3(b) shows the corresponding fitted Wasserstein mean densities, along with density bands. These density bands reflect sampling variability at the density level and aid in inferring the relevance of local features seen in the fitted densities. As indicated by Figure 1, the magnitude of the horizontal variability is the smaller of the two, and this is reflected in the confidence bands. From a neurological point of view, the physiology dictates that larger hematomas tend to be more dense and homogeneous, since the pressure put on them from the surrounding tissue restricts growth and causes voids within the hematoma to be filled, confirming the observed differences between fitted cdfs/densities in Figure 3.

7. Discussion. We have studied the regression of density response curves on vector predictors under the Fréchet regression model and the Wasserstein geometry of optimal transport. The targets are the conditional Wasserstein mean densities, which can be estimated without the need of a tuning parameter, akin to a parametric model. By replacing the additive error term in ordinary linear regression with a random optimal transport map, intuitive test statistics are proposed for testing null hypotheses of both no and partial predictor effects. In the spirit of regression, asymptotic distributions are derived conditional on the observed predictors. The covariance of the random transport map is a nuisance parameter that can be consistently estimated and thus used to form a rejection region with correct asymptotic size and which are uniformly consistent against classes of contiguous alternatives.

Confidence bands are also derived for the fitted Wasserstein mean densities in two forms. Due to the intimate connection between the Wasserstein metric and quantile functions for

univariate distributions, the first type of confidence band is formed in terms of the fitted Wasserstein mean quantile function (equivalently, the Wasserstein mean cdf), and forms a bracket in the usual stochastic ordering of distributions on the real line. These intrinsic confidence bands are complemented by extrinsic bands in density space, allowing the user to simultaneously quantify sampling variability of all quantiles of the distribution, as well as the uncertainty of local features seen in the conditional Wasserstein mean densities.

As for any regression model, it will be necessary in future work to develop diagnostic tools to assess the validity of the Wasserstein regression model for a given data set. It is likely that tools for functional regression models can similarly be adapted to the setting of density response curves (Chiou and Müller (2007)). Likewise, model selection or regularized estimation procedures are clearly desirable, especially in cases where the number of predictors is large (Barber, Reimherr and Schill (2017)).

While our theoretical developments have assumed the density response curves are known, in the majority of practical situations these will need to be estimated from a collection of univariate samples, each generated by one of the random densities. In the simulations, we have demonstrated how this can be done using local linear smoothing of empirical quantile functions, while smoothed histograms could also be used as in the application to head CT densities. Some previous theoretical work in the analysis of density samples has accounted for this preprocessing step, similar to the effects of presmoothing in functional data analysis when curves are measured sparsely in time and often contaminated with noise (Kneip and Utikal (2001), Panaretos and Zemel (2016), Petersen and Müller (2016)). An important point of future research in this area will include identifying a division of regimes between dense and sparse samples for density functions, similar to Zhang and Wang (2016) for classical functional data, and their implications on inferential procedures such as those proposed in this paper.

SUPPLEMENTARY MATERIAL

Supplementary Material to "Wasserstein F-tests and confidence bands for the Fréchet regression of density response curves" (DOI: 10.1214/20-AOS1971SUPP; .pdf). The supplement includes computational algorithms, additional results on confidence bands, simulations for indirectly observed densities and proofs of all theoretical results.

REFERENCES

AGUEH, M. and CARLIER, G. (2011). Barycenters in the Wasserstein space. *SIAM J. Math. Anal.* **43** 904–924. MR2801182 https://doi.org/10.1137/100805741

AITCHISON, J. (1986). *The Statistical Analysis of Compositional Data. Monographs on Statistics and Applied Probability.* CRC Press, London. MR0865647 https://doi.org/10.1007/978-94-009-4109-0

AL-MUFTI, F., THABET, A. M., SINGH, T., EL-GHANEM, M., AMULURU, K. and GANDHI, C. D. (2018). Clinical and radiographic predictors of intracerebral hemorrhage outcome. *Interv. Neurol.* **7** 118–136. https://doi.org/10.1159/000484571

AMBROSIO, L., GIGLI, N. and SAVARÉ, G. (2008). *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, 2nd ed. *Lectures in Mathematics ETH Zürich*. Birkhäuser, Basel. MR2401600

BARBER, R. F., REIMHERR, M. and SCHILL, T. (2017). The function-on-scalar LASSO with applications to longitudinal GWAS. *Electron. J. Stat.* **11** 1351–1389. MR3635916 https://doi.org/10.1214/17-EJS1260

BARDEN, D., LE, H. and OWEN, M. (2013). Central limit theorems for Fréchet means in the space of phylogenetic trees. *Electron. J. Probab.* **18** 1–25. MR3035753 https://doi.org/10.1214/EJP.v18-2201

BARRAS, C. D., TRESS, B. M., CHRISTENSEN, S., MACGREGOR, L., COLLINS, M., DESMOND, P. M., SKOLNICK, B. E., MAYER, S. A., BRODERICK, J. P. et al. (2009). Density and shape as CT predictors of intracerebral hemorrhage growth. *Stroke* **40** 1325–1331.

BIGOT, J., GOUET, R., KLEIN, T. and LÓPEZ, A. (2017). Geodesic PCA in the Wasserstein space by convex PCA. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 1–26. MR3606732 https://doi.org/10.1214/15-AIHP706

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.

BOULOUIS, G., MOROTTI, A., BROUWERS, H. B., CHARIDIMOU, A., JESSEL, M. J., AURIEL, E., PONTES-NETO, O., AYRES, A., VASHKEVICH, A. et al. (2016). Noncontrast computed tomography hypodensities predict poor outcome in intracerebral hemorrhage patients. *Stroke* **47** 2511–2516.

BROADHURST, R. E., STOUGH, J., PIZER, S. M. and CHANEY, E. L. (2006). A statistical appearance model based on intensity quantile histograms. In 3*rd IEEE International Symposium on Biomedical Imaging*: *Nano to Macro*, 2006 422–425. IEEE Press, New York.

CAO, G., YANG, L. and TODEM, D. (2012). Simultaneous inference for the mean function based on dense functional data. *J. Nonparametr. Stat.* **24** 359–377. MR2921141 https://doi.org/10.1080/10485252.2011.638071

CHIOU, J.-M. and MÜLLER, H.-G. (2007). Diagnostics for functional regression via residual processes. *Comput. Statist. Data Anal.* **51** 4849–4863. MR2364544 https://doi.org/10.1016/j.csda.2006.07.042

CLAESKENS, G. and VAN KEILEGOM, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.* **31** 1852–1884. MR2036392 https://doi.org/10.1214/aos/1074290329

CORNEA, E., ZHU, H., KIM, P. and IBRAHIM, J. G. (2017). Regression models on Riemannian symmetric spaces. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **79** 463–482. MR3611755 https://doi.org/10.1111/rssb.12169

DEGRAS, D. A. (2011). Simultaneous confidence bands for nonparametric regression with functional data. *Statist. Sinica* **21** 1735–1765. MR2895997 https://doi.org/10.5705/ss.2009.207

DELCOURT, C., ZHANG, S., ARIMA, H., SATO, S., SALMAN, R. A.-S., WANG, X., DAVIES, L., STAPF, C., ROBINSON, T. et al. (2016). Significance of hematoma shape and density in intracerebral hemorrhage: The intensive blood pressure reduction in acute intracerebral hemorrhage trial study. *Stroke* **47** 1227–1232.

DUBEY, P. and MÜLLER, H.-G. (2019). Fréchet analysis of variance for random objects. *Biometrika* **106** 803–821. MR4031200 https://doi.org/10.1093/biomet/asz052

EGOZCUE, J. J., DÍAZ-BARRERO, J. L. and PAWLOWSKY-GLAHN, V. (2006). Hilbert space of probability density functions based on Aitchison geometry. *Acta Math. Sin.* (*Engl. Ser.*) **22** 1175–1182. MR2245249 https://doi.org/10.1007/s10114-005-0678-2

EUBANK, R. L. and SPECKMAN, P. L. (1993). Confidence bands in nonparametric regression. *J. Amer. Statist. Assoc.* **88** 1287–1301. MR1245362

FAN, J. and ZHANG, W. (2008). Statistical methods with varying coefficient models. *Stat. Interface* **1** 179–195. MR2425354 https://doi.org/10.4310/SII.2008.v1.n1.a15

FARAWAY, J. J. (1997). Regression analysis for a functional response. *Technometrics* **39** 254–261. MR1462586 https://doi.org/10.2307/1271130

FLETCHER, P. T. (2013). Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vis.* **105** 171–185. MR3104017 https://doi.org/10.1007/s11263-012-0591-y

FLETCHER, P. T., LU, C., PIZER, S. M. and JOSHI, S. (2004). Principal geodesic analysis for the study of nonlinear statistics of shape. *IEEE Trans. Med. Imag.* **23** 995–1005.

FRÉCHET, M. (1948). Les éléments aléatoires de nature quelconque dans un espace distancié. *Ann. Inst. Henri Poincaré* **10** 215–310. MR0027464

HEVESI, M., BERSHAD, E. M., JAFARI, M., MAYER, S. A., SELIM, M., SUAREZ, J. I. and DIVANI, A. A. (2018). Untreated hypertension as predictor of in-hospital mortality in intracerebral hemorrhage: A multicenter study. *J. Crit. Care* **43** 235–239.

HRON, K., MENAFOGLIO, A., TEMPL, M., HRUZOVA, K. and FILZMOSER, P. (2016). Simplicial principal component analysis for density functions in Bayes spaces. *MOX-report* **25** 2014.

HSING, T. and EUBANK, R. (2015). *Theoretical Foundations of Functional Data Analysis*, *with an Introduction to Linear Operators*. *Wiley Series in Probability and Statistics*. Wiley, Chichester. MR3379106 https://doi.org/10.1002/9781118762547

KNEIP, A. and UTIKAL, K. J. (2001). Inference for density families using functional principal component analysis. *J. Amer. Statist. Assoc.* **96** 519–542. MR1946423 https://doi.org/10.1198/016214501753168235

LEVINA, E. and BICKEL, P. (2001). The Earth mover's distance is the Mallows distance: Some insights from statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV* 2001 **2** 251–256. IEEE Press, New York.

MARRON, J. S. and ALONSO, A. M. (2014). Overview of object oriented data analysis. *Biom. J.* **56** 732–753. MR3258083 https://doi.org/10.1002/bimj.201300072

MORGENSTERN, L. B., HEMPHILL, J. C., III, ANDERSON, C., BECKER, K., BRODERICK, J. P., CONNOLLY, E. S., JR., GREENBERG, S. M., HUANG, J. N., MACDONALD, R. L., MESSÉ, S. R. et al. (2010). Guidelines for the management of spontaneous intracerebral hemorrhage: A guideline for healthcare professionals from the American Heart Association/American Stroke Association *Stroke* **41** 2108–2129.

NERINI, D. and GHATTAS, B. (2007). Classifying densities using functional regression trees: Applications in oceanology. *Comput. Statist. Data Anal.* **51** 4984–4993. MR2364554 https://doi.org/10.1016/j.csda.2006.09.028

NIETHAMMER, M., HUANG, Y. and VIALARD, F.-X. (2011). Geodesic regression for image time-series. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI* 2011 655–662. Springer, Berlin.

PANARETOS, V. M., PHAM, T. and YAO, Z. (2014). Principal flows. *J. Amer. Statist. Assoc.* **109** 424–436. MR3180574 https://doi.org/10.1080/01621459.2013.849199

PANARETOS, V. M. and ZEMEL, Y. (2016). Amplitude and phase variation of point processes. *Ann. Statist.* **44** 771–812. MR3476617 https://doi.org/10.1214/15-AOS1387

PANARETOS, V. M. and ZEMEL, Y. (2019). Statistical aspects of Wasserstein distances. *Annu. Rev. Stat. Appl.* **6** 405–431. MR3939527 https://doi.org/10.1146/annurev-statistics-030718-104938

PARZEN, E. (1979). Nonparametric statistical data modeling. *J. Amer. Statist. Assoc.* **74** 105–121. MR0529528

PASS, B. (2013). Optimal transportation with infinitely many marginals. *J. Funct. Anal.* **264** 947–963. MR3004954 https://doi.org/10.1016/j.jfa.2012.12.002

PATRANGENARU, V. and ELLINGSON, L. (2016). *Nonparametric Statistics on Manifolds and Their Applications to Object Data Analysis*. CRC Press, Boca Raton, FL. MR3444169

PETERSEN, A., LIU, X. and DIVANI, A. A (2021). Supplement to "Wasserstein F-tests and confidence bands for the Fréchet regression of density response curves." https://doi.org/10.1214/20-AOS1971SUPP.

PETERSEN, A. and MÜLLER, H.-G. (2016). Functional data analysis for density functions by transformation to a Hilbert space. *Ann. Statist.* **44** 183–218. MR3449766 https://doi.org/10.1214/15-AOS1363

PETERSEN, A. and MÜLLER, H.-G. (2019). Fréchet regression for random objects with Euclidean predictors. *Ann. Statist.* **47** 691–719. MR3909947 https://doi.org/10.1214/17-AOS1624

SALAZAR, P., DI NAPOLI, M., JAFARI, M., JAFARLI, A., ZIAI, W., PETERSEN, A., MAYER, S. A., BERSHAD, E. M., DAMANI, R. and DIVANI, A. A. (2020). Exploration of multiparameter hematoma 3D image analysis for predicting outcome after intracerebral hemorrhage. *Neurocritical Care* **32** 539–549. https://doi.org/10.1007/s12028-019-00783-8

SALMAN, R. A.-S., LABOVITZ, D. L. and STAPF, C. (2009). Spontaneous intracerebral haemorrhage. *BMJ* **339** b2586.

SATTERTHWAITE, F. E. (1941). Synthesis of variance. *Psychometrika* **6** 309–316. MR0005564 https://doi.org/10.1007/BF02288586

SHEN, Q. and FARAWAY, J. (2004). An F test for linear models with functional responses. *Statist. Sinica* **14** 1239–1257. MR2126351

SRIVASTAVA, A., JERMYN, I. and JOSHI, S. (2007). Riemannian analysis of probability density functions with applications in vision. *Proceedings from IEEE Conference on Computer Vision and Pattern Recognition* **25** 1–8.

TALSKÁ, R., MENAFOGLIO, A., MACHALOVÁ, J., HRON, K. and FIŠEROVÁ, E. (2018). Compositional regression with functional response. *Comput. Statist. Data Anal.* **123** 66–85. MR3777086 https://doi.org/10.1016/j.csda.2018.01.018

VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*. *Springer Series in Statistics*. Springer, New York. MR1385671 https://doi.org/10.1007/978-1-4757-2545-2

VILLANI, C. (2003). *Topics in Optimal Transportation*. *Graduate Studies in Mathematics* **58**. Amer. Math. Soc., Providence, RI. MR1964483 https://doi.org/10.1090/gsm/058

WANG, J. and YANG, L. (2009). Polynomial spline confidence bands for regression curves. *Statist. Sinica* **19** 325–342. MR2487893

ZHANG, Z. and MÜLLER, H.-G. (2011). Functional density synchronization. *Comput. Statist. Data Anal.* **55** 2234–2249. MR2786984 https://doi.org/10.1016/j.csda.2011.01.007

ZHANG, X. and WANG, J.-L. (2016). From sparse to dense functional data and beyond. *Ann. Statist.* **44** 2281–2321. MR3546451 https://doi.org/10.1214/16-AOS1446