Cox Point Process Regression

Álvaro Gajardo and Hans-Georg Müller

***Abstract*— Point processes in time have a wide range of applications that include the claims arrival process in insurance or the analysis of queues in operations research. Due to advances in technology, such samples of point processes are increasingly encountered. A key object of interest is the local intensity function. It has a straightforward interpretation that allows to understand and explore point process data. We consider functional approaches for point processes, where one has a sample of repeated realizations of the point process. This situation is inherently connected with Cox processes, where the intensity functions of the replications are modeled as random functions. Here we study a situation where one records covariates for each replication of the process, such as the daily temperature for bike rentals. For modeling point processes as responses with vector covariates as predictors we propose a novel regression approach for the intensity function that is intrinsically nonparametric. While the intensity function of a point process that is only observed once on a fixed domain cannot be identified, we show how covariates and repeated observations of the process can be utilized to make consistent estimation possible, and we also derive asymptotic rates of convergence without invoking parametric assumptions.**

***Index Terms*— Cox process, Fréchet regression, intensity function, nonparametric regression, Wasserstein metric.**

I. Introduction

TEMPORAL point processes are encountered in insurance in the form of the claim arrival process, risk processes and ruin theory which targets the solvency of the insurer [1]; queue theory in operations research [2]; seismology; demand patterns in bike sharing systems [3]; or bid arrivals in online auctions [4], [5]. Single realizations of point processes have been well studied in the literature [2], [6]–[9]. One important target is the intensity function, due to its straightforward interpretation as the rate of occurrence of points per unit time [2]. In the context of seismology, one expects the intensity function of the aftershock arrival process to depend on the size of the earthquake that triggered the aftershocks. This exemplifies point process data for which the intensity function depends on covariates, and provides the motivation to develop flexible nonparametric methods for such data. Specifically, we propose a nonparametric regression method for point processes as responses, coupled with Euclidean predictors in $\mathbb{R}^p$.

Poisson processes are one of the most important point processes as they have been shown to provide successful models for a wide range of scientific applications involving random phenomena and allow the construction of more complex processes [2] such as the Cox or doubly stochastic Poisson process [6]. The estimation of the intensity function of a non-homogeneous Poisson process (NHPP) has found much interest in the literature. For a single realization of a NHPP, [9] proposed an approach based on penalized projection estimators but consistency towards the intensity function can only be achieved when the expected number of observed events diverges. A standard asymptotic framework has been to assume that the intensity of the observed process $\tilde{\lambda}(t)$ can be written as a scalar multiple of an underlying intensity of interest $\lambda(t)$, where the scalar is allowed to diverge and thus enables to observe increasingly more points [10]–[12]. For multiple realizations or replicated NHPP, [13] consider the situation where the intensities are common across replications but only differ in that they are time shifted at random i.i.d. time points with a known density function. In [14] a replicated NHPP framework with a common non-random underlying intensity function is considered with results on consistent estimation of the cumulative intensity function, while in [15] convergence results towards the common intensity function are obtained. These previous approaches do not incorporate covariates and thus do not study a regression framework. The study of second order summary statistics such as the pair correlation function [16] in the context of replications of point processes has also received interest [17].

In the context of stationary Cox processes [6], nonparametric kernel estimation of the intensity function for just one observed point process has been proposed [7], connecting this problem to kernel density estimation. However, when one has just one realization of the point process, no consistent estimator of the intensity function exists [18], due to the unavailability of a consistent estimator of the scale factor of the intensity function, i.e. $\int \lambda(t)dt$. In other work that explores the interface of spatio-temporal point processes and functional data analysis, [19] proposed a semi-parametric generalized linear mixed model with a latent process component, and established asymptotic properties under increasing domain asymptotics, a design assumption that is commonly used for spatial processes. We consider here a different scenario, where replications of a temporal point process are available, along with an Euclidean covariate $X \in \mathbb{R}^p$. A previous replicated point process regression approach [20] also dealt

Manuscript received November 28, 2020; revised October 14, 2021; accepted October 20, 2021. Date of publication November 8, 2021; date of current version January 20, 2022. This work was supported in part by the National Science Foundation under Grant DMS-2014626. An earlier version of this paper was presented in part at the Joint Virtual Conference CFE-CMStatistics 2020 (13th International Conference of the ERCIM WG on Computational and Methodological Statistics and 14th International Conference on Computational and Financial Econometrics). *(Corresponding author: Álvaro Gajardo.)*

The authors are with the Department of Statistics, University of California at Davis, Davis, CA 95616 USA (e-mail: aegajardo@ucdavis.edu).

Communicated by O. Johnson, Associate Editor for Probability and Statistics.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TIT.2021.3126466.

Digital Object Identifier 10.1109/TIT.2021.3126466

0018-9448 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

with repeatedly observed non-homogeneous Poisson processes as responses; however this previous approach relied heavily on parametric specifications, for which diagnostics is very difficult, whereas we aim here at a flexible nonparametric approach that applies much more generally.

In the context of Cox processes, where the intensity function $\Lambda(t)$ is a positive locally integrable random function, a common approach when dealing with repeatedly observed point processes (but without a covariate X) has been to combine likelihood methods and techniques from Functional Data Analysis [3], [21]–[24]. For example, in [23] a Karhunen-Loève expansion [25], [26] was applied to the log intensity random functions,

$$\log(\Lambda(t)) = \mu(t) + \sum_{k=1}^{\infty} U_k \phi_k(t), \quad 0 \le t \le T,$$

where $0 < T < \infty$ and $\mu \in L^2([0,T])$ is the mean function, the U_k are uncorrelated random variables and $\phi_1, \phi_2, \ldots$ form an orthonormal basis of $L^2([0,T])$. Then, by using techniques such as Functional Principal Component Analysis (FPCA), a truncated version of the expansion is considered with only p components and thus the functional problem of estimating μ and the ϕ_k is reduced to a multivariate approach, modeling the functions in a finite dimensional function space of basis functions like B-splines. Distributional assumptions such as Gaussianity of the U_k are also introduced in order to justify a likelihood approach to obtain estimates for the basis coefficients. The previous transformation approach is extrinsic since it does not target directly intensity functions, which are subject to a positivity constraint. This constraint makes the intensity space convex but not linear and thus functional data analysis (FDA) methods and especially functional principal components analysis (FPCA) directly applied to $\Lambda(t)$ are not well suited [27].

Similarly, [28] proposed a functional approach that decomposes the intensity function into an intensity factor and a shape function and then performed a Karhunen-Loève expansion for the shape function, borrowing strength across the replications. These methods face constraints due to the non-negative nature of the intensity function and cannot be directly extended to establish regression models for point processes. Since the random intensity functions are not observed, the event arrival times are used in the estimation procedures. For Cox processes, the key relation that allows this is that conditional on n events occurring in an interval $[0,T]$ and $\Lambda = \lambda$, the unordered arrival times form an i.i.d. sample with density $\lambda / \int \lambda$; this is a well-known property for non-homogeneous Poisson processes [2]. Furthermore, as noted in [28], the intensity function λ can be decomposed into a shape function $f = \lambda / \int \lambda$ and an intensity factor $\tau = \int \lambda$. This decomposition is the key relationship that will enable us to split the problem of estimating the intensity function conditional on covariates into two parts: Estimating the conditional shape function and estimating the conditional intensity factor of the process. While this decomposition is defining the structure of the problem, in order to achieve consistent estimation of the shape function, τ is assumed to diverge so that increasingly more points are available. Therefore, we observe a Cox process $(N^{(n)}, \Lambda)$ such that, conditional on Λ, $N^{(n)}(T)$ has a Poisson distribution with rate $\alpha_n \tau$ for some positive sequence $\alpha_n \to \infty$.

The proposed regression approach for the intensity function of replicated temporal point processes on Euclidean predictors utilizes conditional Fréchet means [29] for a suitable metric on the space of intensity functions, where this metric can be decomposed into two parts, one that quantifies differences in shape and a second part that quantifies differences in the intensity factors. As we need to estimate the density function associated with the arrival times of the point process, our asymptotic consistency results for the intensity function also utilize tools that were developed in [30]. The common assumption of letting the observation window $T \to \infty$ is often not applicable, including the data scenarios we consider below to motivate our methods. Therefore, we consider an asymptotic framework where T remains fixed and the number of replicates of the point process increases.

While in general the intensity function of a doubly stochastic Poisson process cannot be consistently estimated as it is random, we demonstrate here that the situation is different for point processes conditional on a covariate, as we establish asymptotic consistency with rates of convergence for conditional intensity functions. We illustrate the implementation of the proposed point process regression with simulations and show that it leads to well interpretable results for the Chicago Divvy bike trips and the New York yellow taxi trips data. An application to the earthquake aftershock process in Chile is presented in section VI-C.

The main innovations presented in this paper are: (1) We develop the first fully nonparametric regression method that features point processes as responses with Euclidean predictors; (2) We obtain asymptotic rates of convergence for conditional intensity functions, while such a result is not achievable for intensity functions unconditionally; (3) Our approach does not require functional principal components and does not require distributional assumptions, as it is not utilizing likelihoods; (4) The proposed approach is shown to work well in relevant applications.

II. The Space of Intensity Functions

Let $\{N(t), t \ge 0\}$ be a temporal point process where $N(t)$ represents the number of events that occur in the time interval $[0,t]$ and $N(0) = 0$. We suppose that $N(t)$ is observed on the time window $[0,T]$ for some endpoint $T > 0$, and is such that $m(t) := E(N(t)) < \infty$ for $0 \le t \le T$. In the context of replicated point processes that we consider here, it is natural to work within the framework of a doubly stochastic Poisson process (N, Λ) where one assumes that there is an underlying stochastic intensity process $\Lambda(t)$ that generates non-negative integrable functions on $[0,T]$ such that conditional on a realization $\Lambda = \lambda$, $N|\Lambda = \lambda$ is a non-homogeneous Poisson process with intensity function λ [6]. A feature that greatly facilitates analysis of such processes [3], [23], [28], [30] is the fact that a Poisson process has the order statistics property, i.e., conditional on m events being observed in $[0,T]$, the

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

successive event times are distributed as the order statistics of m independent and identically distributed random variables with a density that is proportional to the intensity function [2].

Denoting the space of intensity functions as

$$\Omega = \left\{ \Lambda \colon [0,T] \to \mathbb{R}^+ \text{ such that } \int_0^T \Lambda(t)dt < \infty \right\},$$

two key quantities for each $\Lambda \in \Omega$ are the *intensity factor*, the scalar

$$\tau := \int_0^T \Lambda(t)dt,$$

which is the expected number of events in $[0,T]$, conditional on $\Lambda(\cdot)$; and the *shape function*

$$f(t) := \frac{\Lambda(t)}{\tau}, \quad \text{where} \quad \int_0^T f(t)dt = 1,\ f(t) \geq 0, \ t \in [0,T],$$

which is a density function. Hence, $\Lambda(\cdot) = \tau f(\cdot)$ and since there is a one-to-one correspondence between Λ and (τ, f) we may regard Ω as a product space $\Omega = \Omega_{\mathcal{T}} \times \Omega_{\mathcal{F}}$ where $\Omega_{\mathcal{T}} = (0, \infty)$ and

$$\Omega_{\mathcal{F}} = \{ f \colon [0,T] \to \mathbb{R} \text{ such that } f \geq 0 \text{ and } \int_0^T f(t)dt = 1 \}.$$

For our theoretical results, we focus on a subspace of $\Omega_{\mathcal{F}}$ consisting of densities that are well behaved and bounded away from zero, see assumptions (S1) and (S2) in section IV. With a slight abuse of notation, we will continue to refer to this subspace as $\Omega_{\mathcal{F}}$.

Furthermore, if we endow $\Omega_{\mathcal{T}}$ and $\Omega_{\mathcal{F}}$ with metrics $d_{\mathcal{T}}$ and $d_{\mathcal{F}}$, respectively, we may regard (Ω, d) as a product metric space $(\Omega_{\mathcal{T}}, d_{\mathcal{T}}) \times (\Omega_{\mathcal{F}}, d_{\mathcal{F}})$, where

$$d((\tau_1, f_1), (\tau_2, f_2)) := \sqrt{d_{\mathcal{T}}^2(\tau_1, \tau_2) + d_{\mathcal{F}}^2(f_1, f_2)}. \quad (1)$$

In the context of metric geometry such product metric spaces for which the distance arises as an l^2-type norm between the underlying metrics have been extensively studied. In particular, it is well known that Ω is a geodesic space if and only if $\Omega_{\mathcal{T}}$ and $\Omega_{\mathcal{F}}$ are geodesic spaces [31]. This decomposition enables us to measure differences in shape and magnitude separately. We choose the Euclidean metric $d_{\mathcal{T}}(\tau_1, \tau_2) := d_E(\tau_1, \tau_2)$ and the 2-Wasserstein metric $d_{\mathcal{F}}(f_1, f_2) := d_W(\mu_1, \mu_2)$, which for two probability measures μ_1, μ_2 on $[0,T]$ with associated density functions f_1, f_2 and quantile functions Q_1, Q_2 is defined as [32]

$$d_W^2(\mu_1, \mu_2) = d_{L^2}^2(Q_1, Q_2) = \int_0^1 (Q_1(t) - Q_2(t))^2 \, dt,$$

where we assume throughout that these quantities exist and are well defined. The Wasserstein metric has been shown to be a most useful metric in practical applications that involve samples of distributions [33]. The 2-Wasserstein metric on the space of density functions $\Omega_{\mathcal{F}}$ has very rich geometrical interpretations due to its connections with optimal transport [34]. Although one could consider a metric based on the vertical alignment such as the L^2 metric, these metrics are not well suited for densities due to their inherent constraints $f \geq 0$ and $\int_0^T f(t)dt = 1$, which imply that the space $\Omega_{\mathcal{F}}$, while convex, is not a linear space [27].

A basic notion for statistical modeling is the mean of a random variable, where for random objects in a metric space $(\tilde{\Omega}, \tilde{d}, P)$ it has proved advantageous to adopt the barycenter or Fréchet mean [35], defined as $w_\oplus := \arg\min_{w \in \tilde{\Omega}} E(\tilde{d}^2(Y, w))$ where $Y \in \tilde{\Omega}$ is a random object. The Fréchet mean may be regarded as an extension of the standard concept of mean in Euclidean space to abstract metric spaces in the sense that when $\tilde{\Omega}$ is a convex subset of the Euclidean space and $\tilde{d}$ is the Euclidean metric, then the ordinary mean and the Fréchet mean coincide. The barycenter and its estimation has attracted much interest for distribution spaces with the Wasserstein metric [30], [34], [36]–[38]; we adopt these spaces here for the shape part of the intensity function.

III. Framework for Intensity Function Regression

A. Preliminaries

Our goal is to model the regression relation between random intensity functions Y in the above space (Ω, d) as responses and an Euclidean predictor $X \in \mathbb{R}$, for which we adopt the recently developed framework of Fréchet regression [29], which can be viewed as a generalization of Fréchet means to the more general notion of conditional Fréchet means. Formally, define the regression or conditional intensity function $m_\oplus(x)$ as

$$m_\oplus(x) := \arg\min_{w \in \Omega} M_\oplus(w, x),$$
$$M_\oplus(w, x) := E(d^2(Y, w)|X = x),$$

where $w = (\tau_0, f_0) \in \Omega = \Omega_{\mathcal{T}} \times \Omega_{\mathcal{F}}$ and $Y = (\tau, f) \in \Omega_{\mathcal{T}} \times \Omega_{\mathcal{F}}$, so that by (1),

$$M_\oplus(w, x) = E(d_{\mathcal{T}}^2(\tau, \tau_0)|X = x) + E(d_{\mathcal{F}}^2(f, f_0)|X = x).$$

Hence, the optimization problem is separable with optimal solution $m_\oplus(x) = (\tau_\oplus(x), f_\oplus(x))$, where

$$\tau_\oplus(x) = \arg\min_{\tau_0 \in \Omega_{\mathcal{T}}} E(d_{\mathcal{T}}^2(\tau, \tau_0)|X = x) = \max\{E(\tau|X = x), 0\}; \quad (2)$$

$$f_\oplus(x) = \arg\min_{f_0 \in \Omega_{\mathcal{F}}} E(d_{\mathcal{F}}^2(f, f_0)|X = x). \quad (3)$$

As we focus on a subspace of $\Omega_{\mathcal{F}}$ consisting of densities that are well behaved and bounded away from zero, the space of corresponding quantile functions $Q(\Omega_{\mathcal{F}})$ is a closed and convex subset of the Hilbert space $L^2([0,1])$ (see assumptions $(S1)$, $(S2)$ and Lemma 1 in section IV-A). By equivalently casting the optimization problem (3) in terms of quantile functions, as per the definition of the 2-Wasserstein metric, and employing properties of the L^2-inner product, Lemma S.10 in the Appendix implies the existence and uniqueness of the solution to this program. The solution admits a closed form in terms of the quantile functions $Q_\oplus(x)$ corresponding to $f_\oplus(x)$, and is given by $Q_\oplus(x) = E(Q|X = x)$. This shows that under regularity conditions $m_\oplus(x)$ exists and is unique. Note that since we measure the differences in the

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

intensity factor from the changes in shape separately, the order of magnitude between the two metrics is not relevant. More precisely, $m_\oplus(x)$ remains the same for weighted metrics $d_\lambda^2 = \alpha d_\mathcal{T}^2 + \beta d_\mathcal{F}^2$ with $\alpha, \beta > 0$.

The local Fréchet regression function $m_\oplus(x)$ operates directly on the space of intensity functions so that the regression is performed in the geometric space corresponding to the optimal transport geometry that is induced by the 2-Wasserstein metric on the shape components. Thus $m_\oplus(x)$ provides a notion of conditional center that is shaped by the underlying geometry of the data generating mechanism that produces random intensities. We show through a simulation example in section H in the Appendix that the standard Euclidean conditional intensity $E(\Lambda|X = x)$ can be severely distorted, while $m_\oplus(x)$ captures the underlying geometry of the problem and provides a better notion of center. As $m_\oplus(x)$ is a valid intensity function, it still enjoys the usual interpretation of rate of events per unit time, while representing the (conditional) point process barycenter at predictor level x.

A basic difficulty is that the function $\Lambda(\cdot)$ is not observed. If sufficiently many arrival times are observed for each replication of the point process, then it is well known that consistent estimation of the density function f may be achieved by classical density estimation techniques [7], [30], while some work-arounds exist for sparsely observed point processes [28]. However, the situation is much less benign regarding estimation of the intensity parameter τ. It would be natural to employ the total count $N(T)$, which is conditionally unbiased for τ in the sense that $E(N(T)|\Lambda) = \tau$ but is not (conditionally) consistent as $\mathrm{Var}(N(T)|\Lambda) = \tau$. When just one replicate of a point process is observed, consistent estimation of the intensity function is therefore not possible. Further motivating conditional intensity function modeling, we show in the following that the situation is different when considering conditional intensity functions. We demonstrate that the counts $N_i(T)$ can be used as initial estimates for the random intensity factors τ_i, from which consistent estimators can then be derived. This phenomenon is analogous to classical linear regression modeling, where one has errors in the responses and yet consistent estimation of the conditional expectation that corresponds to the true regression function is achieved. This provides strong motivation for the proposed methods and the study of conditional point processes.

B. Local Regression for Intensity Functions

Suppose that a sample of replicates (X_i, N_i, f_i, τ_i) is drawn from the joint distribution of (X, N, f, τ), $i = 1, \ldots, n$, where $N|\Lambda = f \times \tau$ is a Poisson process with intensity function $\lambda = f \times \tau$. We employ empirical weights from local linear regression [39] that are inherent to the local Fréchet regression approach [29], and are given by

$$s_{in}(x, h) = \frac{1}{\hat{\sigma}_0^2} K_h(X_i - x) \left[\hat{u}_2 - \hat{u}_1 (X_i - x)\right], \qquad (4)$$

where $\hat{u}_j = n^{-1} \sum_{i=1}^n K_h(X_i - x)(X_i - x)^j$ with $j \in \{0, 1, 2\}$, $\hat{\sigma}_0^2 = \hat{u}_0 \hat{u}_2 - \hat{u}_1^2$ and $K_h(\cdot) = h^{-1} K(\cdot/h)$, the kernel K is a continuous and symmetric density function with support $[-1, 1]$, and $h = h_n$ is a sequence of bandwidths.

As we can factorize the response into a density function and an intensity factor, one regression component is the conditional mean $E(\tau|X = x)$. For estimating this quantity we employ local linear regression. If the τ_i were observed, then a naive estimator for (2) would be given by

$$\hat{\tau}_\oplus(x) = \max\{0, n^{-1} \sum_{i=1}^n s_{in}(x, h) \tau_i\}, \qquad (5)$$

as local linear regression is a linear estimator, assigning weights $s_{in}(x, h)$ to the responses. However, this estimator is based on the intensity factors τ_i, which are not observed; we only observe the counts of arrivals $N_i(T)$ for each replicate of the point process. This difficulty can be resolved by noting that the observed counts $N_i(T)$ satisfy the relationship $E(N_i(T)|\Lambda_i) = \tau_i$, which enables us to replace τ_i by $N_i(T)$ in (5) since $E(N(T)|X) = E(E(N(T)|X, \Lambda)|X) = E(\tau|X)$ is on target. Hence, under suitable regularity conditions one readily obtains well known non-parametric convergence rates for the corresponding locally weighted least squares estimator of $\tau_\oplus(x)$ (2) [39]. As we require an increasing asymptotic intensity framework, our final empirical estimate for the intensity factor part will be presented in section IV-B.

If the densities f_i associated with point processes N_i were completely observed, we could implement local Fréchet regression on the space of densities for (3) [29],

$$\hat{f}_\oplus(x) = \operatorname*{arg\,min}_{f_0 \in \Omega_\mathcal{F}} n^{-1} \sum_{i=1}^n s_{in}(x, h) d_\mathcal{F}^2(f_i, f_0). \qquad (6)$$

We however only observe the arrival times, from which the densities f_i must be estimated, and this will induce an additional error that needs to be accounted for when analyzing the final estimator. Existence and uniqueness of the solution to the optimization program in (6) can be obtained by considering corresponding quantile functions, which we discuss next.

As noted by [40], one of the main differences compared to classical density estimation techniques is that in the context of point processes we cannot let the number of observations go to infinity as it is a random feature of the point process itself. Instead it is useful to consider an asymptotic framework where the intensity factors τ_i diverge to infinity, while the observation window $[0, T]$ remains fixed; such frameworks have been considered before in the literature and allow to add information everywhere on $[0, T]$ as opposed to the common domain asymptotics $T \to \infty$, which are often not applicable, see also [30] or [41]. This framework will be introduced in section IV. From now on, τ will denote a generic random intensity factor such that $\tau_i = \int_0^T \Lambda_i(s) ds \overset{iid}{\sim} \tau$, $i = 1, \ldots, n$.

The minimization problem (6) is easily solved by considering quantile functions. If Q_i is the quantile function corresponding to f_i, $i = 1, \ldots, n$ and $\hat{Q}_\oplus(\cdot, x) : [0, 1] \to [0, T]$ is the quantile function corresponding to the density $\hat{f}_\oplus(x)$ in (6),

$$\hat{Q}_\oplus(\cdot, x) = \operatorname*{arg\,min}_{q \in Q(\Omega_\mathcal{F})} n^{-1} \sum_{i=1}^n s_{in}(x, h) ||Q_i - q||^2_{L^2([0,1])},$$

where $Q(\Omega_\mathcal{F})$ is the space of quantile functions corresponding to densities in $\Omega_\mathcal{F}$. Standard properties of the $L^2([0, 1])$ inner

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

product imply (Proposition 1 in [29])

$$\hat{Q}_{\oplus}(\cdot,x) = \underset{q\in Q(\Omega_{\mathcal{F}})}{\arg\min}\ ||q - n^{-1}\sum_{i=1}^{n} s_{in}(x,h)Q_i||^2_{L^2([0,1])}. \quad (7)$$

Existence and uniqueness of the solution of (7) and therefore of (6) is guaranteed as $\hat{Q}_{\oplus}(\cdot,x)$ corresponds to the orthogonal projection of $n^{-1}\sum_{i=1}^{n} s_{in}(x,h)Q_i$ as an element of the Hilbert space $L^2([0,1])$ on the closed and convex set $Q(\Omega_{\mathcal{F}})$ as shown in Lemma 1 under regularity conditions on the space $\Omega_{\mathcal{F}}$.

To discuss estimation of the Q_i, which are needed in (7) but are not directly available, it is helpful to consider auxiliary probability measures $\hat{\mu}_i$ on $[0,T]$ that correspond to the empirical measure of the arrival times when the total count $N_i(T) \geq 1$ and to the uniform measure on $[0,T]$ otherwise (see [30]). That is,

$$\hat{\mu}_i = \begin{cases} \dfrac{1}{N_i(T)} \displaystyle\sum_{j=1}^{N_i(T)} \delta_{Z_{ij}}, & \text{if } N_i(T) \geq 1 \\ \frac{1}{T}\ \mathcal{L}, & \text{if } N_i(T) = 0, \end{cases}$$

where Z_{ij} are the arrival times of the point process N_i and $\mathcal{L}$ is the Lebesgue measure on $[0,T]$. For a probability measure μ on $[0,T]$ with cdf F_μ we consider its quantile function $Q_\mu(t) := \inf\{x \in [0,T]\colon F_\mu(x) \geq t\}$. Let $\hat{Q}_i := Q_{\hat{\mu}_i}$, then replacing Q_i by $\hat{Q}_i$ in (7) leads to the empirical estimate

$$\tilde{Q}_{\oplus}(\cdot,x) := \underset{q\in Q(\Omega_{\mathcal{F}})}{\arg\min}\ ||q - n^{-1}\sum_{i=1}^{n} s_{in}(x,h)\hat{Q}_i||^2_{L^2([0,1])}. \quad (8)$$

IV. Asymptotic Results

A. Convergence of the Shape Function Estimates

Let $\tilde{f}_{\oplus}(x)$ be the density function corresponding to the quantile function $\tilde{Q}_{\oplus}(\cdot,x)$. Thus, $\tilde{f}_{\oplus}(x)$ corresponds to the empirical estimate for (6). We require the following assumptions, which guarantee that $Q(\Omega_{\mathcal{F}})$ is a closed and convex subset of the Hilbert space $L^2([0,1])$, yielding existence and uniqueness of the naive and empirical estimators in (7) and (8), respectively.

(S1) Suppose that there exists $0 < M < L < \infty$ such that $Q \in Q(\Omega_{\mathcal{F}})$ if

$$M|x-y| \leq |Q(x) - Q(y)| \leq L|x-y|,$$

for all $x, y \in [0,1]$.

(S2) Suppose that for any $Q \in Q(\Omega_{\mathcal{F}})$ it holds that $Q(0) = 0$ and $Q(1) = T$.

These assumptions are needed to ensure that the quantile functions do not increase too rapidly or too slowly, which is equivalent to constraining the corresponding density functions to be well behaved and bounded away from zero.

Lemma 1: Under $(S1) - (S2)$, $Q(\Omega_{\mathcal{F}})$ is a closed and convex set on the Hilbert space $L^2([0,1])$.

Lemma 1 guarantees existence and uniqueness of the local Fréchet regression function on shape space $f_{\oplus}(x)$ defined in (3). By employing the Hilbert space structure of $L^2([0,1])$ and properties of the 2-Wasserstein metric, Lemma S.10 in the Appendix shows that $f_{\oplus}(x)$ corresponds to the density function with associated quantile function $E(Q|X = x)$; the latter can be shown to reside in $Q(\Omega_{\mathcal{F}})$. Regarding the convergence of $\tilde{f}_{\oplus}(x)$ towards $f_{\oplus}(x)$ in the 2-Wasserstein metric, by the triangle inequality

$$d_{\mathcal{F}}(f_{\oplus}(x), \tilde{f}_{\oplus}(x)) \leq d_{\mathcal{F}}(f_{\oplus}(x), \hat{f}_{\oplus}(x)) + d_{\mathcal{F}}(\hat{f}_{\oplus}(x), \tilde{f}_{\oplus}(x)), \quad (9)$$

and for the term $d_{\mathcal{F}}(\hat{f}_{\oplus}(x), \tilde{f}_{\oplus}(x))$ we observe that properties of the orthogonal projection on a closed and convex set in the Hilbert space $L^2([0,1])$ imply that

$$\begin{aligned} d_{\mathcal{F}}(\tilde{f}_{\oplus}(x), \hat{f}_{\oplus}(x)) &= ||\tilde{Q}_{\oplus}(\cdot,x) - \hat{Q}_{\oplus}(\cdot,x)||_{L^2([0,1])} \\ &\leq n^{-1}\sum_{i=1}^{n} |s_{in}(x,h)|\ ||\hat{Q}_i - Q_i||_{L^2([0,1])}. \end{aligned} \quad (10)$$

Therefore, convergence hinges on consistent estimation of the quantile functions Q_i. For the asymptotic framework, we consider a new sequence of Cox point processes $N_i^{(n)}$ which have the same shape function f_i but an increasing intensity factor $\alpha_n\tau_i$ so that increasingly more points are observed per point process as sample size increases [30]. The asymptotic framework is then as follows:

1) A first random mechanism generates pairs of predictors X_i and intensity functions Λ_i, $(X_1,\Lambda_1),\ldots,(X_n,\Lambda_n) \overset{iid}{\sim} (X,\Lambda)$ which encapsulates the dependency between these random quantities. While the X_i are observed, the Λ_i are not observed.
2) Given the random intensity functions Λ_i, a second independent random mechanism then generates the observable number of arrivals $N_i^{(n)}(T)$ and the arrival times $Z_{i1},\ldots,Z_{iN_i^{(n)}(T)}$ for the i-th point process $N_i^{(n)}$, $i = 1,\ldots,n$.
3) Conditional on $N_i^{(n)}(T)$, the (unordered) arrival times $Z_{i1},\ldots,Z_{iN_i^{(n)}(T)} \overset{iid}{\sim} f_i$.
4) Given the random intensity function Λ_i, $N_i^{(n)}(T) \sim \mathcal{P}(\alpha_n\tau_i)$ for a positive sequence $\alpha_n \to \infty$, where $\mathcal{P}(\eta)$ denotes a Poisson random variable with rate η.

We note that conditions 1-3 are standard in the context of Cox processes while condition 4 allows for the observable number of arrivals $N_i^{(n)}(T)$ to diverge as n increases and to avoid empty point processes [30], which is the key for consistent estimation of Q_i by using the empirical measure of the arrival times $\hat{\mu}_i$ with $N_i^{(n)}(T)$ in place of $N_i(T)$. A similar fixed domain asymptotic point process framework was considered in [30] but without covariates.

The following result shows that the second term on the right hand side of (9) is $O_p(\alpha_n^{-1/4})$ provided that the support of τ is bounded away from zero and α_n grows fast enough.

(S3) There exists a scalar $\kappa > 0$ such that $\tau \geq \kappa$ almost surely.

Proposition 1: Suppose that $(S1)$, $(S2)$ and $(S3)$ hold, the marginal density f_X of X satisfies $f_X(\cdot) > 0$ and is twice continuously differentiable, $\dfrac{\alpha_n}{\log n} \to \infty$ as $n \to \infty$, and the

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

bandwidth sequence $h = h_n$ satisfies $h_n \to 0$ and $nh_n \to \infty$ as $n \to \infty$. Then

$$d_{\mathcal{F}}(\hat{f}_{\oplus}(x), \tilde{f}_{\oplus}(x)) = O_p(\alpha_n^{-1/4}).$$

The term $d_{\mathcal{F}}(f_{\oplus}(x), \hat{f}_{\oplus}(x))$ on the right hand side of (9) was shown to be $O_p(n^{-2/5})$ under the following regularity condition [29]:

(L1) The marginal density f_X of X, as well as the conditional densities g_y of $X|\tilde{Y} = y$, exist for $\tilde{Y} \in \Omega_{\mathcal{F}}$ and are twice continuously differentiable, the latter for all $y \in \Omega_{\mathcal{F}}$, and $\sup_{x,y} |g_y''(x)| < \infty$. Additionally, for any open $U \subset \Omega_{\mathcal{F}}$, $\int_U \mathrm{d}F_{\tilde{Y}|X}(x, y)$ is continuous as a function of x.

Summarizing these results, we obtain

Theorem 1: Suppose that $(S1)$, $(S2)$, $(S3)$ and $(L1)$ hold, the density function satisfies $f_X(\cdot) > 0$, $\dfrac{\alpha_n}{\log n} \to \infty$ as $n \to \infty$ and $h = h_n \sim c_0 n^{-1/5}$ for some constant $c_0 > 0$. Then

$$d_{\mathcal{F}}(f_{\oplus}(x), \tilde{f}_{\oplus}(x)) = O_p(n^{-2/5} + \alpha_n^{-1/4}).$$

Accordingly, consistent estimation in the 2-Wasserstein metric for the shape part of the conditional intensity function can be achieved at the rate $O_p(n^{-2/5})$ as long as $n^{8/5}\alpha_n^{-1}$ is bounded above. If this assumption holds, the well-known rate of convergence for local linear regression with real valued responses is thus obtainable.

B. Convergence of the Intensity Factor Estimates

In the increasing asymptotics framework that was introduced in the previous section, we assumed that there is a common intensity factor multiplier α_n such that $\alpha_n \to \infty$ as $n \to \infty$. This led to consistent estimation of the conditional intensity functions for the shape function part of the intensity function, where one works in the density space $\Omega_{\mathcal{F}}$. Since intensity functions can be factorized into a shape part, which corresponds to a density function, and an intensity factor, it remains to construct an estimator for the intensity factor (2) conditional on predictors X.

It turns out that this is a challenge, as an estimator for (5) is not easily available. This is because in order to estimate the shape functions consistently, it is necessary to assume that the expected number of events increases without bound. We show in the following how this challenge can be overcome and consistent estimation of $E(\tau|X)$ is nevertheless still possible up to the constant $E(\tau)$, so that relative intensities can be estimated consistently. The key to achieve this is to utilize the average observed number of arrivals $\bar{N}(T) := n^{-1}\sum_{i=1}^n N_i^{(n)}(T)$. We require the following regularity conditions.

(LL1) The regression function $m(x) = E(\tau|X = x)$, the density function $f_X(x) > 0$ of X and $\sigma^2(x) = E(e^2|X = x)$, where $e = \tau - m(X)$, are twice continuously differentiable in x.

(LL2) For the bandwidth sequence h, $nh^5 = O(1)$, as $n \to \infty$.

(LL3) There exists $\delta > 0$ and $\bar{\sigma} > 0$ such that $E(|e|^{2+\delta}|X) \leq \bar{\sigma}$, for all predictors X.

We note that assumption $(LL1)$ is a basic smoothness assumption that is needed to expand the bias for local linear smoothing, while $(LL2)$ is also a common assumption and implies that as $n \to \infty$ and for $q \in \mathbb{N}$, we have $nh^q \to \infty$ if $0 \leq q \leq 4$ and $nh^q \to 0$ for $q > 5$. Assumption $(LL3)$ will be used for an application of the central limit theorem.

Our main result on conditional intensity estimation is as follows. For two sequences β_n and γ_n, denote by $\beta_n \asymp \gamma_n$ if $c_1\beta_n \leq \gamma_n \leq c_2\beta_n$ for some constants $c_1, c_2 > 0$. Denote by $\tilde{\tau}_{\oplus}(x)$ the empirical and standardized estimate of the intensity factor part, which is given by

$$\tilde{\tau}_{\oplus}(x) = \max\left(0, \frac{n^{-1}\sum_{i=1}^n s_{in}(x, h)N_i^{(n)}(T)}{\bar{N}(T)}\right). \quad (11)$$

Theorem 2: Under $(LL1) - (LL3)$ and $(S3)$, suppose that $\tau \leq M_1$ almost surely for a constant $M_1 \in [\kappa, \infty)$ with κ as in assumption (S3). If $\psi(\alpha_n) \asymp n^{2/5}$ for some function $\psi : \mathbb{R}^+ \to \mathbb{R}$, $h = h_n = c_0 n^{-1/5}$ for some constant $c_0 > 0$ and $\dfrac{\alpha_n}{\log n} \to \infty$ as $n \to \infty$, then

$$\tilde{\tau}_{\oplus}(x) = \frac{1}{E(\tau)}\tau_{\oplus}(x) + O_p(n^{-2/5}).$$

This means that $\tau_{\oplus}(x)$ can still be consistently estimated up to the constant $E(\tau)$ by using the observable numbers of arrivals $N_i^{(n)}(T)$ of each replication of the point process instead of the true intensity factors τ_i, which are not observed. Furthermore, as the observed counts $N_i^{(n)}(T)$ grow with α_n as $n \to \infty$, we can stabilize the local linear estimator by employing comparisons against the average number of arrivals $\bar{N}(T)$. We remark that even though the quantity $E(\tau)$ is unknown, relative intensity factors at different covariate levels can still be recovered consistently, which is a key result of interest in our framework.

The assumptions require that $\psi(\alpha_n)$ does not increase faster than $n^{2/5}$ for some function $\psi : \mathbb{R}^+ \to \mathbb{R}$. This is due to the fact that local linear regression estimators with real valued responses are employed. These have a well-known optimal rate of convergence $O_p(n^{-2/5})$ under mild assumptions, which is obtained under our assumptions for general growth rates of α_n. For example, if $\alpha_n = c_1 n^{\rho}$ has a polynomial growth rate, where $c_1, \rho > 0$, then our assumptions are satisfied by taking $\psi\colon t \to t^{2/(5\rho)}$, $t > 0$, leading to the optimal rate $O_p(n^{-2/5})$. Similarly, if $\alpha_n = c_1 \exp(n\gamma)$ has an exponential growth rate, where $c_1, \gamma > 0$, then the conditions are satisfied by taking $\psi\colon t \to (\log(t)/\gamma)^{2/5}$, $t > 0$.

C. Convergence of the Conditional Intensity Function Estimates

We are now in position to construct an estimate for the conditional intensity function by combining our previous results. Recall that the regression or conditional intensity function satisfies $m_{\oplus}(x) = \tau_{\oplus}(x)f_{\oplus}(x)$ where $\tau_{\oplus}(x)$ and $f_{\oplus}(x)$ are defined in (2) and (3), respectively, and

$$\tilde{\tau}_{\oplus}(x) = \max\left(0, \frac{n^{-1}\sum_{i=1}^n s_{in}(x, h)N_i^{(n)}(T)}{\bar{N}(T)}\right),$$

which corresponds to the estimate of $\tau_{\oplus}(x)$, up to the constant $E(\tau)$, as per Theorem 2.

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

Since $\Lambda_\oplus(x) = \tau_\oplus(x) f_\oplus(x)$, we obtain an estimate of $\Lambda_\oplus(x)$ by plugging in the previously obtained estimates of the intensity factor $\tau_\oplus(x)$ and of the shape function $f_\oplus(x)$, leading to

$$\hat{\Lambda}_\oplus(x) = \tilde{\tau}_\oplus(x)\tilde{f}_\oplus(x). \tag{12}$$

Here $\tilde{f}_\oplus(x)$ is the density corresponding to the quantile function defined in (8). This estimator is consistent for the conditional intensity function up to $E(\tau)$, as per the following result.

Corollary 1: Under the regularity conditions of Theorems 1 and 2, the estimate $\hat{\Lambda}_\oplus(x)$ of the conditional intensity function $m_\oplus(x)$ is consistent up to the constant $E(\tau)$ in the sense that

$$d(m_\oplus(x), E(\tau)\hat{\Lambda}_\oplus(x)) = O_p(n^{-2/5} + \alpha_n^{-1/4}),$$

where $\alpha_n / \log n \to \infty$ as $n \to \infty$ and $\psi(\alpha_n) \asymp n^{2/5}$ for some function $\psi : \mathbb{R}^+ \to \mathbb{R}$.

This result follows directly from Theorems 1 and 2. If the sequence α_n has a polynomial growth rate $\alpha_n / n^\rho \to C$ as $n \to \infty$ for some $\rho > 0$ and $C > 0$, then the convergence rate in the Corollary is $O_p(n^{-\rho/4})$ if $\rho \in (0, 8/5)$ while the well-known non-parametric rate for local linear regression with real valued responses $O_p(n^{-2/5})$ is achieved whenever $\rho \geq 8/5$. The fastest convergence rate achievable is obtained when ρ is at least $8/5$ which leads to $d(m_\oplus(x), E(\tau)\hat{\Lambda}_\oplus(x)) = O_p(n^{-2/5})$. In this case, both the estimation of the intensity factor part, up to the constant $E(\tau)$, and the shape part of the conditional intensity function can be recovered at the rate $O_p(n^{-2/5})$.

We remark that in the special case where $\mathrm{Var}(\tau) = 0$ and the distribution of the random density f corresponds to a point mass in the Wasserstein space of probability distributions $\Omega_\mathcal{F}$ endowed with the 2-Wasserstein metric, one has that $f = g$ almost surely for some density g with corresponding quantile function $Q_g \in Q(\Omega_\mathcal{F})$ and $\tau = \eta_0$ almost surely for some constant $\eta_0 \in [\kappa, M_1]$ with κ, M_1 as in Theorem 2. This setting corresponds to the situation when there is no regression of the point process on X and is still covered by Corollary 1. Moreover, in this special case our framework is equivalent to that of replicated Poisson processes as the underlying intensity functions are non-random and identical. Lemma S.12 in the Appendix shows that $\tilde{\tau}_\oplus(x) = \eta_0$ and $\tilde{f}_\oplus(x) = g$ so that $m_\oplus(x)$ is equivalent to the underlying common intensity. Thus the problem translates into one of consistent estimation of the common underlying intensity function across independent replications of a single Poisson process, which we obtain up to a constant. In this direction, several works exist such as [15] where a non-parametric estimate of the underlying intensity function is considered and pointwise as well as MSE convergence results are derived. Parametric approaches have also been extensively studied; see [42]–[46] for further details. Non-parametric approaches using wavelets have also been explored in [47] and semiparametric approaches in [48]. When replications of a non-homogeneous Poisson process with a common and non-random underlying intensity function are available [15], one can readily exploit this fact so that an asymptotic infill framework is not required in this situation; rather, one can pull observations together to estimate the common shape or intensity factor components, however the situation is different in the regression framework that we study here.

It is often of interest to study the association between categorical predictors and point processes, for which the previously studied local regression approaches are not applicable as continuity of the predictor X is required. The next section is devoted to the construction of a second regression model that hinges on a generalization of the classical parametric multivariate linear regression model in the Euclidean case, and allows to address this problem.

D. Global Regression Framework for Intensity Functions

We briefly demonstrate here a generalization of multiple linear regression to the case where responses are point processes that allows the inclusion of categorical predictors while responses are objects residing in intensity space (Ω, d). The key is a characterization of multiple linear regression as a weighted sum of the responses, which can then be generalized to the case of weighted Fréchet means [29].

Consider an Euclidean predictor $X \in \mathbb{R}^p$ and assume that $\mu := E(X)$ and $\Sigma := \mathrm{Var}(X)$ exist, with Σ positive definite. In particular, this allows to consider either continuous or categorial predictors. The standard linear regression setting for $(X, \tilde{Y}) \in \mathbb{R}^p \times \mathbb{R}$ is that the regression function $E(\tilde{Y}|X = x) = \beta_0 + \beta_1^T(x - \mu)$ is linear in x, where β_0 and β_1 are the scalar intercept and slope vector, respectively. [29] recharacterized the linear regression function as $E(\tilde{Y}|X = x) = \operatorname{arg\,min}_{y \in \mathbb{R}} E(s(X, x) d_E^2(\tilde{Y}, y))$, where $s(X, x) := 1 + (X - \mu)^T \Sigma^{-1}(x - \mu)$ are weights that vary with x and d_E is the Euclidean metric. This allows a direct generalization to linear regression in intensity space (Ω, d) by simply replacing $\tilde{Y}$ by the object $Y \in \Omega$ and the standard Euclidean distance d_E by the metric d in intensity space, which inherits properties of the standard linear regression setup as we show below.

The global regression function of $Y \in \Omega$ on X is given by

$$m_{\mathrm{G}\oplus}(x) := \operatorname*{arg\,min}_{w \in \Omega} E(s(X, x) d^2(Y, w)).$$

Although Ω is not a linear space due to the non-negative nature of the intensity functions, the global regression curve $m_{\mathrm{G}\oplus}(x)$ passes through the Fréchet mean of Y at $x = \mu$ since $s(X, \mu) = 1$, a feature inherent to linear regression models. Moreover, the weights $s(X, x)$ can be negative, do not necessarily decay to zero away from x, and do not depend on a tuning parameter like local methods do. Arguments similar to those outlined in section III-A show that $m_{\mathrm{G}\oplus}(x) = (\tau_{\mathrm{G}\oplus}(x), f_{\mathrm{G}\oplus}(x))$, where

$$\begin{aligned} \tau_{\mathrm{G}\oplus}(x) &= \operatorname*{arg\,min}_{\tau_0 \in \Omega_\mathcal{T}} E(s(X, x) d_\mathcal{T}^2(\tau, \tau_0)) \\ &= \max\{E(s(X, x)\tau), 0\}; \end{aligned} \tag{13}$$

$$f_{\mathrm{G}\oplus}(x) = \operatorname*{arg\,min}_{f_0 \in \Omega_\mathcal{F}} E(s(X, x) d_\mathcal{F}^2(f, f_0)). \tag{14}$$

Lemma 1 guarantees existence and uniqueness of the global Fréchet regression function on shape space $f_{\mathrm{G}\oplus}(x)$. Similarly as in the local framework, Lemma S.11 in the Appendix shows

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

that the Hilbert space structure allows to characterize the quantile function corresponding to $f_{G\oplus}(x)$ as the orthogonal projection of $Q_x = E(s(X,x)Q)$ as an element of $L^2([0,1])$ on the closed and convex set $Q(\Omega_{\mathcal{F}})$. Suppose that a sample of replicates $(X_i, N_i, f_i, \tau_i) \overset{iid}{\sim} (X, N, f, \tau)$, where $N|\Lambda = f \times \tau$ is a Poisson process with intensity function $\lambda = f \times \tau$, is available and consider the same asymptotic framework as outlined in section IV-A. To obtain empirical estimates, define $s_{in}(x) := 1 + (X_i - \bar{X})^T \hat{\Sigma}^{-1}(x - \bar{X})$, where $\bar{X} := n^{-1}\sum_{i=1}^n X_i$ and $\hat{\Sigma} := n^{-1}\sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^T$. The next result shows that the global regression function in $\Omega_{\mathcal{T}}$ space, $\tau_{G\oplus}(x)$, can be consistently estimated up to the constant $E(\tau)$.

Theorem 3: Suppose that (S3) holds and $\tau \leq M_1$ almost surely for some constant $M_1 \in [\kappa, \infty)$ with κ as in assumption (S3). If $\psi(\alpha_n) \asymp n^{1/2}$ for some function $\psi : \mathbb{R}^+ \to \mathbb{R}$ and $\alpha_n / \log n \to \infty$ as $n \to \infty$, then

$$\max\left(0, \frac{n^{-1}\sum_{i=1}^n s_{in}(x) N_i^{(n)}(T)}{\bar{N}(T)}\right) = \frac{1}{E(\tau)}\tau_{G\oplus}(x) + O_p(n^{-1/2}).$$

Theorem 3 shows that the parametric $\sqrt{n}$-convergence rate can be obtained under mild conditions on the growth of α_n. Thus faster rates are obtained compared to the local setting. For example, if $\alpha_n = c_1 n^{\rho}$ has a polynomial growth rate, where $c_1, \rho > 0$, then the assumptions are satisfied by taking $\psi\colon t \to t^{1/(2\rho)}$, $t > 0$, which leads to the optimal $\sqrt{n}$-rate. Similarly, if $\alpha_n = c_1 \exp(n\gamma)$ has an exponential growth rate, where $c_1, \gamma > 0$, then the conditions are satisfied by taking $\psi\colon t \to (\log(t)/\gamma)^{1/2}$, $t > 0$.

Similarly as in the local regression setup, the shape components f_i remain unobserved and must be estimated from the arrival times across each replication. We consider the same estimation scheme for the shape functions as outlined in section III-B but replacing the local weights $s_{in}(x,h)$ by the global weights $s_{in}(x)$. This leads to the empirical estimate $\tilde{f}_{G\oplus}(x)$ of $f_{G\oplus}(x)$. The following result shows consistency of the estimated global regression function in $\Omega_{\mathcal{F}}$ space.

Theorem 4: Suppose that $(S1)$, $(S2)$ and $(S3)$ hold, and $\dfrac{\alpha_n}{\log n} \to \infty$ as $n \to \infty$. Then

$$d_{\mathcal{F}}(f_{G\oplus}(x), \tilde{f}_{G\oplus}(x)) = O_p(n^{-1/2} + \alpha_n^{-1/4}).$$

Thus, if α_n has a polynomial growth rate $\alpha_n / n^{\rho} \to C$ as $n \to \infty$ for some $\rho > 0$ and $C > 0$, we obtain the $\sqrt{n}$-rate as long as $\rho \geq 2$. If $\rho \in (0,2)$, then the rate achieved is $O_p(n^{-\rho/4})$. The following corollary summarizes the consistency, up to the constant $E(\tau)$, of the empirical estimate $\hat{\Lambda}_{G\oplus}(x) := \max\left(0, \frac{n^{-1}\sum_{i=1}^n s_{in}(x) N_i^{(n)}(T)}{\bar{N}(T)}\right) \tilde{f}_{G\oplus}(x)$ of the global regression function $m_{G\oplus}(x)$.

Corollary 2: Under the regularity conditions of Theorems 3 and 4, the estimate $\hat{\Lambda}_{G\oplus}(x)$ of the conditional intensity function $m_{G\oplus}(x)$ is consistent up to the constant $E(\tau)$ in the sense that

$$d(m_{G\oplus}(x), E(\tau)\hat{\Lambda}_{G\oplus}(x)) = O_p(n^{-1/2} + \alpha_n^{-1/4}),$$

where $\alpha_n / \log n \to \infty$ as $n \to \infty$ and $\psi(\alpha_n) \asymp n^{1/2}$ for some function $\psi : \mathbb{R}^+ \to \mathbb{R}$.

Thus, when α_n has a polynomial growth rate $\alpha_n / n^{\rho} \to C$ as $n \to \infty$ for some $\rho > 0$ and $C > 0$, the parametric $\sqrt{n}$-rate is achieved whenever $\rho \geq 2$ and otherwise the rate is $O_p(n^{-\rho/4})$ if $\rho \in (0,2)$, which is attained by the estimation of the corresponding shape component part.

Similarly as in section IV-C, the special case of no regression on X when $\tau = \eta_0$ almost surely for some positive constant $\eta_0 \in [\kappa, M_1]$ and $f = g$ almost surely for some density g with corresponding quantile function $Q_g \in Q(\Omega_{\mathcal{F}})$ is still covered by Corollary 2. Lemma S.13 in the Appendix shows that in this case $\tau_{G\oplus}(x) = \eta_0$ and $f_{G\oplus}(x) = g$. Thus the global Fréchet regression function coincides with the local version and is equivalent to the common underlying intensity function across independent replications of a single Poisson process. Here convergence towards $m_{G\oplus}(x) = \eta_0 g$, up to a constant, can be achieved at the parametric rate if α_n grows faster than n^2.

V. Simulations

A. Numerical Approximation for the Shape Component Estimates

The minimization problem for the local Fréchet regression on the shape component as in (8) is solved numerically and similar to the quadratic optimization problem considered in [29]. Recall that

$$\tilde{Q}_{\oplus}(\cdot, x) = \underset{q \in Q(\Omega_{\mathcal{F}})}{\arg\min} \; ||q - n^{-1}\sum_{i=1}^n s_{in}(x,h)\hat{Q}_i||^2_{L^2([0,1])}.$$

Let r_j, $j = 1, \ldots, \nu$, be an equispaced grid in $(0,1)$, where $\Delta r_\nu = 1/(\nu+1)$ is the grid spacing and ν is a positive integer. Denote the objective function by $\mathcal{M}(q) = ||q - W||^2_{L^2([0,1])}$, where $W(\cdot) = n^{-1}\sum_{i=1}^n s_{in}(x,h)\hat{Q}_i(\cdot)$ and $q \in Q(\Omega_{\mathcal{F}})$. We employ Riemann sum approximations to numerically solve (8) as follows. Letting $w_j := W(r_j) = n^{-1}\sum_{i=1}^n s_{in}(x,h)\hat{Q}_i(r_j)$, $j = 1, \ldots, \nu$, the Riemann sum approximation of $\mathcal{M}(q)$ is given by $\mathcal{M}_\nu(q) := \sum_{j=1}^{\nu}(q_j - w_j)^2 \Delta r_\nu$, where $q_j = q(r_j)$. Replacing $\mathcal{M}(q)$ by $\mathcal{M}_\nu(q)$ in (8) leads to an intermediate discretized optimization problem

$$\mathcal{S}_\nu := \underset{q \in Q(\Omega_{\mathcal{F}})}{\arg\min} \; \mathcal{M}_\nu(q).$$

Here $\mathcal{S}_\nu$ is the set of all such optimal solutions in $Q(\Omega_{\mathcal{F}})$. This set is non-empty, which can be seen by considering an auxiliary quadratic convex optimization program

$$q_\nu^* := \underset{\tilde{q} \in \mathbb{R}^\nu}{\arg\min} \; \|\tilde{q} - w\|_E^2, \tag{15}$$

subject to the constraints $0 < \tilde{q}_1 \leq \cdots \leq \tilde{q}_\nu < T$, $M\Delta r_\nu \leq \tilde{q}_{j+1} - \tilde{q}_j \leq L\Delta r_\nu$, $j = 1, \ldots, \nu - 1$, $M\Delta r_\nu \leq \tilde{q}_1 \leq L\Delta r_\nu$, and $M\Delta r_\nu \leq 1 - \tilde{q}_\nu \leq L\Delta r_\nu$, where $w = (w_1, \ldots, w_\nu)^T$. Therefore any function $q \in Q(\Omega_{\mathcal{F}})$ that interpolates the values q_ν^* at the grid points $r_1, \ldots, r_\nu$ belongs to $\mathcal{S}_\nu$. The next proposition shows that $\tilde{Q}_{\oplus}(\cdot, x)$ can be well recovered in the L^2-norm by choosing a sufficiently fine grid. Let $Q_{\oplus\nu}$ be any (fixed) element in $\mathcal{S}_\nu$, which can be selected by the axiom of choice.

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

Proposition 2: Suppose that $(S1)$ and $(S2)$ hold. Then

$$||\tilde{Q}_{\oplus}(\cdot,x) - Q_{\oplus\nu}||_{L^2([0,1])} = o(1),$$

as $\nu \to \infty$.

A natural element in $\mathcal{S}_\nu$ corresponds to the standard linear interpolation function constructed from q_ν^*, which is given by $Q_\nu^*(t) = q_{\nu,j}^* + (t - r_j)(q_{\nu,j+1}^* - q_{\nu,j}^*)/\Delta r_\nu$ for $t \in [r_j, r_{j+1})$, $j = 0, \ldots, \nu$, where $q_{\nu,j}^*$ is the jth coordinate of q_ν^*, $j = 1, \ldots, \nu$, and $q_{\nu,0}^* = 0$, $r_0 = 0$, $q_{\nu,\nu+1}^* = T$ and $r_{\nu+1} = 1$. By continuity, we define $Q_\nu^*(1) := \lim_{t\to 1-} Q_\nu^*(t) = T$ as the left-limit. Lemma S.9 in the Appendix shows that $Q_\nu^* \in Q(\Omega_{\mathcal{F}})$ and thus Q_ν^* lies in $\mathcal{S}_\nu$. In practice, M, L are taken as very small/large constants, $M = 10^{-10}$ and $L = 10^{10}$. This choice works very well in practice. The optimization problem (15) is a quadratic convex program (QP) with linear constraints similar to the one considered in [29] but slightly modifying the constraint matrix associated with the QP, and can be solved using state of the art optimization routines. The linear interpolation Q_ν^* of the optimal discrete solution q_ν^* corresponds to a discretized version of $\tilde{Q}_{\oplus}(\cdot,x)$ which is then mapped back to density space to obtain a discrete approximation of the corresponding density function $\tilde{f}_{\oplus}(\cdot,x)$. The latter step is performed by first constructing the cdf associated with q_ν^* and then utilizing local linear smoothing methods [49]. The implementation of the global regression is similar.

B. Simulations for Local Fréchet Regression

To assess the finite sample performance of the proposed conditional intensity function estimates, we constructed a generative model that produces simulated random intensity functions $\Lambda(\cdot) = f(\cdot)\tau$ along with an Euclidean predictor $X \in \mathbb{R}$. First, to generate a random density function f we consider the transformation to a Hilbert space approach using the log quantile density transformation (LQD) [27], where a Karhunen-Loève (KL) decomposition is employed for the transformed density, which is an element of L^2, and the latter curve is mapped back to density space. Specifically, denoting by $\psi : f \to -\log[f(Q(\cdot))]$ the LQD transform of f, where Q is the quantile function corresponding to the density f, we consider a truncated KL decomposition [50] conditionally on $X = x$

$$\psi(f)(\cdot) = \mu(\cdot,x) + \sum_{k=1}^{K} \xi_k(x)\phi_k(\cdot), \tag{16}$$

where $\mu(\cdot,x)$ is the (conditional) mean function of the L^2 process ψ, which we assume Gaussian, the $\xi_k(x)$ are independent across k such that $\xi_k(x) \sim N(0, \upsilon_k(x))$, where the eigenvalues $\upsilon_1(x)$ and $\upsilon_K(x)$ are strictly positive for all x in the support of F_X, and the eigenfunctions ϕ_k are orthonormal in L^2. Thus the mean function and the scores are allowed to vary with x while the eigenfunctions are independent of x, which provides better interpretability of the dependency of $\psi(f)$ on $X = x$. Performing the KL decomposition in the transformed Hilbert space rather than in density space is well suited due to the former being a linear vector space whereas the latter lacks linearity structure, and therefore a truncated KL expansion applied directly to the density process may not reside in density space [27].

The data generation mechanism for the shape part of the intensity function is as follows: First generate the covariate $X \sim \mathcal{U}(0,1)$. Then a random element $\mathcal{X}(s)$, $s \in (0,1)$, in L^2 is generated from (16) by sampling the Functional Principal Component (FPC) scores ξ_k. The random density function f with support $[0,T]$, where $T = 1$, is obtained by mapping back $\mathcal{X}(\cdot)$ to density space using the inverse LQD transform, i.e. $Q(t) = \theta_{\mathcal{X}}^{-1} \int_0^t \exp(\mathcal{X}(s))ds$, where $\theta_{\mathcal{X}} = \int_0^1 \exp(\mathcal{X}(s))ds$ and f is the density corresponding to the quantile function Q.

For the intensity factor τ, we consider a linear regression setting $E(\tau|X = x) = a_1 + b_1 x$ such that the values on the right hand side are all positive. The conditional intensity factors τ for covariate level X were obtained through a linear regression model $\tau = a_1 + b_1 X + \varepsilon$, where ε is independent of X and has a truncated normal distribution with mean zero, variance σ_1^2 and support $[c_1, d_1]$. The choice of the constants above are such that $a_1 + b_1 x + \varepsilon > 0$ for all x in the support of F_X.

Next, random samples of data (X_i, τ_i, Q_i), $i = 1, \ldots, n$, were generated following the above procedure, where $a_1 = 1$, $b_1 = 0.2$, $c_1 = -0.2$, $d_1 = 0.2$, $\sigma_1 = 1.5$, $K = 2$, $\mu(s,x) = \exp(1.5x) + \exp(1.5s)$, $\phi_1(s) = -\sqrt{2}\cos(\pi s)$, $\phi_2(s) = \sqrt{2}\sin(\pi s)$, $s \in [0,1]$, $\upsilon_1^2(x) = 3 + 2x$ and $\upsilon_2^2(x) = (2 - x)^2$. A triangular array of point processes was then obtained as follows: Conditional on (X_i, τ_i, Q_i), the observable number of arrivals for the i-th point process $N_i^{(n)}(T)$ was sampled from a Poisson distribution with rate $\alpha_n \tau_i$, where $\alpha_n = 40n^{4/5}$. Then, conditional on $N_i^{(n)}(T)$ and Q_i, the arrival times were generated as an i.i.d. sample of size $N_i^{(n)}(T)$ from Q_i. For this step, we utilize inverse sampling method by generating $N_i^{(n)}(T)$ i.i.d. uniform in $(0,1)$ random variables $u_1, \ldots, u_{N_i^{(n)}(T)}$, independent of all other random quantities, and then the arrival times are obtained from the $Q(u_k)$, $k = 1, \ldots, N_i^{(n)}(T)$. We generate Q_i over a dense grid on $(0,1)$ and used the bandwidth sequence $h = h_n = n^{-1/5}$ for the local Fréchet regression step.

Figure 1 shows the "oracle" regression function with intensity factors $E(\tau|X = x)$ and shape (density) function defined through the corresponding quantile function $E(Q(\cdot)|X = x)$, where we consider a grid of 50 equispaced predictor values in $(0,1)$. Here $E(Q(\cdot)|X = x)$ is approximated through a Monte Carlo approach where we average across random quantiles Q_i generated at predictor level x.

We ran 1000 simulations for sample sizes $n = 100, 200$ and 500. For the r^{th} simulation, we measure the performance of the method by comparing against the "oracle" conditional intensity function as defined before. Denoting by $\tilde{f}_{\oplus}^r(x)$ and $\tilde{\tau}_{\oplus}^r(x)$ the empirical estimates for the shape function and intensity factor parts of the conditional intensity function in the r-th simulation, respectively, we measured the quality of the estimation by integrated squared errors similar to [29], using the metric as in (1). Since $\tau_{\oplus}(x)$ can be consistently estimated up to the constant $E(\tau)$, we expect the estimates and $\tau_{\oplus}(x)$ to

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

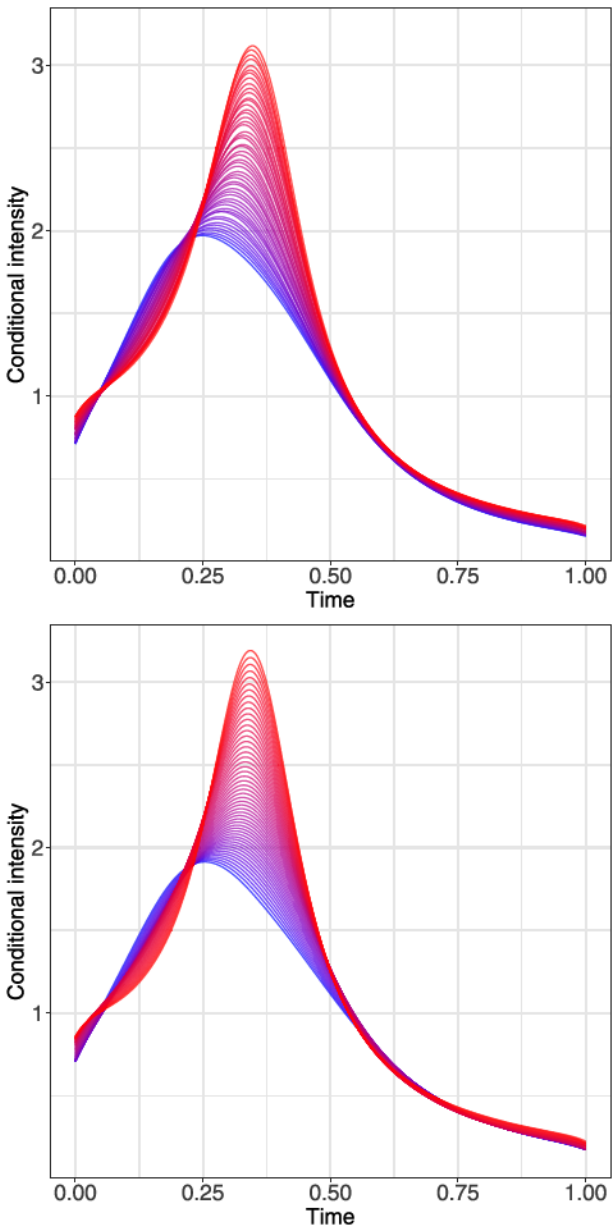

Fig. 1. Conditional intensity functions in the simulation setting over a dense grid of predictor levels x, displayed in blue when $x = 0$ to red when $x = 1$ in the online version (dark grey and light grey in the print version, resp.). The upper panel shows the approximation of $\Lambda_{\oplus}(x)$ through a Monte Carlo approach while the bottom panel shows the estimate $\hat{\Lambda}_{\oplus}(x)$ adjusted by the constant $E(\tau)$ using $n = 2000$.

differ by a positive constant and the estimates $\tilde{\tau}^r_{\oplus}(x)$ of $\tau_{\oplus}(x)$ as in Theorem 2 by $E(\tau)$. This leads to

$$\begin{aligned} ISE_r &= \int_0^1 \Big\{ d^2_{L^2}(\tilde{Q}_{\oplus}(x), E(Q(\cdot)|X = x)) \\ &\qquad + d^2_E(E(\tau)\tilde{\tau}^r_{\oplus}(x), E(\tau|X = x)) \Big\} dx \\ &= ISE^{\mathcal{F}}_r + ISE^{\mathcal{T}}_r . \end{aligned}$$

The previous integrals are obtained numerically over a dense grid of predictor values consisting of 200 equidistant points in $(0, 1)$, where $E(Q(\cdot)|X = x)$ is obtained through

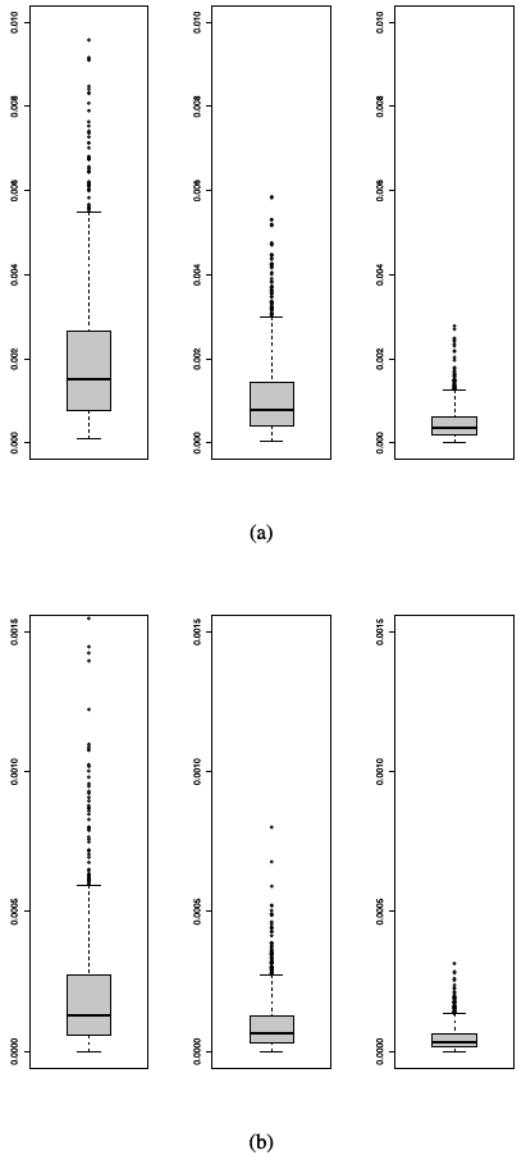

Fig. 2. Boxplots of the errors for the conditional shape function estimates $ISE^{\mathcal{F}}_r$ (upper panels) and the conditional intensity factors $ISE^{\mathcal{T}}_r$ (bottom panels), in the simulation setting for $n = 100$ (left), $n = 200$ (middle) and $n = 500$ (right). Here four and three outliers were removed for the shape and intensity factor boxplots when $n = 100$, respectively.

a Monte Carlo approach for each x in the dense grid as explained before. The boxplots of $ISE^{\mathcal{F}}_r$ and $ISE^{\mathcal{T}}_r$ are presented in Figure 2. As sample size increases, these error estimates are seen to decrease towards 0. This indicates that the estimated conditional intensity functions converge to their true counterparts, up to the constant $E(\tau)$.

C. Simulations for Global Fréchet Regression

In this section we assess the finite sample performance of the global Fréchet regression estimates. The data generation mechanism is as follows. First generate the covariate $X \sim \mathcal{U}(0, 1)$. Then the random density f corresponds to a truncated normal random variable with support $[0, T]$, $T = 1$, mean $\mu(x) = a_2 + b_2 x + \varepsilon_1$ and standard deviation $\sigma(x) = a_3 + b_3 x + \varepsilon_2$, where ε_k, $k = 1, 2$, is independent of all other random

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

quantities and has a truncated normal distribution with mean zero, standard deviation σ_k and support $[e_k, f_k]$ such that $\sigma(x)$ is positive for all x in the support of F_X. Thus both the mean and standard deviation change linearly with x. We choose $a_2 = 0.3$, $b_2 = 0.4$, $a_3 = 0.1$, $b_3 = -0.01$, $e_1 = -0.1$, $f_1 = 0.1$, $e_2 = -0.01$, $f_2 = 0.01$ and $\sigma_1 = \sigma_2 = 0.5$. These settings reflect a situation where the shape components are Gaussian and pushed to the right as x increases while the intensity factor becomes larger. Figure 3 shows the "oracle" global Fréchet regression function over a dense grid of predictor values and the estimated counterpart adjusted by the constant $E(\tau)$. Lemma S.11 in the appendix shows that $Q(f_{G\oplus}(x))$ is the orthogonal projection of $Q_G(\cdot, x) = E(s(X, x)Q(\cdot))$ onto $Q(\Omega_{\mathcal{F}})$, where Q is the quantile function corresponding to the generic random density f. We obtain $Q_G(\cdot, x)$ at each value of x in the grid by employing a Monte Carlo approach similarly as in section V-B by averaging across random trajectories $s(X, x)Q_i$.

We ran 1000 simulations for sample sizes $n = 100, 200$ and 500. For the r^{th} simulation, we measure the performance of the method by comparing against the "oracle" global intensity function as defined before and using integrated squared errors analogous to the ones outlined in section V-B. The boxplots of the error metrics against the "oracle" global intensity function are presented in Figure 4 and are clearly seen to converge to zero as sample size increases.

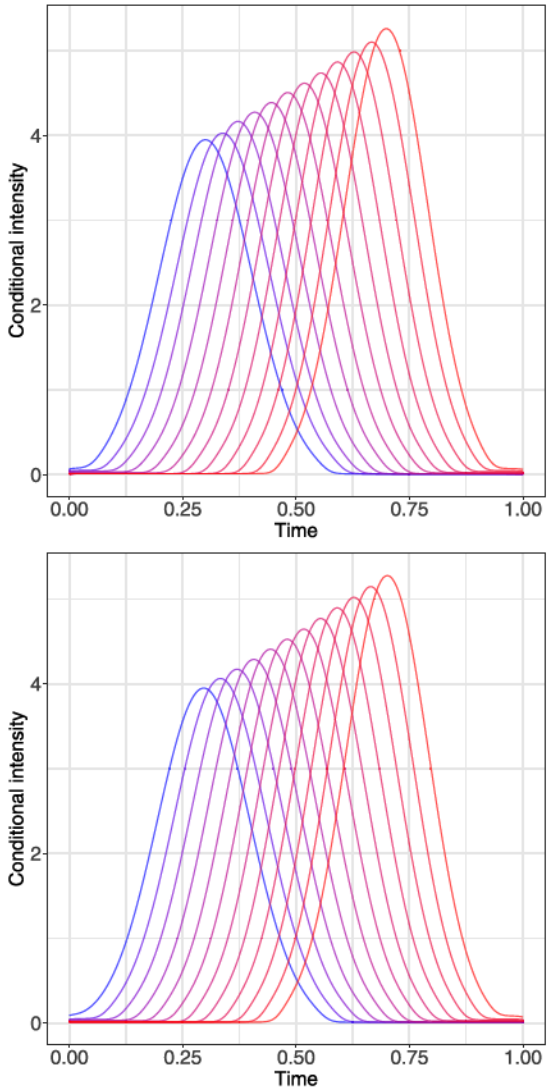

Fig. 3. Conditional global intensity functions $m_{G\oplus}(x)$ in the simulation setting over a dense grid of predictor levels x, displayed in blue when $x = 0$ to red when $x = 1$ in the online version (dark grey and light grey in the print version, resp.). The upper panel shows the approximation of $m_{G\oplus}(x)$ through a Monte Carlo approach while the bottom panel shows the estimate $\hat{\Lambda}_{G\oplus}(x)$ adjusted by the constant $E(\tau)$ using $n = 2000$.

VI. Data Applications and Extensions

A. Chicago's Divvy Bike System

We illustrate our approach for the bike trips records of the Chicago Divvy bike system, which are publicly available at https://www.divvybikes.com/system-data. The bike trip records contain information such as the bike pickup and drop-off location, date and time, between more than 600 bike rental stations in Chicago. In the context of replicated temporal Poisson processes, [3] analyzed this dataset by adapting an additive principal component model to the log-intensity functions of daily pickups and estimating model parameters through a likelihood based approach, and such bike sharing systems have been extensively studied [51]. We considered the point process of daily pickups of bikes in a cluster consisting of 6 stations not far from each other in the Chicago Divvy system during weekdays of 2017, consisting of a station on East South Water street and the five nearest bike rental stations south of the Chicago river.

To study the effect of the temperature on the demand of bikes, we obtained the daily observed temperature in Chicago as recorded at the weather station 'Northerly island' from https://www.ncdc.noaa.gov and fitted a local Fréchet regression model to obtain the conditional intensity functions of the bike rentals, using estimates $\hat{\Lambda}_{\oplus}(x)$ as in (12), where we used a bandwidth of 1.5°C. The results are presented in Figure 5.

A clear difference emerges between days with temperature above 10°C which have a uniformly higher intensity function compared to days with lower temperature. In both cases, the shape of the intensity function appears to be bimodal with peaks at 9am and 5pm, which are likely due to bike rentals for the purpose of commuting to the workplace. Moreover, the conditional intensity function estimate is higher around 5pm compared to the early peak at 9am, which may be explained by the fact that it is warmer and easier to bike in the afternoon than early in the morning, so perhaps commuters use public or shared transport in the morning and a bike in the afternoon.

There appears to be a "shoulder" or minor peak of bike rental demand at around 12pm on warm days only, which is likely related to more or less optional lunch break related bicycle travel, for leisure or to catch some food. Overall, we find that increasing temperature boosts the bike rental demand in this region of Chicago. Figure 5 (b) shows the quotient between the estimated conditional intensity function at $x = 2$°C and $x = 18$°C. We observe that the ratio is much higher around noon compared to the ratios at 9am and 5pm; moreover, this ratio is higher for the morning

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

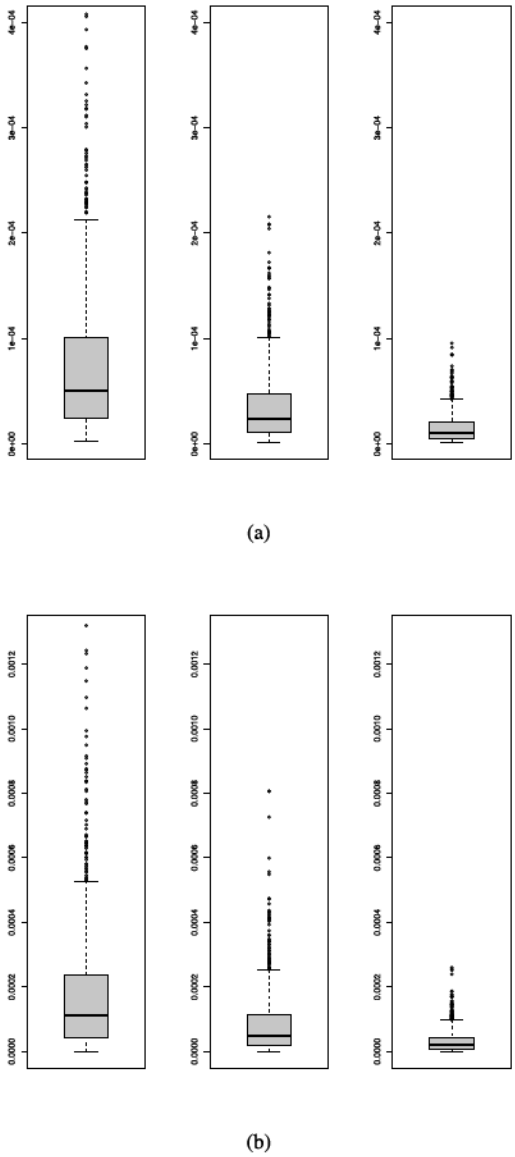

Fig. 4. Boxplots of the errors for the conditional shape function estimates $ISE_r^{\mathcal{F}}$ (upper panels) and the conditional intensity factors $ISE_r^{\mathcal{T}}$ (bottom panels), in the simulation setting for $n = 100$ (left), $n = 200$ (middle) and $n = 500$ (right). Five outliers were removed for both the shape and intensity factor boxplots when $n = 100$, respectively.

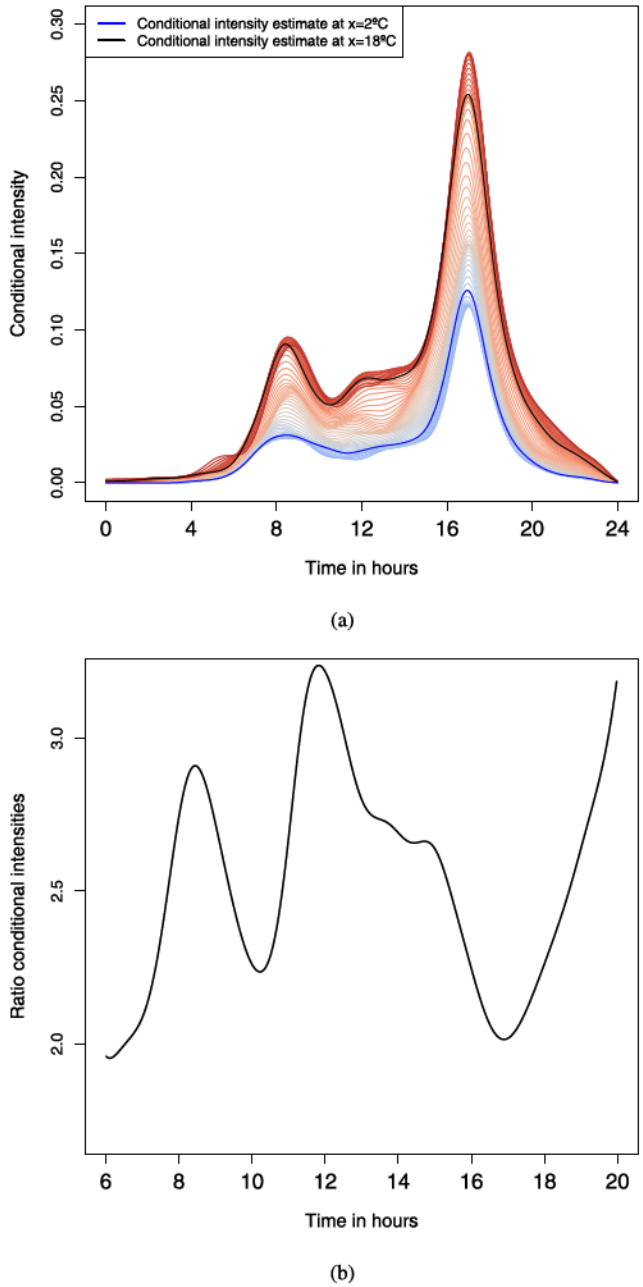

Fig. 5. (a) Estimated conditional intensity functions in dependence on temperature values between $-2°$C (blue) and $24°$C (red) in the online version (dark grey and light grey in the print version, resp.), using bandwidth $h = 1.5°$C, with highlighted conditional intensity functions at temperatures $2°$C and $18°$C. (b) Ratio between the estimated conditional intensity functions at temperature $18°$C to that at $2°$C.

commute compared to the evening commute, indicating that the afternoon demand is not just a reflection of the morning demand. This ratio can be characterized as the degree to which bike travel is optional, where the obvious alternatives to making a trip by bike are not making a trip or making the trip by other means of transportation.

B. New York Yellow Taxi System

The New York yellow taxi trip records is a rich and large scale database that contains information such as the taxi pickup and drop-off latitude and longitude locations as well as the date and time, among several other variables. The data is available from the NYC Taxi and Limousine Commission (TLC) at https://www1.nyc.gov/site/tlc/about/data.page. Using point processes, this data has been studied by several authors in an applied setting; see for example [52] and the references therein for a review and comparison of different intensity models. The Poisson process as a working model for the taxi pickups at a fixed location is well justified theoretically as a superposition of many independent and sufficiently sparse point processes, similar to the case of call arrivals at a telephone exchange [6]. It is of interest to study how the demand of taxis is associated with the day of the week and for this we employ a regression approach.

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

We consider the point process of daily pickups of yellow taxis that occurred at Penn station in Manhattan during 2017. Penn station is a major train station located in Midtown Manhattan that serves commuters from New York City, New Jersey and Long Island, and connects NYC with several other cities. We view these data as a sample of replicated point processes, as each day produces a replication of the underlying data generation mechanism. To study the effect of weekdays and weekends on the demand of taxis at Penn station, we consider a categorical predictor X that indicates whether the day corresponds to a Monday-Thursday, Friday, Saturday or Sunday. Since local smoothing does not apply to indicator type predictors, we instead consider the global regression framework that was introduced in section IV-D and is well suited for categorical predictors.

Fitting a global regression model for the intensity function on the day of week X by using the estimates $\hat{\Lambda}_{G\oplus}(x)$ defined after Theorem 4 leads to the results as presented in Figure 6. For Sundays, the intensity function is highest late in the day and after 4pm is higher than on all other days, likely due to people returning to New York City from an out of town trip. The weekday (Monday through Thursday) intensity function is bimodal with a higher first mode. These modes likely correspond to the commuter traffic, where the 8am mode would be due to commuters who live outside New York City and arrive for work in the City in the morning, while the evening mode likely corresponds to people who return from out of town at Penn Station and hail a taxi there. On Fridays, the same modes are present but with a reversal of their height, as the second mode is now higher than the first mode, likely corresponding to reverse commuters who live in New York City and return from outside, perhaps having a work place away from the City.

The patterns for Saturday are also bimodal but with different locations and levels of the modes, indicating that relative large numbers of people arrive at Penn Station around noon and at 8pm, perhaps indicating leisure and shopping trips.

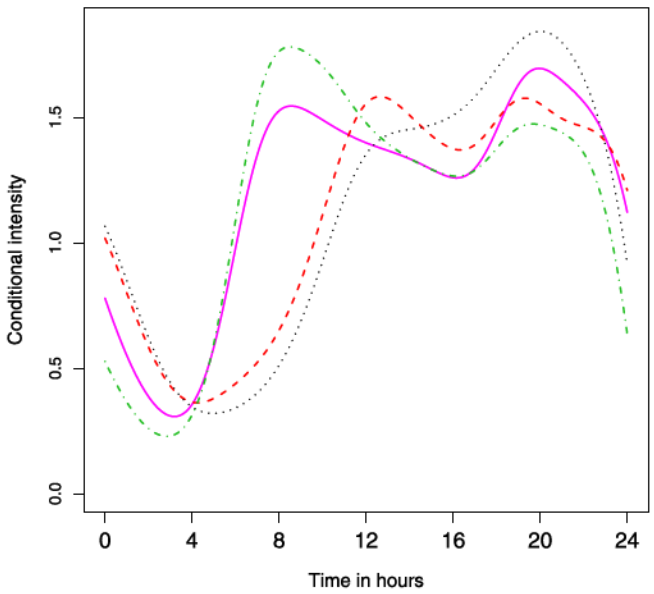

Fig. 6. Estimated intensity functions using global regression with type of day as predictors. The fitted response functions are shown for Monday-Thursday (green dot-dashed), Friday (magenta solid), Saturday (red dashed) and Sunday (black dotted) in the online version. The Y-axis is scaled by a factor of 10^{-5}.

TABLE I
DISTRIBUTION OF EARTHQUAKES IN CHILE BETWEEN 1980 AND 2017

Magnitude	Number of Earthquakes
$[5.0, 5.5)$	1,679
$[5.5, 6.0)$	476
$[6, 6.5)$	160
$[6.5, 7.0)$	53
$[7.0, 9.0)$	25

C. Aftershock Earthquake Process in Chile

The proposed Cox process regression also has applications in seismology, where the times earthquakes strike naturally forms a point processes. So it is not surprising that earthquake activity has met with major interest in the point process literature [2]. Chile is widely known for its strong seismic activity in both frequency and intensity. Our goal is to study the aftershock process that follows a major earthquake occurring at time $t = 0$, where the major earthquake that may trigger aftershocks is referred to as the mainshock. We focus on the arrival of aftershocks that occur in a time window of 2 months after the mainshock so that the aftershock process $N(t)$ is observed in $[0, T]$ for $T = 2$ months.

To demonstrate the proposed regression methods, we considered the mainshocks that occurred between 1980 and 2017 in Chile. These include some strong earthquakes such as the magnitude 8.8 earthquake on the moment magnitude scale in February 2010, 8.2 in April 2014, 8.0 in March 1985, and other strong earthquakes. The data was obtained from the U.S. Geological Survey web page https://earthquake.usgs.gov/earthquakes/search/ which provides information concerning the location, magnitude and date of the earthquakes. We classified each earthquake in terms of its magnitude category and selected 15 mainshocks at random from each category in the order strongest to weakest. This enables us to consider the strongest earthquakes above magnitude 7; see Table I. Table II shows some of the strongest aftershocks along with their arrival time after the 2010 earthquake.

In order to avoid including an aftershock that could correspond to more than one mainshock, we adopted the following selection scheme: If we select an earthquake that occurs at calendar time t_0 as mainshock, then we cannot choose any earthquake that occurs in the interval $(t_0 - T, t_0 + T)$ as mainshock. Furthermore, we consider an earthquake to be a mainshock if the sequence of earthquakes that occur during the following 2 months after its arrival time have strictly smaller magnitudes.

We implemented conditional intensity function estimation using the magnitude of the mainshock as a one-dimensional predictor. Figure 7 shows the estimated conditional intensity functions (12) for different levels of magnitudes between 5 and 8.5, along with the conditional intensity for the mean magnitude level $x = 6.3$ and for a strong earthquake at magnitude

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

TABLE II

SOME OF THE STRONGEST AFTERSHOCKS WITHIN 4 HOURS FROM THE MAGNITUDE 8.8 EARTHQUAKE ON 27 FEBRUARY, 2010, CHILE

Magnitude	6.2	6.0	6.0	7.4	6.1	6.0	6.0
Aftershock Time [hours]	0.3	0.6	1.1	1.5	1.9	3.9	3.9

$x = 8.0$. We observe that the conditional intensity function for the strong earthquake has an exponential decay, suggesting that most aftershocks occur closer to the mainshock. The shape of the regression curve agrees with the traditional model for aftershock sequences that correspond to large earthquakes, with intensity function declining as a power law, which is known as the modified Omori law [2], [53].

The regression curve for a magnitude 6.3 earthquake is mostly flat but presents a slow decay towards the end of the time window of observation. This could be due to the fact that medium or low magnitude earthquakes do not tend to produce substantially more aftershocks compared to the natural seismic background activity in Chile. In fact, the conditional intensity function for the strong earthquake is uniformly higher than the one for the magnitude 6.3 mainshock. Finally, the conditional intensity functions for earthquakes below magnitude 6 are mostly flat, which indicates that those earthquakes tend to produce aftershocks that are more uniformly scattered and at least partly correspond to the background seismic activity in Chile.

VII. CONCLUSION

We develop here a novel fully non-parametric regression method that features point processes as responses coupled with Euclidean predictors $X \in \mathbb{R}^p$ by establishing a connection to conditional barycenters. Crucially, our model is based on the availability of repeated realizations of the same point process. The random objects for which we construct conditional barycenters are the intensity functions of Cox processes, which we can represent as taking values in a product metric space. A novelty in point processes is that the regression setting makes it possible to achieve consistent estimation of intensity functions (up to a constant scale factor for the intensity that is common to all observed realizations of the point process).

Obtaining such consistency has been an elusive goal and in fact is not possible when one has one realization of the point process over a fixed domain. What we show here is that this lack of consistency can be overcome in a regression setting where one can harness concepts of conditional barycenters that have been developed for Fréchet regression. For each point process, one may have a continuous one-dimensional or general vector predictor that is a random variable associated with the point process. In the former case we can use a nonparametric smoothing method under minimal assumptions, while in the latter case we target a global model that is akin to multiple linear regression and makes it possible to include indicators as predictors.

Our approach relies on straightforward computations and does not require the use of functional principal components, a tool that is not well suited for intensity functions as they do not reside in a linear space due to their non-negative nature. We show that the proposed regression model is applicable to many data situations where one is interested to study the behavior of point processes in dependence on covariates, including applications in transportation and seismology.

APPENDIX

A. Proofs of Results in Section IV-A

Proof of Lemma 1: Let $Q_1, Q_2 \in Q(\Omega_{\mathcal{F}})$, $\lambda \in [0,1]$ and $Q_\lambda = \lambda Q_1 + (1-\lambda)Q_2$ and $x, y \in [0,1]$. It is clear that $Q_\lambda(0) = 0$, $Q_\lambda(1) = T$, and by the triangle inequality $|Q_\lambda(x) - Q_\lambda(y)| \leq L|x-y|$. Suppose now that $x \leq y$, then since Q_λ is non-decreasing we have $|Q_\lambda(x) - Q_\lambda(y)| = Q_\lambda(y) - Q_\lambda(x)$. Furthermore, for M as in (S1),

$$\begin{aligned} &Q_\lambda(y) - Q_\lambda(x) \\ &= \lambda(Q_1(y) - Q_1(x)) + (1-\lambda)(Q_2(y) - Q_2(x)) \\ &\geq \lambda M(y-x) + (1-\lambda)M(y-x) \\ &= M(y-x), \end{aligned}$$

which implies $|Q_\lambda(x) - Q_\lambda(y)| \geq M|x-y|$ for all $x, y \in [0,1]$. Hence, $Q(\Omega_{\mathcal{F}})$ is convex and it is clearly a subset of $L^2([0,1])$ since the quantile functions are bounded. Next, let $Q_1, Q_2, \ldots$ be a sequence in $Q(\Omega_{\mathcal{F}})$ such that $Q_n \xrightarrow{L^2} Q \in L^2([0,1])$ as $n \to \infty$. We show that $Q \in Q(\Omega_{\mathcal{F}})$. In fact, since the family $\{Q_n\}_{n=1}^{\infty}$ has a common Lipschitz constant L, it follows that it is uniformly equicontinuous. Moreover, since $[0,1]$ is compact and $Q_n \xrightarrow{L^2} Q \in L^2([0,1])$ then $Q_n \to Q$ uniformly as $n \to \infty$. This implies $Q(0) = 0$ and $Q(1) = T$. Next, for any $x, y \in [0,1]$ and $\epsilon > 0$ we have that $||Q_n - Q||_\infty \leq \epsilon/2$ for n large enough and

$$\begin{aligned} &|Q(x) - Q(y)| \\ &\leq |Q(x) - Q_n(x)| + |Q_n(x) - Q_n(y)| + |Q_n(y) - Q(y)| \\ &\leq 2||Q_n - Q||_\infty + L|x-y| \\ &\leq \epsilon + L|x-y|, \end{aligned}$$

for large enough $n(\epsilon)$, using that $Q_n \in Q(\Omega_{\mathcal{F}})$ by assumption and the functions in this space satisfy the Lipschitz condition with constant L. Taking $\epsilon \downarrow 0$ we obtain that Q is also Lipschitz with constant L. Similarly, for n large enough

$$\begin{aligned} M|x-y| &\leq |Q_n(x) - Q_n(y)| \\ &\leq 2||Q_n - Q||_\infty + |Q(x) - Q(y)| \\ &\leq \epsilon + |Q(x) - Q(y)|, \end{aligned}$$

so that Q satisfies condition $(S1)$. Therefore, $Q(\Omega_{\mathcal{F}})$ is closed.

The following lemma shows that there are no empty point processes and follows by adopting analogous arguments as the ones outlined in the proof of Lemma 3 in [30]. We present it here only for completeness and without proof. In what follows, we introduce auxiliary quantities $s_i(x,h) = \sigma_0^{-2} K_h(X_i - x)[u_2 - u_1(X_i - x)]$, $i = 1, \ldots, n$, where $u_j = E(K_h(X_1 - x)(X_1 - x)^j)$, $j \in \{0,1,2\}$ and $\sigma_0^2 = u_0 u_2 - u_1^2$.

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

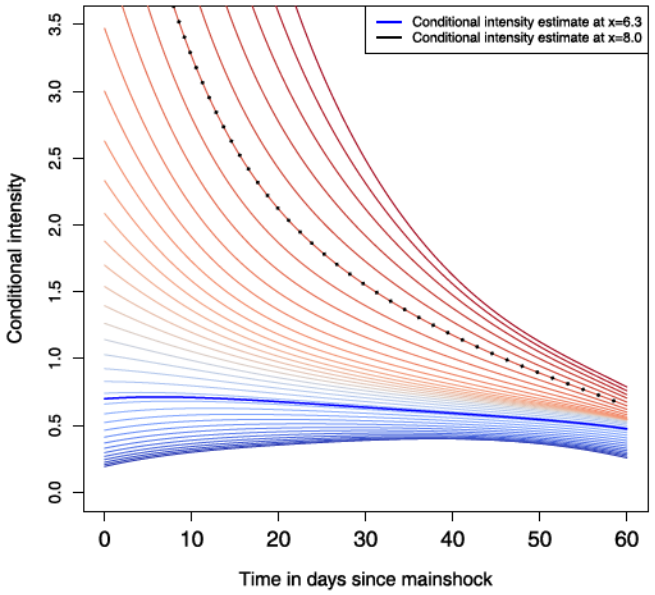

Fig. 7. Estimated conditional intensity functions for the aftershock process for different levels of magnitude between 5.0 (blue) and 8.5 (red) in the online version, with specific conditional intensity functions at magnitudes 6.3 and 8.0, and bandwidth 0.8 in magnitude. The Y-axis is scaled by a factor of 10^{-3}.

Lemma S.1: Suppose that $\tau \geq \kappa > 0$ almost surely with κ as in assumption $(S3)$ and $\alpha_n / \log n \to \infty$ as $n \to \infty$. Then

$$\liminf_{n\to\infty} \frac{\min_{1\leq i\leq n} N_i^{(n)}(T)}{\alpha_n} \geq \frac{(1-e^{-1})\kappa}{2} \quad \text{a.s..}$$

Proof of Proposition 1: From (10) we have

$$d_{\mathcal{F}}(\hat{f}_{\oplus}(x), \tilde{f}_{\oplus}(x)) \leq n^{-1} \sum_{i=1}^{n} |s_{in}(x,h)| \; ||\hat{Q}_i - Q_i||_{L^2([0,1])}. \tag{17}$$

From the proof of Lemma 2 in [29], we have $s_{in}(x,h) - s_i(x,h) = W_{0n} K_h(X_i - x) + W_{1n} K_h(X_i - x)(X_i - x)$, where $W_{0n} = \hat{u}_2/\hat{\sigma}_0^2 - u_2/\sigma_0^2 = O_p((nh)^{-1/2})$ and $W_{1n} = \hat{u}_1/\hat{\sigma}_0^2 - u_1/\sigma_0^2 = O_p((nh^3)^{-1/2})$. Thus

$$\begin{aligned}
&n^{-1} \sum_{i=1}^{n} |s_{in}(x,h)| \; ||\hat{Q}_i - Q_i||_{L^2([0,1])} \\
&\leq n^{-1} \sum_{i=1}^{n} |s_{in}(x,h) - s_i(x,h)| \; ||\hat{Q}_i - Q_i||_{L^2([0,1])} \\
&+ n^{-1} \sum_{i=1}^{n} |s_i(x,h)| \; ||\hat{Q}_i - Q_i||_{L^2([0,1])} \\
&\leq W_{0n} \; n^{-1} \sum_{i=1}^{n} K_h(X_i - x) ||\hat{Q}_i - Q_i||_{L^2([0,1])} \\
&+ W_{1n} \; n^{-1} \sum_{i=1}^{n} K_h(X_i - x) |X_i - x| \; ||\hat{Q}_i - Q_i||_{L^2([0,1])} \\
&+ |u_2/\sigma_0^2| \; n^{-1} \sum_{i=1}^{n} K_h(X_i - x) ||\hat{Q}_i - Q_i||_{L^2([0,1])} \\
&+ |u_1/\sigma_0^2| \; n^{-1} \sum_{i=1}^{n} K_h(X_i - x) |X_i - x| \; ||\hat{Q}_i - Q_i||_{L^2([0,1])}.
\end{aligned} \tag{18}$$

Let μ_i be the probability measure on $[0,T]$ with corresponding quantile function Q_i. Note that

$$\begin{aligned}
&||\hat{Q}_i - Q_i||_{L^2([0,1])} \\
&= d_W(\hat{\mu}_i, \mu_i) 1_{(N_i^{(n)}(T)>0)} + d_W(\hat{\mu}_i, \mu_i) 1_{(N_i^{(n)}(T)=0)} \\
&\leq d_W(\hat{\mu}_i, \mu_i) 1_{(N_i^{(n)}(T)>0)} + T 1_{(N_i^{(n)}(T)=0)} \\
&\leq d_W(\hat{\mu}_i, \mu_i) 1_{(N_i^{(n)}(T)>0)},
\end{aligned}$$

almost surely and for large enough n, where the last inequality is due to Lemma S.1 above using the condition $\alpha_n / \log n \to \infty$ as $n \to \infty$. This along with the fact that

$$\begin{aligned}
&E(d_W(\hat{\mu}_i, \mu_i) 1_{(N_i^{(n)}(T)>0)} | \Lambda_i, X_i) \\
&= E(d_W(\hat{\mu}_i, \mu_i) 1_{(N_i^{(n)}(T)>0)} | \Lambda_i) \\
&\leq (T/(2\kappa))^{1/4} \alpha_n^{-1/4},
\end{aligned}$$

for large enough n, which follows by similar arguments as the ones outlined in the proof of Theorem 2 in [30], shows that

$$E(||\hat{Q}_i - Q_i||_{L^2([0,1])} | \Lambda_i, X_i) = O(\alpha_n^{-1/4}),$$

for large enough n, where the $O(\alpha_n^{-1/4})$ term is uniform in i. Thus, for $j \in \{0,1\}$ and by a conditioning argument, we have

$$\begin{aligned}
&E\left(n^{-1} \sum_{i=1}^{n} K_h(X_i - x) |X_i - x|^j \; ||\hat{Q}_i - Q_i||_{L^2([0,1])} \right) \\
&= n^{-1} \sum_{i=1}^{n} E\Big(K_h(X_i - x) |X_i - x|^j \\
&\qquad E(||\hat{Q}_i - Q_i||_{L^2([0,1])} | \Lambda_i, X_i) \Big) \\
&= O(\alpha_n^{-1/4}) \; n^{-1} \sum_{i=1}^{n} E\left(K_h(X_i - x) |X_i - x|^j \right) \\
&= O(h^j \alpha_n^{-1/4}),
\end{aligned}$$

which implies

$$\begin{aligned}
&n^{-1} \sum_{i=1}^{n} K_h(X_i - x) |X_i - x|^j \; ||\hat{Q}_i - Q_i||_{L^2([0,1])} \\
&= O_p(h^j \alpha_n^{-1/4}).
\end{aligned} \tag{19}$$

Define auxiliary quantities $\gamma_k = \int_{-1}^{1} u^k K(u) du$, $k = 0, 1, 2$. By Taylor expansion it is easy to see that $u_0 = f(x) + h \int_{-1}^{1} u K(u) f'(\xi_{0u}) du$, $u_1 = h^2 (\int_{-1}^{1} u^2 K(u) f'(\xi_{1u}) du)$, $u_2 = h^2 \gamma_2 f(x) + h^3 \int_{-1}^{1} u^3 K(u) f'(\xi_{2u}) du$, where ξ_{1u}, ξ_{2u} and ξ_{3u} are between x and $x + uh$. Since $a_k := \int_{-1}^{1} u^{k+1} K(u) f'(\xi_{ku}) du$, $k = 0, 1, 2$, satisfies $|a_k| \leq \sup_{s\in[x-1,x+1]} |f'(s)| < \infty$ for large enough n, it follows that

$$\begin{aligned}
&|u_2/\sigma_0^2| \\
&= (\gamma_2 f(x) + h a_2)/(\gamma_2 f^2(x) + h(a_2 f(x) + a_0 \gamma_2 f(x)) \\
&\qquad + h^2(a_0 a_2 - a_1^2)) = O(1), \\
&|u_1/\sigma_0^2| \\
&= a_1/(\gamma_2 f^2(x) + h(a_2 f(x) + a_0 \gamma_2 f(x)) + h^2(a_0 a_2 - a_1^2)) \\
&= O(1),
\end{aligned}$$

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

as $n \to \infty$. Combining with (18), (19) and the fact that $W_{0n} = O_p((nh)^{-1/2})$ and $W_{1n} = O_p((nh^3)^{-1/2})$ leads to

$$n^{-1}\sum_{i=1}^{n}|s_{in}(x,h)|\ ||\hat{Q}_i - Q_i||_{L^2([0,1])} = O_p(\alpha_n^{-1/4}),$$

and the result then follows from (17).

Proof of Theorem 1: Under the assumptions $(K0)$, $(L1)$ and $f_X > 0$ with unbounded support and the choice of h_n, we have from Corollary 1 in [29] that

$$d_{\mathcal{F}}(f_{\oplus}(x), \hat{f}_{\oplus}(x)) = O_p(n^{-2/5}).$$

Next, from Proposition 1 we obtain $d_{\mathcal{F}}(\hat{f}_{\oplus}(x), \tilde{f}_{\oplus}(x)) = O_p(\alpha_n^{-1/4})$. Combining with the triangle inequality (9) leads to

$$\begin{aligned} d_{\mathcal{F}}(f_{\oplus}(x), \tilde{f}_{\oplus}(x)) &\le d_{\mathcal{F}}(f_{\oplus}(x), \hat{f}_{\oplus}(x)) + d_{\mathcal{F}}(\hat{f}_{\oplus}(x), \tilde{f}_{\oplus}(x)) \\ &= O_p(n^{-2/5} + \alpha_n^{-1/4}), \end{aligned}$$

which shows the result.

B. Proofs of Results in Section IV-B

We begin by establishing four auxiliary lemmas.

Lemma S.2: Suppose that there exists $M_1 < \infty$ such that $\tau \le M_1$ almost surely and let $i \in \{1, \ldots, n\}$. Then, conditionally on a realization of Λ_i, $N_i^{(n)}(T)/\alpha_n - \tau_i$ is mean zero with (conditional) variance bounded above by M_1/α_n.

Proof of Lemma S.2: Since $N_i^{(n)}$ is a Cox process, it is generated by two independent random mechanisms: First the generation of the intensity function Λ_i, and then, conditional on Λ_i, of the realizations of $N_i^{(n)}$ corresponding to those of a Poisson process $N_i^{(n)}(\cdot|\Lambda_i)$ with intensity function $\alpha_n\Lambda_i$ [2]. Thus, we may regard the probability space associated with the generation of the intensity function and the Poisson process as a product probability space $\mathcal{W}_1 \times \mathcal{W}_2$ such that $\Lambda_i = \Lambda_i(w_1)$, where $w_1 \in \mathcal{W}_1$ leads to a realization of the intensity function and then $N_i^{(n)}(\cdot) = N_i^{(n)}(\cdot, w_2)$, where $w_2 \in \mathcal{W}_2$ leads to a realization of a Poisson process with intensity function $\alpha_n\Lambda_i$. For $\tau_i = \int_0^T \Lambda_i(s)ds$, we have that given $\Lambda_i(w_1)$, $N_i^{(n)}(T) \sim \mathcal{P}(\alpha_n\tau_i(w_1))$, with (conditional) variance $\alpha_n\tau_i(w_1)$. Thus $E_{\mathcal{W}_2}((N_i^{(n)}(T)/\alpha_n - \tau_i(w_1))|\Lambda_i(w_1)) = 0$ and

$$\begin{aligned} &E_{\mathcal{W}_2}((N_i^{(n)}(T)/\alpha_n - \tau_i(w_1))^2|\Lambda_i(w_1)) \\ &= \alpha_n^{-2}\mathrm{Var}_{\mathcal{W}_2}(N_i^{(n)}(T)\ |\Lambda_i(w_1)) \\ &= \tau_i(w_1)/\alpha_n \\ &\le M_1/\alpha_n, \end{aligned}$$

which shows that

$$\mathrm{Var}_{\mathcal{W}_2}(N_i^{(n)}(T)/\alpha_n - \tau_i(w_1)|\Lambda_i(w_1)) \le M_1/\alpha_n.$$

The result follows.

Lemma S.3: Suppose that $(S3)$ holds. If there exists $M_1 < \infty$ such that $\kappa \le \tau \le M_1$ almost surely with κ as in assumption $(S3)$, then

$$\sqrt{\alpha_n}\ n^{-1}\sum_{i=1}^{n} s_{in}(x,h)\left(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\right) = O_p((nh)^{-1/2}).$$

Proof of Lemma S.3: From the proof of Lemma 2 in [29], we have $s_{in}(x,h) - s_i(x,h) = W_{0n}K_h(X_i - x) + W_{1n}K_h(X_i - x)(X_i - x)$, where $W_{0n} = \hat{u}_2/\hat{\sigma}_0^2 - u_2/\sigma_0^2 = O_p((nh)^{-1/2})$ and $W_{1n} = \hat{u}_1/\hat{\sigma}_0^2 - u_1/\sigma_0^2 = O_p((nh^3)^{-1/2})$. Thus

$$\begin{aligned} &\left|\sqrt{\alpha_n}\ n^{-1}\sum_{i=1}^{n} s_{in}(x,h)\Big(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\Big)\right| \\ &\le \left|\sqrt{\alpha_n}\ n^{-1}\sum_{i=1}^{n}(s_{in}(x,h) - s_i(x,h))\ \Big(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\Big)\right| \\ &+ \left|\sqrt{\alpha_n}\ n^{-1}\sum_{i=1}^{n} s_i(x,h)\ \Big(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\Big)\right| \\ &\le \left|\sqrt{\alpha_n}W_{0n}\ n^{-1}\sum_{i=1}^{n} K_h(X_i - x)\ \Big(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\Big)\right| \\ &+ \left|\sqrt{\alpha_n}W_{1n}n^{-1}\sum_{i=1}^{n} K_h(X_i - x)(X_i - x)\Big(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\Big)\right| \\ &+ \left|\sqrt{\alpha_n}\ n^{-1}\sum_{i=1}^{n} s_i(x,h)\ \Big(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\Big)\right|. \end{aligned} \tag{20}$$

Defining $Z_{jn} := n^{-1}\sum_{i=1}^{n} K_h(X_i - x)(X_i - x)^j[N_i^{(n)}(T)/\alpha_n - \tau_i]$, $j = 0, 1$, then by independence and a conditioning argument we obtain

$$\begin{aligned} &E(Z_{jn}^2) \\ &= E\Big(n^{-2}\sum_{i=1}^{n} K_h^2(X_i - x)(X_i - x)^{2j}\Big(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\Big)^2\Big) \\ &+ n^{-2}\sum_{i=1}^{n}\sum_{k\neq i} E\Big[K_h(X_i - x)(X_i - x)^j\Big(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\Big)\Big] \\ &\qquad E\Big[K_h(X_k - x)(X_k - x)^j\Big(\frac{N_k^{(n)}(T)}{\alpha_n} - \tau_k\Big)\Big] \\ &= E\Big(n^{-2}\sum_{i=1}^{n} K_h^2(X_i - x)(X_i - x)^{2j} \\ &\qquad E\left((N_i^{(n)}(T)/\alpha_n - \tau_i)^2|\Lambda_i, X_i\right)\Big) \\ &\le (M_1/\alpha_n)E\Big(n^{-2}\sum_{i=1}^{n} K_h^2(X_i - x)(X_i - x)^{2j}\Big) \\ &= \Big(M_1h^{2j-1}/(n\alpha_n)\Big)\int_{-1}^{1} u^{2j}K^2(u)f(x+uh)du \\ &= O(h^{2j-1}/(n\alpha_n)), \end{aligned}$$

where the second equality and third inequality are due to Lemma S.2 above and using that $E((N_i^{(n)}(T)/\alpha_n - \tau_i)|\Lambda_i, X_i) = E((N_i^{(n)}(T)/\alpha_n - \tau_i)|\Lambda_i) = 0$. This implies $Z_{0n} = O_p((nh\alpha_n)^{-1/2})$ and $Z_{1n} = O_p((n\alpha_n/h)^{-1/2})$. Thus

$$\begin{aligned} &\sqrt{\alpha_n}W_{0n}\ n^{-1}\sum_{i=1}^{n} K_h(X_i - x)\ \Big(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\Big) \\ &= O_p((nh)^{-1}), \end{aligned} \tag{21}$$

$$\begin{aligned} &\sqrt{\alpha_n}W_{1n}\ n^{-1}\sum_{i=1}^{n} K_h(X_i - x)(X_i - x)\ \Big(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\Big) \\ &= O_p((nh)^{-1}). \end{aligned} \tag{22}$$

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

Similarly, due to $s_i(x,h) = \sigma_0^{-2} K_h(X_i - x)[u_2 - u_1(X_i - x)]$,

$$\begin{aligned}\sqrt{\alpha_n}\, n^{-1} \sum_{i=1}^{n} s_i(x,h) \left(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\right) \\ = (u_2/\sigma_0^2)\sqrt{\alpha_n} Z_{0n} - (u_1/\sigma_0^2)\sqrt{\alpha_n} Z_{1n} \\ = O_p((nh)^{-1/2}),\end{aligned}$$

where the last equality is due to $u_2/\sigma_0^2 = O(1)$ and $u_1/\sigma_0^2 = O(1)$, which were shown in the proof of Proposition 1. This along with (20), (21) and (22) leads to the result.

For the following, recall that $\bar{N}(T) := n^{-1}\sum_{i=1}^{n} N_i^{(n)}(T)$.

Lemma S.4: Suppose that the same assumptions as in Lemma S.3 hold. If $\alpha_n \to \infty$ as $n \to \infty$, then

$$\sqrt{n}\left(\frac{\bar{N}(T)}{\alpha_n} - E(\tau)\right) = O_p(1). \tag{23}$$

Proof of Lemma S.4: The result follows from an application of a central limit theorem for triangular arrays. First consider the case when $\mathrm{Var}(\tau) > 0$. Let $a_i = N_i^{(n)}(T)/\alpha_n$, then by conditioning on Λ_i it follows that $E(a_i) = E(\tau)$ and

$$\sigma_i^2 := \mathrm{Var}(a_i) = \mathrm{Var}(\tau) + E(\tau)/\alpha_n.$$

Setting $s_n^2 = \sum_{i=1}^{n} \sigma_i^2 = n(\mathrm{Var}(\tau) + E(\tau)/\alpha_n)$, we show that

$$\frac{1}{s_n}\sum_{i=1}^{n}(a_i - E(\tau)) \xrightarrow{\mathcal{L}} N(0,1), \tag{24}$$

whence we may infer

$$\sqrt{n}\left(\frac{\bar{N}(T)}{\alpha_n} - E(\tau)\right) \Big/ \sqrt{\mathrm{Var}(\tau) + E(\tau)/\alpha_n} = O_p(1),$$

and furthermore (23), since $\mathrm{Var}(\tau) + E(\tau)/\alpha_n$ is bounded above as τ is uniformly bounded and the positive sequence α_n satisfies $1/\alpha_n \to 0$ as $n \to \infty$. The Lyapunov condition

$$\lim_{n\to\infty} \frac{1}{s_n^4} \sum_{i=1}^{n} E(N_i^{(n)}(T)/\alpha_n - E(\tau))^4 = 0 \tag{25}$$

implies (24) and will hold if we show $\lim_{n\to\infty} \frac{1}{n\alpha_n^4} E(N_1^{(n)}(T) - \alpha_n E(\tau))^4 = 0$. Noting that $N_1^{(n)}(T)|\Lambda_1 \sim \mathcal{P}(\alpha_n \tau_1)$ and

$$\begin{aligned}&(N_1^{(n)}(T) - \alpha_n E(\tau))^4 \\ &= [N_1^{(n)}(T)]^4 - 4[N_1^{(n)}(T)]^3 \alpha_n E(\tau) \\ &\quad + 6[N_1^{(n)}(T)]^2 \alpha_n^2 E(\tau)^2 - 4[N_1^{(n)}(T)]\alpha_n^3 E(\tau)^3 \\ &\quad + \alpha_n^4 E(\tau)^4,\end{aligned} \tag{26}$$

conditional on Λ_1, the higher order moments of $N_1^{(n)}(T)|\Lambda_1$ are given by $E([N_1^{(n)}(T)]^4|\Lambda_1) = (\alpha_n\tau_1)^4 + 6(\alpha_n\tau_1)^3 + 7(\alpha_n\tau_1)^2 + (\alpha_n\tau_1)$, $E([N_1^{(n)}(T)]^3|\Lambda_1) = (\alpha_n\tau_1)^3 + 3(\alpha_n\tau_1)^2 + (\alpha_n\tau_1)$ and $E([N_1^{(n)}(T)]^2|\Lambda_1) = (\alpha_n\tau_1)^2 + (\alpha_n\tau_1)$. Thus, by taking expectation of the conditional moments and using the fact that τ_1 is uniformly bounded along with equation (26) leads to $\lim_{n\to\infty} \frac{1}{n\alpha_n^4} E(N_1^{(n)}(T) - \alpha_n E(\tau))^4 = 0$, completing the proof for the case when $\mathrm{Var}(\tau) > 0$. Next, if $\mathrm{Var}(\tau) = 0$, then $\tau = \tau_0$ almost surely, for some $\tau_0 \in [\kappa, M_1]$. By a conditioning argument we obtain

$$\begin{aligned}\mathrm{Var}(\bar{N}(T)) &= n^{-1}\,\mathrm{Var}(N_1^{(n)}(T)) \\ &= n^{-1}\{E[\,\mathrm{Var}(N_1^{(n)}(T)|\Lambda_1)] \\ &\qquad + \mathrm{Var}(E[N_1^{(n)}(T)|\Lambda_1])\} \\ &= n^{-1}\alpha_n\tau_0.\end{aligned}$$

Letting $\upsilon_n = \sqrt{n}[\bar{N}(T)/\alpha_n - E(\tau)]$, due to $E(\bar{N}(T)/\alpha_n) = E(\tau) = \tau_0$, it follows that $E(\upsilon_n^2) = \tau_0/\alpha_n$. This implies $\upsilon_n = O_p(1)$, which shows the result.

Lemma S.5: Suppose that the same assumptions as in Lemma S.3 hold. If $\psi(\alpha_n) = O(n^{1/2})$ for some function $\psi : \mathbb{R}^+ \to \mathbb{R}$ such that $\psi(\alpha_n) \to \infty$ as $n \to \infty$ and $\frac{\alpha_n}{\log n} \to \infty$ as $n \to \infty$, then

$$\begin{aligned}&\psi(\alpha_n)(\bar{N}(T)/\alpha_n)^{-1} E(\tau|X = x) \\ &= \psi(\alpha_n)E(\tau|X = x)/E(\tau) + O_p(1).\end{aligned} \tag{27}$$

Proof of Lemma S.5: From Lemma S.4 we have $\bar{N}(T)/\alpha_n = E(\tau) + O_p(n^{-1/2})$, and a Taylor expansion leads to

$$\begin{aligned}(\bar{N}(T)/\alpha_n)^{-1} = \frac{1}{E(\tau)} - \frac{1}{E(\tau)^2}\left(\frac{\bar{N}(T)}{\alpha_n} - E(\tau)\right) \\ + o_p(n^{-1/2}).\end{aligned}$$

With $m(x) = E(\tau|X = x)$ one obtains

$$\begin{aligned}&\psi(\alpha_n)(\bar{N}(T)/\alpha_n)^{-1} m(x) \\ &= \frac{\psi(\alpha_n)}{E(\tau)} m(x) - \frac{m(x)}{E(\tau)^2}\frac{\psi(\alpha_n)}{\sqrt{n}}\sqrt{n}\left(\frac{\bar{N}(T)}{\alpha_n} - E(\tau)\right) \\ &\quad + o_p\left(\psi(\alpha_n)/\sqrt{n}\right).\end{aligned}$$

The results follows since $\psi(\alpha_n) = O(n^{1/2})$ implies $\psi(\alpha_n)/n^{1/2} = O(1)$ as $n \to \infty$ and by using Lemma S.4.

Proof of Theorem 2: Observe

$$\begin{aligned}&\psi(\alpha_n)\,\frac{n^{-1}\sum_{i=1}^{n} s_{in}(x,h) N_i^{(n)}(T)}{\bar{N}(T)} \\ &= \psi(\alpha_n)(\bar{N}(T)/\alpha_n)^{-1} \\ &\qquad n^{-1}\sum_{i=1}^{n} s_{in}(x,h)(N_i^{(n)}(T)/\alpha_n - \tau_i) \\ &\quad + \psi(\alpha_n)(\bar{N}(T)/\alpha_n)^{-1} n^{-1}\sum_{i=1}^{n} s_{in}(x,h)\tau_i \\ &= (\psi(\alpha_n)/\sqrt{\alpha_n})(\bar{N}(T)/\alpha_n)^{-1} O_p((nh)^{-1/2}) \\ &\quad + (\psi(\alpha_n)/\sqrt{nh_n})(\bar{N}(T)/\alpha_n)^{-1}\Big(\sqrt{nh_n}E(\tau|X = x) \\ &\quad + O_p(1)\Big) \\ &= O_p(\alpha_n^{-1/2}) + \psi(\alpha_n)(\bar{N}(T)/\alpha_n)^{-1} E(\tau|X = x) \\ &\quad + (\psi(\alpha_n)/\sqrt{nh_n})(\bar{N}(T)/\alpha_n)^{-1} O_p(1) \\ &= \psi(\alpha_n)\,\frac{E(\tau|X = x)}{E(\tau)} + O_p(1),\end{aligned}$$

where the second equality follows from Lemma S.3 and by applying Lemma S.6 below with $h = h_n = c_0 n^{-1/5}$ for some constant $c_0 > 0$; and the third and last equalities

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

follows from Lemma S.5 and since $\psi(\alpha_n) = O(\sqrt{nh_n})$ with $nh_n = n^{4/5} \to \infty$ as $n \to \infty$. The result then follows using that $\psi(\alpha_n) \asymp n^{2/5}$.

C. Consistency of Local Linear Estimator

In this section we include for completeness a well known result regarding the asymptotic normality of the local linear regression estimate for the conditional mean function. Suppose that $(X_1, Y_1), \ldots, (X_n, Y_n) \overset{iid}{\sim} (X, Y)$ where X and Y are real valued. Let $m(x) = E(Y|X = x)$ be the regression function and $f(x) > 0$ be the design density function. Then, the local linear regression estimate [39] $\hat{m}(x)$ of $m(x)$ is given by

$$\hat{m}(x) = b_1^T \left(\frac{1}{n}\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1} \left(\frac{1}{n}\mathbf{X}^T\mathbf{W}\mathbf{Y}\right), \quad b_1 = (1\ 0)^T,$$

where $\mathbf{W} = \text{diag}(K_h(X_i - x))$, $K_h(\cdot) = K(\cdot)/h$ with $K(\cdot)$ a kernel function, $\mathbf{Y} = (Y_1, \ldots, Y_n)^T$, and the i^{th} row of $\mathbf{X}$ is given by $(1, X_i - x)$, $i = 1, \ldots, n$. Now, if we let $\hat{u}_j := n^{-1}\sum_{i=1}^n K_h(X_i - x)(X_i - x)^j$ and $\hat{\sigma}_0^2 := \hat{u}_0\hat{u}_2 - \hat{u}_1^2$,

$$\left(\frac{1}{n}\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1} = \begin{bmatrix} \frac{\hat{u}_2}{\hat{\sigma}_0^2} & -\frac{\hat{u}_1}{\hat{\sigma}_0^2} \\ -\frac{\hat{u}_1}{\hat{\sigma}_0^2} & \frac{\hat{u}_0}{\hat{\sigma}_0^2} \end{bmatrix}$$

and

$$\frac{1}{n}\mathbf{X}^T\mathbf{W}\mathbf{Y} = \begin{bmatrix} n^{-1}\sum_{i=1}^n K_h(X_i - x)Y_i \\ n^{-1}\sum_{i=1}^n K_h(X_i - x)(X_i - x)Y_i \end{bmatrix},$$

whence

$$\hat{m}(x) = \begin{bmatrix} \frac{\hat{u}_2}{\hat{\sigma}_0^2} & -\frac{\hat{u}_1}{\hat{\sigma}_0^2} \end{bmatrix} \begin{bmatrix} n^{-1}\sum_{i=1}^n K_h(X_i - x)Y_i \\ n^{-1}\sum_{i=1}^n K_h(X_i - x)(X_i - x)Y_i \end{bmatrix}. \tag{28}$$

We make the following assumptions:

- (A1) The regression function $m(x) = E(Y|X = x)$, the design density function $f(x) > 0$ and $\sigma^2(x) = E(e^2|X = x)$, where $e = Y - m(X)$, are twice continuously differentiable.
- (A2) The kernel $K(\cdot)$ is bounded and corresponds to a density function which is symmetric around zero and has compact support $[-1, 1]$.
- (A3) As $n \to \infty$, $nh^5 = O(1)$.
- (A4) There exists $\delta > 0$ and $\bar{\sigma} > 0$ such that $E(|e_i|^{2+\delta}|X_i) \leq \bar{\sigma}$, $i = 1, \ldots, n$, where $e_i := Y_i - m(X_i)$.

Assumption $(A3)$ implies that as $n \to \infty$ we have $h \to 0$, $nh^q \to \infty$ for $q \in \{0, 1, 2, 3, 4\}$ and $nh^p \to 0$ for $p \geq 6$.

By a second order Taylor expansion of $m(\cdot)$ around x

$$\begin{aligned} Y_i &= m(X_i) + e_i \\ &= m(x) + m'(x)(X_i - x) + \frac{m''(x)}{2}(X_i - x)^2 \\ &\quad + \frac{(m''(\xi(X_i, x)) - m''(x))}{2}(X_i - x)^2 + e_i, \end{aligned}$$

where $\xi(X_i, x)$ lies between x and X_i. By replacing the previous expression in (28) and after some algebra we obtain

$$\begin{aligned} &\hat{m}(x) \\ &= m(x) + \begin{bmatrix} \frac{\hat{u}_2}{\hat{\sigma}_0^2} & -\frac{\hat{u}_1}{\hat{\sigma}_0^2} \end{bmatrix} \\ &\begin{bmatrix} \hat{u}_2\frac{m''(x)}{2} + n^{-1}\sum_{i=1}^n K_h(X_i - x)e_i + R_1 \\ \hat{u}_3\frac{m''(x)}{2} + n^{-1}\sum_{i=1}^n K_h(X_i - x)(X_i - x)e_i + R_2 \end{bmatrix}, \end{aligned} \tag{29}$$

where the remainder terms are given by

$$R_1 = \frac{1}{n}\sum_{i=1}^n K_h(X_i - x)\frac{(m''(\xi(X_i, x)) - m''(x))}{2}(X_i - x)^2,$$

$$R_2 = \frac{1}{n}\sum_{i=1}^n K_h(X_i - x)\frac{(m''(\xi(X_i, x)) - m''(x))}{2}(X_i - x)^3.$$

We now study the asymptotic distribution of $\hat{m}(x)$ with a suitably chosen scaling factor. For this, we introduce the following quantities. Let i, j be non-negative integers and define $k_{ij} := \int_{-1}^1 K(u)^i u^j du$. The auxiliary lemma on the asymptotic normality of $\hat{m}(x)$ is listed for completeness only. Its proof follows by standard arguments. See for example Theorem 5.2 in [39] and the references therein.

Lemma S.6: Under assumptions $(A1) - (A4)$,

$$\sqrt{nh}\left(\hat{m}(x) - m(x) - h^2\frac{m''(x)}{2}k_{12}\right) \xrightarrow{\mathcal{L}} N\left(0, k_{20}\frac{\sigma^2(x)}{f(x)}\right),$$

as $n \to \infty$.

D. Proofs of Results in Section IV-D

We need the following auxiliary lemma.

Lemma S.7: Suppose that the conditions of Theorem 3 hold. Then

$$n^{-1}\sum_{i=1}^n s_i(x)\left(\frac{N_i^{(n)}(T)}{\alpha_n} - \tau_i\right) = O_p((n\alpha_n)^{-1/2}),$$

where $s_i(x) := 1 + (X_i - \mu)^T\Sigma^{-1}(x - \mu)$, $i = 1, \ldots, n$.

Proof of Lemma S.7: Letting $S_n = n^{-1}\sum_{i=1}^n s_i(x)(N_i^{(n)}(T)/\alpha_n - \tau_i)$, by independence we have

$$\begin{aligned} E(S_n^2) &= n^{-2}\sum_{i=1}^n E(s_i^2(x)(N_i^{(n)}(T)/\alpha_n - \tau_i)^2) \\ &\quad + n^{-2}\sum_{i=1}^n\sum_{k\neq i} E(s_i(x)(N_i^{(n)}(T)/\alpha_n - \tau_i)) \\ &\qquad E(s_k(x)(N_k^{(n)}(T)/\alpha_n - \tau_k)) \\ &= n^{-2}\sum_{i=1}^n E(s_i^2(x)(N_i^{(n)}(T)/\alpha_n - \tau_i)^2) \\ &= n^{-1}E(s_1^2(x)(N_1^{(n)}(T)/\alpha_n - \tau_1)^2), \end{aligned}$$

where the second equality is due to $E(s_i(x)(N_i^{(n)}(T)/\alpha_n - \tau_i)) = E[s_i(x)E(N_i^{(n)}(T)/\alpha_n - \tau_i|\Lambda_i X_i)] = 0$. Next, from Lemma S.2 and by a conditioning argument, it follows that

$$E(s_1^2(x)(N_1^{(n)}(T)/\alpha_n - \tau_1)^2)$$

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

$$= E(s_1^2(x)E[(N_1^{(n)}(T)/\alpha_n - \tau_1)^2|\Lambda_1, X_1])$$
$$\leq (M_1/\alpha_n)E(s_1^2(x))$$
$$= (M_1/\alpha_n)(1 + (x-\mu)^T\Sigma^{-1}(x-\mu)).$$

This implies $E(S_n^2) = O((n\alpha_n)^{-1})$ and the result follows.

Proof of Theorem 3: Consider the auxiliary quantities $s_i(x) := 1 + (X_i - \mu)^T\Sigma^{-1}(x-\mu)$, $W_{0n}(x) := \bar{X}^T\hat{\Sigma}^{-1}(x - \bar{X}) - \mu^T\Sigma^{-1}(x-\mu)$ and $W_{1n}(x) := \Sigma^{-1}(x-\mu) - \hat{\Sigma}^{-1}(x-\bar{X})$. The arguments in the proof of Theorem 1 in [29] show that $W_{0n} = O_p(n^{-1/2})$, $\|W_{1n}\|_2 = O_p(n^{-1/2})$ and $s_{in}(x) - s_i(x) = -W_{0n} - W_{1n}^T X_i$. Thus

$$n^{-1}\sum_{i=1}^n s_{in}(x)N_i^{(n)}(T)/\alpha_n$$
$$= -W_{0n}\bar{N}(T)/\alpha_n - n^{-1}\sum_{i=1}^n X_i^T(N_i^{(n)}(T)/\alpha_n)W_{1n}$$
$$+ n^{-1}\sum_{i=1}^n s_i(x)N_i^{(n)}(T)/\alpha_n.$$

Next, note that

$$W_{0n}\bar{N}(T)/\alpha_n = W_{0n}(\bar{N}(T)/\alpha_n - E(\tau)) + W_{0n}E(\tau)$$
$$= O_p(n^{-1/2}),$$

by Lemma S.4. Further, from the fact that

$$E\left(\|X_1\|_2 N_1^{(n)}(T)/\alpha_n\right) \leq \left(E\|X_1\|_2^2\ E(N_1^{(n)}(T)/\alpha_n)^2\right)^{1/2}$$
$$= \sqrt{(\mathrm{tr}(\Sigma) + \|\mu\|_2^2)(E(\tau^2) + o(1))},$$

is uniformly bounded and observing the inequality

$$\|n^{-1}\sum_{i=1}^n X_i^T(N_i^{(n)}(T)/\alpha_n)\|_2$$
$$\leq n^{-1}\sum_{i=1}^n (N_i^{(n)}(T)/\alpha_n)\|X_i\|_2,$$

it follows that $n^{-1}\sum_{i=1}^n X_i^T(N_i^{(n)}(T)/\alpha_n) = O_p(1)$ and thus $n^{-1}\sum_{i=1}^n X_i^T(N_i^{(n)}(T)/\alpha_n)W_{1n} = O_p(n^{-1/2})$. This shows that

$$n^{-1}\sum_{i=1}^n s_{in}(x)N_i^{(n)}(T)/\alpha_n = n^{-1}\sum_{i=1}^n s_i(x)N_i^{(n)}(T)/\alpha_n + O_p(n^{-1/2}). \quad (30)$$

Next, note that

$$n^{-1}\sum_{i=1}^n s_i(x)N_i^{(n)}(T)/\alpha_n$$
$$= n^{-1}\sum_{i=1}^n s_i(x)(N_i^{(n)}(T)/\alpha_n - \tau_i) + n^{-1}\sum_{i=1}^n s_i(x)\tau_i$$
$$= n^{-1}\sum_{i=1}^n s_i(x)(N_i^{(n)}(T)/\alpha_n - \tau_i) + E(s(X,x)\tau)$$
$$+ O_p(n^{-1/2})$$
$$= E(s(X,x)\tau) + O_p(n^{-1/2}),$$

where the second equality follows from the central limit theorem and the third is due to Lemma S.7 along with the fact that $\alpha_n \to \infty$ as $n \to \infty$. Combining this with (30) and the fact that $\psi(\alpha_n) \asymp n^{1/2}$ leads to

$$\psi(\alpha_n)\ n^{-1}\sum_{i=1}^n s_{in}(x)N_i^{(n)}(T)/\alpha_n = \psi(\alpha_n)E(s(X,x)\tau) + O_p(1).$$

Finally, arguments similar to those in the proof of Lemma S.5 lead to

$$\psi(\alpha_n)\ \frac{n^{-1}\sum_{i=1}^n s_{in}(x)N_i^{(n)}(T)}{\bar{N}(T)} = \psi(\alpha_n)\ \frac{E(s(X,x)\tau)}{E(\tau)} + O_p(1),$$

whence the result follows since $\psi(\alpha_n) \asymp n^{1/2}$.

Proof of Theorem 4: The proof follows by arguments similar to those in the proof of Proposition 1 and Theorem 2 in [29] and is therefore omitted.

E. Proofs of Results in Section V-A

We require the following auxiliary lemma.

Lemma S.8: Suppose that $(S1)$ and $(S2)$ hold. Let ν be a positive integer and r_j, $j = 1, \ldots, \nu$, be an equispaced grid in $(0,1)$, where $\Delta r_\nu = 1/(\nu+1)$ is the grid spacing. Then

$$\sup_{Q_1, Q_2 \in Q(\Omega_{\mathcal{F}})} \Big| \int_0^1 (Q_1(t) - Q_2(t))^2 dt - \sum_{j=1}^{\nu} (Q_1(r_j) - Q_2(r_j))^2 \Delta r_\nu \Big|$$
$$= o(1),$$

as $\nu \to \infty$.

Proof of Lemma S.8: Since $r_j = j/(\nu+1)$ and denoting by $r_0 = 0$, we have

$$\Big| \int_0^1 (Q_1(t) - Q_2(t))^2 dt - \sum_{j=1}^{\nu} (Q_1(r_j) - Q_2(r_j))^2 \Delta r_\nu \Big|$$
$$= \Big| \sum_{j=1}^{\nu} \Big[\int_{r_{j-1}}^{r_j} (Q_1(t) - Q_2(t))^2 dt - (Q_1(r_j) - Q_2(r_j))^2 \Delta r_\nu \Big] + \int_{r_\nu}^1 (Q_1(t) - Q_2(t))^2 dt \Big|$$
$$\leq \frac{T^2}{\nu+1} + \sum_{j=1}^{\nu} \Big| \int_{r_{j-1}}^{r_j} (Q_1(t) - Q_2(t))^2 dt - (Q_1(r_j) - Q_2(r_j))^2 \Delta r_\nu \Big|.$$

Next, using that $Q_1, Q_2 \in Q(\Omega_{\mathcal{F}})$ along with simple calculations shows that

$$\Big| \int_{r_{j-1}}^{r_j} (Q_1(t) - Q_2(t))^2 dt - (Q_1(r_j) - Q_2(r_j))^2 \Delta r_\nu \Big|$$
$$= \Big| \int_{r_{j-1}}^{r_j} (Q_1(t) - Q_2(t))^2 - (Q_1(r_j) - Q_2(r_j))^2 dt \Big|$$

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

$$\leq 2T \int_{r_{j-1}}^{r_j} |Q_1(t) - Q_1(r_j) + Q_2(r_j) - Q_2(t)|dt$$
$$\leq 4LT \int_{r_{j-1}}^{r_j} (r_j - t)dt = 2LT\Delta r_\nu^2.$$

Thus

$$\sum_{j=1}^{\nu} \Big| \int_{r_{j-1}}^{r_j} (Q_1(t) - Q_2(t))^2 dt - (Q_1(r_j) - Q_2(r_j))^2 \Delta r_\nu \Big| \leq 2LT\Delta r_\nu,$$

whence the result follows.

Proof of Proposition 2: We will show convergence along subsequences, which is a similar idea as in Proposition 4.1 in [54] or [55]. Recall that $Q_{\oplus\nu}$ is any (fixed) element in $\mathcal{S}_\nu$, which can be selected by the axiom of choice. Note that any sequence $q_n \in Q(\Omega_\mathcal{F})$ is uniformly bounded and uniformly equicontinuous. Since the q_n are continuous functions defined on $[0,1]$, an application of the Arzela-Ascoli theorem shows that $Q(\Omega_\mathcal{F})$ is a compact set in $L^2([0,1])$. Consider a sequence of positive integers $\nu = \nu_m$ such that $\nu_m \to \infty$ as $m \to \infty$. Note that

$$|\mathcal{M}(q) - \mathcal{M}_{\nu_m}(q)| = o(1),$$

as $m \to \infty$. Since $Q(\Omega_\mathcal{F})$ is compact, there exists a subsequence $(Q_{\oplus\nu_{m_k}})_{k\in\mathbb{N}}$ of $(Q_{\oplus\nu_m})_{m\in\mathbb{N}}$ which converges to an element $Q^* \in Q(\Omega_\mathcal{F})$ as $k \to \infty$, i.e., $\|Q_{\oplus\nu_{m_k}} - Q^*\|_{L^2([0,1])} \to 0$ as $k \to \infty$. Next, as $k \to \infty$ we have

$$\mathcal{M}(\tilde{Q}_\oplus(\cdot,x)) = \mathcal{M}_{\nu_{m_k}}(\tilde{Q}_\oplus(\cdot,x)) + o(1) \geq \mathcal{M}_{\nu_{m_k}}(Q_{\oplus\nu_{m_k}}) + o(1),$$

where the inequality is due to the fact that $Q_{\oplus\nu_{m_k}} \in \mathcal{S}_{\nu_{m_k}}$ minimizes $\mathcal{M}_{\nu_{m_k}}(\cdot)$ over the class $Q(\Omega_\mathcal{F})$ and $\tilde{Q}_\oplus(\cdot,x)$ is an element of the latter space. Note that

$$\begin{aligned}
&\mathcal{M}_{\nu_{m_k}}(Q_{\oplus\nu_{m_k}}) \\
&= \sum_{j=1}^{\nu_{m_k}} (Q_{\oplus\nu_{m_k}}(r_j) - w_j)^2 \Delta r_{\nu_{m_k}} \\
&= \sum_{j=1}^{\nu_{m_k}} (Q_{\oplus\nu_{m_k}}(r_j) - Q^*(r_j))^2 \Delta r_{\nu_{m_k}} \\
&\quad + \sum_{j=1}^{\nu_{m_k}} (Q^*(r_j) - w_j)^2 \Delta r_{\nu_{m_k}} \\
&\quad + 2\sum_{j=1}^{\nu_{m_k}} (Q_{\oplus\nu_{m_k}}(r_j) - Q^*(r_j))(Q^*(r_j) - w_j)\Delta r_{\nu_{m_k}} \\
&= \|Q_{\oplus\nu_{m_k}} - Q^*\|_{L^2([0,1])} + \mathcal{M}(Q^*) + o(1) \\
&\quad + O(\mathcal{M}(Q^*)\|Q_{\oplus\nu_{m_k}} - Q^*\|_{L^2([0,1])}) \\
&= \mathcal{M}(Q^*) + o(1),
\end{aligned}$$

as $k \to \infty$, where the third equality follows by using the uniform Riemann sum integrability over the class $Q(\Omega_\mathcal{F})$ shown in Lemma S.8 above along with the Cauchy-Schwarz inequality, and the last equality is due to $\|Q_{\oplus\nu_{m_k}} - Q^*\|_{L^2([0,1])} \to 0$ as $k \to \infty$. This shows that

$$\mathcal{M}(\tilde{Q}_\oplus(\cdot,x)) \geq \mathcal{M}(Q^*) + o(1),$$

and taking $k \to \infty$ leads to $\mathcal{M}(\tilde{Q}_\oplus(\cdot,x)) \geq \mathcal{M}(Q^*)$. Since $\tilde{Q}_\oplus(\cdot,x)$ is the unique solution to the optimization problem (8) involving $\mathcal{M}$, then $Q^*(\cdot) = \tilde{Q}_\oplus(\cdot,x)$. The previous arguments show that all convergent subsequences of $(Q_{\oplus\nu_m})_{m\in\mathbb{N}}$ converge to the same limit $\tilde{Q}_\oplus(\cdot,x)$. Since $Q_{\oplus\nu_m} \in Q(\Omega_\mathcal{F})$ for all $m \geq 1$ and $Q(\Omega_\mathcal{F})$ is compact, this implies that $Q_{\oplus\nu_m}$ converges to $\tilde{Q}_\oplus(\cdot,x)$ in the L^2 norm. The result follows.

Recall that $Q_\nu^*(t) = q_{\nu,j}^* + (t - r_j)(q_{\nu,j+1}^* - q_{\nu,j}^*)/\Delta r_\nu$ for $t \in [r_j, r_{j+1})$, $j = 0, \ldots, \nu$, where $q_{\nu,j}^*$ is the jth coordinate of q_ν^*, $j = 1, \ldots, \nu$, $q_{\nu,0}^* = 0$, $r_0 = 0$, $q_{\nu,\nu+1}^* = T$ and $r_{\nu+1} = 1$. By continuity we define $Q_\nu^*(1) := \lim_{t\to 1-} Q_\nu^*(t) = T$. Lemma S.9 below shows that Q_ν^* is in the quantile space $Q(\Omega_\mathcal{F})$.

Lemma S.9: Suppose that $(S1)$ and $(S2)$ hold. The linear interpolation function Q_ν^* satisfies $Q_\nu^* \in Q(\Omega_\mathcal{F})$.

Proof of Lemma S.9: Let $I_j = [r_j, r_{j+1})$, $j = 0, \ldots, \nu$, and $t_1, t_2 \in (0,1)$. If $t_1, t_2 \in I_j$, then $|Q_\nu^*(t_2) - Q_\nu^*(t_1)| = (q_{\nu,j+1}^* - q_{\nu,j}^*)|t_2 - t_1|/\Delta r_\nu$ and the constraints of the optimization problem (15) imply $M|t_2 - t_1| \leq |Q_\nu^*(t_2) - Q_\nu^*(t_1)| \leq L|t_2 - t_1|$. Next, consider the case when $t_1 \in I_j$ and $t_2 \in I_k$ for $j < k$. Note that $Q_\nu^*(r_{j+1}) - Q_\nu^*(t_1) = q_{\nu,j+1}^* - q_{\nu,j}^* - (t_1 - r_j)(q_{\nu,j+1}^* - q_{\nu,j}^*)/\Delta r_\nu = (q_{\nu,j+1}^* - q_{\nu,j}^*)(1 - (t_1 - r_j)/\Delta r_\nu)$, which implies $M(r_{j+1} - t_1) \leq Q_\nu^*(r_{j+1}) - Q_\nu^*(t_1) \leq L(r_{j+1} - t_1)$. Also

$$Q_\nu^*(t_2) - Q_\nu^*(t_1) = Q_\nu^*(t_2) - Q_\nu^*(r_k) + \sum_{l=j+2}^{k} (q_{\nu,l}^* - q_{\nu,l-1}^*) + Q_\nu^*(r_{j+1}) - Q_\nu^*(t_1),$$

where $\sum_{l=j+2}^{k}(q_{\nu,l}^* - q_{\nu,l-1}^*)$ is defined as zero whenever $j = k - 1$, and $M\Delta r_\nu \leq q_{\nu,l}^* - q_{\nu,l-1}^* \leq L\Delta r_\nu$, which is due to the constraints in (15). Combining this with

$$M(t_2 - r_k) \leq Q_\nu^*(t_2) - Q_\nu^*(r_k) \leq L(t_2 - r_k),$$

which is due to $t_2, r_k \in I_k$, leads to

$$M(t_2 - t_1) \leq Q_\nu^*(t_2) - Q_\nu^*(t_1) \leq L(t_2 - t_1).$$

Interchanging the role of t_1 and t_2 shows that $M|t_2 - t_1| \leq |Q_\nu^*(t_2) - Q_\nu^*(t_1)| \leq L|t_2 - t_1|$ for any $t_1, t_2 \in (0,1)$. Finally, by construction it is clear that $Q_\nu^*(0) = 0$ and $Q_\nu^*(1) = \lim_{t\to 1-} Q_\nu^*(t) = T$. Thus $Q_\nu^* \in Q(\Omega_\mathcal{F})$ and the result follows.

F. Additional Theoretical Results

Lemma S.10 and S.11 below present the explicit solution for the local Fréchet regression shape component $f_\oplus(x)$ and the corresponding global regression, respectively.

Lemma S.10: Suppose that $(S1)$ and $(S2)$ hold. The solution $f_\oplus(x)$ to the local Fréchet regression problem on the shape component

$$f_\oplus(x) = \operatorname*{arg\,min}_{f_0 \in \Omega_\mathcal{F}} E(d_\mathcal{F}^2(f, f_0)|X = x),$$

is given by the density function with corresponding quantile function $E(Q|X = x)$.

Proof of Lemma S.10: Denoting by Q and Q_0 the quantile functions corresponding to f and f_0, respectively, and

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

$Q_x := E(Q|X = x)$, then similarly as in the proof of Proposition 1 in [29] it follows that

$$\begin{aligned}
&E(d_{\mathcal{F}}^2(f, f_0)|X = x) \\
&\quad = E(\|Q - Q_0\|_{L^2([0,1])}^2|X = x) \\
&\quad = E(\|Q - Q_x\|_{L^2([0,1])}^2|X = x) + \|Q_x - Q_0\|_{L^2([0,1])}^2 \\
&\qquad + 2E(\langle Q - Q_x, Q_x - Q_0\rangle_{L^2([0,1])}|X = x) \\
&\quad = E(\|Q - Q_x\|_{L^2([0,1])}^2|X = x) + \|Q_x - Q_0\|_{L^2([0,1])}^2,
\end{aligned}$$

and thus the optimal solution Q_0 is achieved by setting $Q_0 = Q_x$, provided that we can show that Q_x lies in the space $Q(\Omega_{\mathcal{F}})$. Indeed, since $Q \in Q(\Omega_{\mathcal{F}})$ we have $M|t - s| \leq |Q(t) - Q(s)| \leq L|t - s|$, $t, s \in (0, 1)$. It is then easy to show that $M|t - s| \leq |Q_x(t) - Q_x(s)| \leq L|t - s|$. It is also clear that $Q_x(0) = 0$ and $Q_x(1) = T$, and the result follows.

Lemma S.11: Suppose that $(S1)$ and $(S2)$ hold. The solution $f_{G\oplus}(x)$ to the global Fréchet regression problem on the shape component

$$f_{G\oplus}(x) = \underset{f_0 \in \Omega_{\mathcal{F}}}{\arg\min}\, E(s(X, x)d_{\mathcal{F}}^2(f, f_0))$$

is given by the density function whose corresponding quantile function is equal to the L^2-orthogonal projection of $E(s(X, x)Q)$ on $Q(\Omega_{\mathcal{F}})$.

Proof of Lemma S.11: Denoting by Q and Q_0 the quantile functions corresponding to f and f_0, respectively, and $Q_x := E(s(X, x)Q)$, then similarly as in the proof of Lemma S.10 or Proposition 1 in [29] we have

$$\begin{aligned}
&E(s(X, x)d_{\mathcal{F}}^2(f, f_0)) \\
&\quad = E(s(X, x)\|Q - Q_x\|_{L^2([0,1])}^2) + \|Q_x - Q_0\|_{L^2([0,1])}^2.
\end{aligned}$$

Since $Q(\Omega_{\mathcal{F}})$ is closed and convex in $L^2([0, 1])$ due to Lemma 1, it follows that the optimal solution Q_0 exists and is unique, and corresponds to the orthogonal projection of Q_x on $Q(\Omega_{\mathcal{F}})$, and the result follows.

The following lemma presents explicit solutions of the local Fréchet regression function for the special case where the distributions associated to the random intensity factor and shape functions are point masses.

Lemma S.12: Suppose that $(S1)$-$(S3)$ hold and there exists $M_1 < \infty$ such that $\kappa \leq \tau \leq M_1$ almost surely with κ as in assumption $(S3)$. Also, suppose that $\mathrm{Var}(\tau) = 0$ and the distribution of the random density f corresponds to a point mass in the space of probability distributions $\Omega_{\mathcal{F}}$ endowed with the 2-Wasserstein metric. Then $f = g$ almost surely for some density g with corresponding quantile function $Q_g \in Q(\Omega_{\mathcal{F}})$, $\tau = \eta_0$ almost surely for some positive constant $\eta_0 \in [\kappa, M_1]$ and the local Fréchet regression function satisfies

$$\tilde{f}_{\oplus}(x) = g, \quad \tilde{\tau}_{\oplus}(x) = \eta_0.$$

Proof of Lemma S.12: Since the probability distribution of the random density f is a point mass in $\Omega_{\mathcal{F}}$, there exists a density function g with corresponding quantile function $Q_g \in Q(\Omega_{\mathcal{F}})$ such that $f = g$ almost surely. Similarly, $\tau = \eta_0$ a.s. for some $\eta_0 \in [\kappa, M_1]$. Let f_0 be a density with

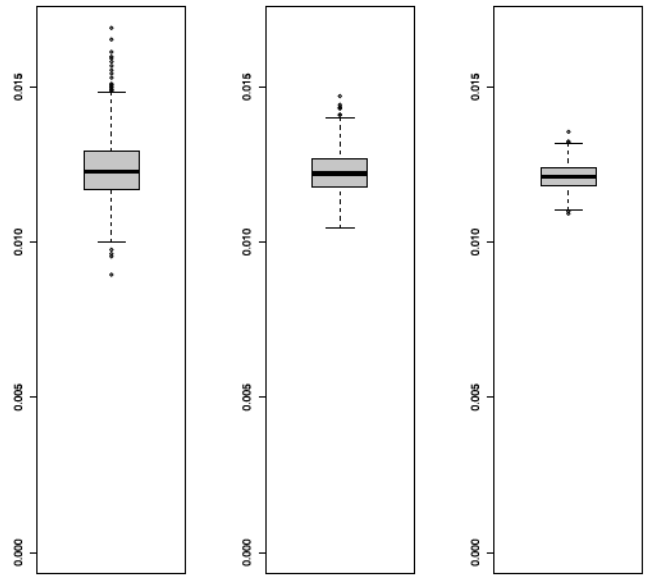

Fig. 8. Boxplots of the errors for the conditional shape function estimates $ISE_\tau^{\mathcal{F}}$ in the simulation setting for local Fréchet regression in section V-B using fixed $\alpha_n = 1$, which does not grow with n, for $n = 1000$ (left), $n = 2000$ (middle) and $n = 5000$ (right).

corresponding quantile function $Q_0 \in Q(\Omega_{\mathcal{F}})$ and let Q be the quantile function associated with f. Then

$$\begin{aligned}
E(d_{\mathcal{F}}^2(f, f_0)|X = x) &= E(||Q - Q_0||_{L^2([0,1])}^2) \\
&= ||Q_g - Q_0||_{L^2([0,1])}^2.
\end{aligned}$$

Thus, from (3) the minimizer has $Q_0 = Q_g$ which implies $\tilde{f}_{\oplus}(x) = g$, and the first result follows. Next, from (13) we have $\tau_{\oplus}(x) = \max\{E(\tau|X = x), 0\} = \eta_0$, implying the second result.

The following lemma shows the corresponding explicit solutions when considering the global Fréchet regression framework and the point mass probability distribution on the components τ and f.

Lemma S.13: Suppose that the same regularity conditions as in Lemma S.12 hold. Then $f = g$ almost surely for some density g with corresponding quantile function $Q_g \in Q(\Omega_{\mathcal{F}})$, $\tau = \eta_0$ almost surely for some positive constant $\eta_0 \in [\kappa, M_1]$ and the following relations hold for the global Fréchet regression function:

$$f_{G\oplus}(x) = g, \quad \tau_{G\oplus}(x) = \eta_0.$$

Proof of Lemma S.13: Analogously as in the proof of Lemma S.12 we have $f = g$ almost surely for some density g with corresponding quantile function $Q_g \in Q(\Omega_{\mathcal{F}})$ and $\tau = \eta_0$ almost surely for some positive constant $\eta_0 \in [\kappa, M_1]$. Denote by Q the quantile function associated with f and let f_0 be a density function with corresponding quantile $Q_0 \in Q(\Omega_{\mathcal{F}})$. Since

$$\begin{aligned}
E(s(X, x)d_{\mathcal{F}}^2(f, f_0) &= E(s(X, x)||Q - Q_0||_{L^2([0,1])}^2) \\
&= E(s(X, x))||Q_g - Q_0||_{L^2([0,1])}^2 \\
&= ||Q_g - Q_0||_{L^2([0,1])}^2,
\end{aligned}$$

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

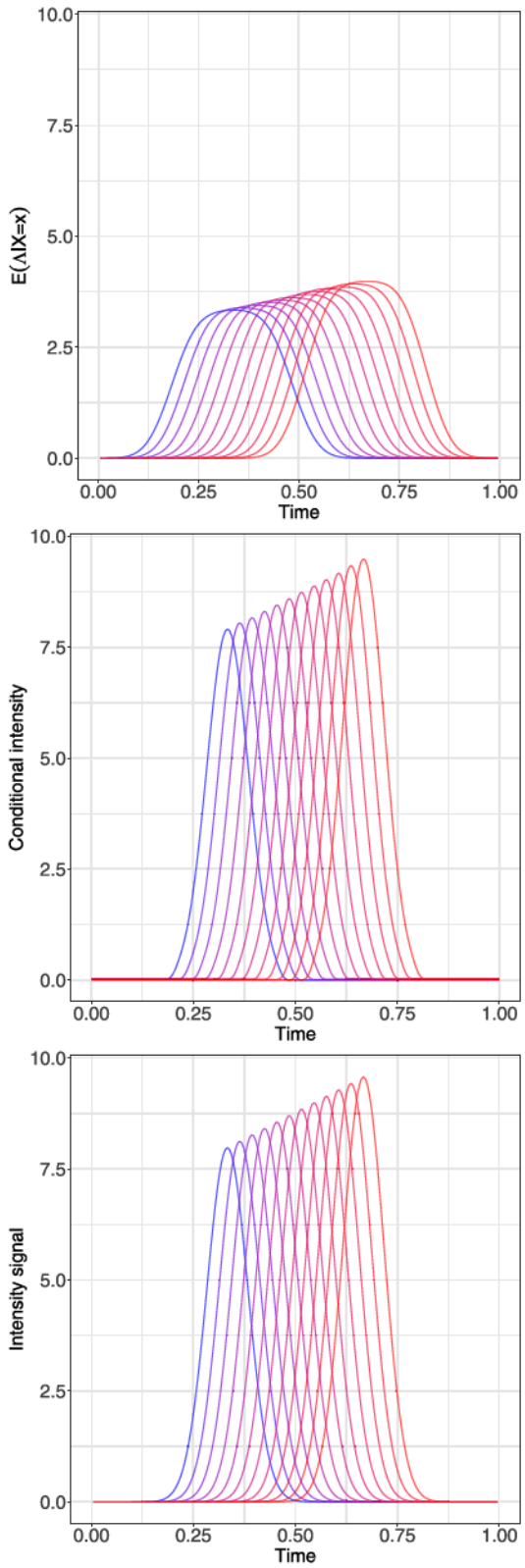

Fig. 9. Conditional intensity functions in the simulation setting over a dense grid of predictor levels x, displayed in blue when $x = 0$ to red when $x = 1$ in the online version (dark grey and light grey in the print version, resp.). The upper panel illustrates the results for classical Euclidean intensity regression function $E(\Lambda|X = x)$ while the middle panel illustrates the results for the conditional intensity $\Lambda_{\oplus}(x)$ that corresponds to the proposed point process regression. The bottom panel shows the true intensity signal Λ_x when there is no noise in either the intensity factor or the mean of the simulated conditional shape component.

which is due to $E(s(X, x)) = 1$, then the minimizer is attained when $Q_0 = Q_g \in Q(\Omega_{\mathcal{F}})$. From (14) it then follows that $f_{G\oplus}(x) = g$. Next, from (13) we have $\tau_{G\oplus}(x) = \max\{E(s(X, x)\tau), 0\} = \eta_0$ by again using that $E(s(X, x)) = 1$. The result follows.

G. Local Fréchet Regression When α_n Does Not Grow With Sample Size n

Figure 8 shows the integrated error metric for the shape part in the simulation settings for local Fréchet regression as outlined in section V-B when fixing $\alpha_n = 1$, so that α_n does not grow with n, and thus violates a basic assumption. We find that consistent recovery of the conditional intensity function is not possible if α_n is not allowed to increase with n as the integrated error box plots for the shape component show no decline in bias and stay well bounded away from zero for increasing sample size.

H. Comparison Between Standard Euclidean Intensity Regression Function and Local Fréchet Regression

In this section we compare through a simulation example the standard Euclidean intensity regression function and the local Fréchet regression counterpart. We consider the same Gaussian data generation mechanism as the one outlined in the global framework in section V-C but modifying the following population parameters: $a_2 = b_2 = 1/3$, $a_3 = 0.05$, $b_3 = 0$, $e_1 = -0.15$, $f_1 = 0.15$, $\sigma_1 = 3$, $\sigma_2 = 0$. Here we do not consider the error ε_2 on the standard deviation $\sigma(x)$ of the generated random densities, but rather keep it constant at $\sigma(x) = a_3$ for all x. This reflects a similar situation of horizontal translation of a Gaussian random variable as here only the mean increases with x while the standard deviation remains small and constant. It is easy to show that the standard Euclidean intensity regression function

$$g_x(\cdot) = E(\Lambda(\cdot)|X = x),$$

is given by $g_x(\cdot) = E(f(\cdot)|X = x)E(\tau|X = x)$. We approximate $E(f(\cdot)|X = x)$ through a Monte Carlo approach where we average across random densities f_i generated at predictor level x. Similarly, for the local Fréchet regression we obtain $E(Q(\cdot)|X = x)$ by averaging the corresponding random quantiles Q_i. To compare both quantities, denote by $f_x(\cdot)$ the density function which corresponds to the truncated Gaussian model considered before but disregarding the error ε_1 affecting its mean. Thus $f_x(\cdot)$ is the true density (without noise) at level $X = x$. The intensity signal $\Lambda_x(\cdot)$ is then constructed as $\Lambda_x = f_x E(\tau|X = x)$ and corresponds to the underlying intensity function in dependence on x after removing noise.

Figure 9 presents g_x in the upper left panel over a grid of values for x while the upper right panel displays the local Fréchet regression function $\Lambda_{\oplus}(x)$. The true intensity signal Λ_x is in the bottom panel. One finds that the shape component of the standard regression function g_x is distorted and does not lie in the Gaussian class where the random densities are situated. This is due to the noise in the mean function $\mu(x)$ which produces Gaussian densities that are centered at

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

the true signal f_x, and thus $E(f|X = x)$ corresponds to a mixture distribution which resides outside the Gaussian class. The conditional Fréchet regression function defined through the 2-Wasserstein barycenter is able to correctly capture the underlying geometry of the intensity space as its shape components remain in the Gaussian ensemble where the true signal was generated from, and thus provides a reasonable notion of center or mean in intensity space. If the variance of the noise ε_1 is very low, then both quantities are similar.

ACKNOWLEDGMENT

The authors thank the reviewers for helpful comments that led to numerous improvements.

REFERENCES

[1] T. Mikosch, *Non-Life Insurance Mathematics: An Introduction With the Poisson Process*. Berlin, Germany: Springer, 2009.
[2] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes: Elementary Theory and Methods*, vol. 1, 2nd ed. New York, NY, USA: Springer, 2003.
[3] D. Gervini and M. Khanal, "Exploring patterns of demand in bike sharing systems via replicated point process models," *J. Roy. Stat. Soc., C, Appl. Statist.*, vol. 68, no. 3, pp. 585–602, Apr. 2019.
[4] S. K. Reddy and M. Dass, "Modeling on-line art auction dynamics using functional data analysis," *Stat. Sci.*, vol. 21, no. 2, pp. 179–193, May 2006.
[5] G. Shmueli, R. P. Russo, and W. Jank, "The BARISTA: A model for bid arrivals in online auctions," *Ann. Appl. Statist.*, vol. 1, no. 2, pp. 412–441, Dec. 2007.
[6] D. R. Cox and V. Isham, "Point processes," in *Monographs on Applied Probability and Statistics*. London, U.K.: Chapman & Hall, 1980.
[7] P. Diggle, "A kernel method for smoothing point process data," *Appl. Statist.*, vol. 34, no. 2, pp. 138–147, 1985.
[8] P. J. Diggle, P. Moraga, B. Rowlingson, and B. M. Taylor, "Spatial and spatio-temporal log-Gaussian Cox processes: Extending the geostatistical paradigm," *Stat. Sci.*, vol. 28, no. 4, pp. 542–563, Nov. 2013.
[9] P. Reynaud-Bouret, "Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities," *Probab. Theory Rel. Fields*, vol. 126, no. 1, pp. 103–153, 2003.
[10] P. Reynaud-Bouret and V. Rivoirard, "Near optimal thresholding estimation of a Poisson intensity on the real line," *Electron. J. Statist.*, vol. 4, pp. 172–238, Jan. 2010.
[11] A. Cowling, P. Hall, and M. J. Phillips, "Bootstrap confidence regions for the intensity of a Poisson point process," *J. Amer. Stat. Assoc.*, vol. 91, no. 436, pp. 1516–1524, Dec. 1996.
[12] R. M. Willett and R. D. Nowak, "Multiscale Poisson intensity and density estimation," *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3171–3187, Sep. 2007.
[13] J. Bigot, S. Gadat, T. Klein, and C. Marteau, "Intensity estimation of non-homogeneous Poisson processes from shifted trajectories," *Electron. J. Statist.*, vol. 7, pp. 881–931, Jan. 2013.
[14] L. M. Leemis, "Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process," *Manage. Sci.*, vol. 37, no. 7, pp. 886–900, 1991.
[15] S. G. Henderson, "Estimation for nonhomogeneous Poisson processes from aggregated data," *Oper. Res. Lett.*, vol. 31, no. 5, pp. 375–382, Sep. 2003.
[16] J. Møller and R. P. Waagepetersen, *Statistical Inference and Simulation for Spatial Point Processes*. Boca Raton, FL, USA: CRC Press, 2003.
[17] G. Xu, C. Zhao, A. Jalilian, R. Waagepetersen, J. Zhang, and Y. Guan, "Nonparametric estimation of the pair correlation function of replicated inhomogeneous point processes," *Electron. J. Statist.*, vol. 14, no. 2, pp. 3730–3765, Jan. 2020.
[18] T. Zhang and S. C. Kou, "Nonparametric inference of doubly stochastic Poisson process data via the kernel method," *Ann. Appl. Statist.*, vol. 4, no. 4, pp. 1913–1941, Dec. 2010.
[19] Y. Li and Y. Guan, "Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance," *J. Amer. Stat. Assoc.*, vol. 109, no. 507, pp. 1205–1215, Jul. 2014.
[20] J. F. Lawless, "Regression methods for Poisson process data," *J. Amer. Stat. Assoc.*, vol. 82, no. 399, pp. 808–815, Sep. 1987.
[21] P. R. Bouzas, M. J. Valderrama, A. M. Aguilera, and N. Ruiz-Fuentes, "Modelling the mean of a doubly stochastic Poisson process by functional data analysis," *Comput. Statist. Data Anal.*, vol. 50, no. 10, pp. 2655–2667, Jun. 2006.
[22] P. Bouzas and N. Ruiz-Fuentes, "A review on functional data analysis for Cox processes," *Boletín Estadística Investigación Operativa*, vol. 31, no. 3, pp. 215–230, Nov. 2015.
[23] D. Gervini, "Multiplicative component models for replicated point processes," 2017, *arXiv:1705.09693*.
[24] G. Xu *et al.*, "Semi-parametric learning of structured temporal point processes," *J. Mach. Learn. Res.*, vol. 21, no. 192, pp. 1–39, 2020.
[25] U. Grenander, "Stochastic processes and statistical inference," *Arkiv Matematik*, vol. 1, pp. 195–277, Oct. 1950.
[26] J. Kleffe, "Principal components of random variables with values in a separable Hilbert space," *Mathematische Operationsforschung Statistik*, vol. 4, pp. 391–406, Jan. 1973.
[27] A. Petersen and H.-G. Müller, "Functional data analysis for density functions by transformation to a Hilbert space," *Ann. Statist.*, vol. 44, no. 1, pp. 183–218, Feb. 2016.
[28] H.-G. Müller, S. Wu, and Z. Zhang, "Functional data analysis for point processes with rare events," *Stat. Sinica*, vol. 23, no. 1, pp. 1–23, 2014.
[29] A. Petersen and H.-G. Müller, "Fréchet regression for random objects with Euclidean predictors," *Ann. Statist.*, vol. 47, no. 2, pp. 691–719, Apr. 2019.
[30] V. M. Panaretos and Y. Zemel, "Amplitude and phase variation of point processes," *Ann. Statist.*, vol. 44, no. 2, pp. 771–812, Apr. 2016.
[31] D. Burago, Y. Burago, and S. Ivanov, *A Course in Metric Geometry*, vol. 33. Providence, RI, USA: American Mathematical Society, 2001.
[32] C. Villani, *Topics in Optimal Transportation*. Providence, RI, USA: American Mathematical Society, 2003.
[33] B. M. Bolstad, R. A. Irizarry, M. Åstrand, and T. P. Speed, "A comparison of normalization methods for high density oligonucleotide array data based on variance and bias," *Bioinformatics*, vol. 19, no. 2, pp. 185–193, 2003.
[34] V. M. Panaretos and Y. Zemel, "Statistical aspects of Wasserstein distances," *Annu. Rev. Statist. Appl.*, vol. 6, no. 1, pp. 405–431, Mar. 2019.
[35] M. Fréchet, "Les éléments aléatoires de nature quelconque dans un espace distancié," *Annales l'Institut Henri Poincaré*, vol. 10, no. 4, pp. 215–310, 1948.
[36] M. Agueh and G. Carlier, "Barycenters in the Wasserstein space," *SIAM J. Math. Anal.*, vol. 43, no. 2, pp. 904–924, 2011.
[37] E. Cazelles, V. Seguy, J. Bigot, M. Cuturi, and N. Papadakis, "Geodesic PCA versus log-PCA of histograms in the Wasserstein space," *SIAM J. Sci. Comput.*, vol. 40, no. 2, pp. B429–B456, Jan. 2018.
[38] J. Bigot, E. Cazelles, and N. Papadakis, "Data-driven regularization of Wasserstein barycenters with an application to multivariate density registration," *Inf. Inference, A, J. IMA*, vol. 8, no. 4, pp. 719–755, Dec. 2019.
[39] J. Fan and I. Gijbels, *Local Polynomial Modelling and Its Applications*. London, U.K.: Chapman & Hall, 1996.
[40] L. Cucala, "Intensity estimation for spatial point processes observed with noise," *Scandin. J. Statist.*, vol. 35, no. 2, pp. 322–334, Jun. 2008.
[41] P. Diggle and J. S. Marron, "Equivalence of smoothing parameter selectors in density and intensity estimation," *J. Amer. Stat. Assoc.*, vol. 83, no. 403, pp. 793–800, Sep. 1988.
[42] P. A. W. Lewis and G. S. Shedler, "Statistical analysis of non-stationary series of events in a data base system," *IBM J. Res. Develop.*, vol. 20, no. 5, pp. 465–482, Sep. 1976.
[43] S. Lee, J. R. Wilson, and M. M. Crawford, "Modeling and simulation of a nonhomogeneous Poisson process having cyclic behavior," *Commun. Statist. Simul. Comput.*, vol. 20, nos. 2–3, pp. 777–809, Jan. 1991.
[44] M. E. Kuhl, J. R. Wilson, and M. A. Johnson, "Estimating and simulating Poisson processes having trends or multiple periodicities," *IIE Trans.*, vol. 29, no. 3, pp. 201–211, Mar. 1997.
[45] M. E. Kuhl and J. R. Wilson, "Least squares estimation of nonhomogeneous Poisson processes," *J. Stat. Comput. Simul.*, vol. 67, no. 1, pp. 699–712, Aug. 2000.
[46] E. P. C. Kao and S.-L. Chang, "Modeling time-dependent arrivals to service systems: A case in using a piecewise-polynomial rate function in a nonhomogeneous Poisson process," *Manage. Sci.*, vol. 34, no. 11, pp. 1367–1379, Nov. 1988.
[47] M. E. Kuhl and P. S. Bhairgond, "Nonparametric estimation of nonhomogeneous Poisson processes using wavelets," in *Proc. Winter Simulation Conf.*, vol. 1, Dec. 2000, pp. 562–571.

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.

[48] M. E. Kuhl and J. R. Wilson, "Modeling and simulating Poisson processes having trends or nontrigonometric cyclic effects," *Eur. J. Oper. Res.*, vol. 133, no. 3, pp. 566–582, 2001.

[49] J. Fan, I. Gijbels, T.-C. Hu, and L.-S. Huang, "A study of variable bandwidth selection for local polynomial regression," *Stat. Sinica*, vol. 6, no. 1, pp. 113–127, 1996.

[50] T. Hsing and R. Eubank, *Theoretical Foundations of Functional Data Analysis, With an Introduction to Linear Operators*. Hoboken, NJ, USA: Wiley, 2015.

[51] P. Borgnat, P. Abry, P. Flandrin, C. Robardet, J.-B. Rouquier, and E. Fleury, "Shared bicycles in a city: A signal processing and data analysis perspective," *Adv. Complex Syst.*, vol. 14, no. 3, pp. 415–438, Jun. 2011.

[52] H. R. Sayarshad and J. Y. J. Chow, "Survey and empirical evaluation of nonhomogeneous arrival process models with taxi data," *J. Adv. Transp.*, vol. 50, no. 7, pp. 1275–1294, Nov. 2016.

[53] T. Utsu, Y. Ogata, and R. S. Matsu'ura, "The centenary of the Omori formula for a decay law of aftershock Activity," *J. Phys. Earth*, vol. 43, no. 1, pp. 1–33, 1995.

[54] G. Peyré and M. Cuturi, "Computational optimal transport: With applications to data science," *Found. Trends Mach. Learn.*, vol. 11, nos. 5–6, pp. 355–607, 2019.

[55] Q. Berthet, M. Blondel, O. Teboul, M. Cuturi, J.-P. Vert, and F. Bach, "Learning with differentiable perturbed optimizers," in *Advances in Neural Information Processing Systems*, vol. 33, H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, Eds. Red Hook, NY, USA: Curran Associates, 2020, pp. 9508–9519.

Álvaro Gajardo is currently pursuing the Ph.D. degree in statistics with the University of California at Davis. His current research interests include functional data analysis, point processes, and Wasserstein analysis and applications to Covid-19 trajectory modeling.

Hans-Georg Müller is currently a Professor of statistics with the University of California at Davis. His main current research interests include functional data analysis, geometric statistics and the statistical analysis of random objects in metric spaces, including Fréchet regression, networks, distributional data analysis, and optimal transport models and applications.

Authorized licensed use limited to: Univ of Calif Davis. Downloaded on May 04,2022 at 09:33:43 UTC from IEEE Xplore. Restrictions apply.