Scientific Research Publishing

Open Journal of Statistics, 2024, 14, 150-162
https://www.scirp.org/journal/ojs
ISSN Online: 2161-7198
ISSN Print: 2161-718X

Using Extreme Value Theory Approaches to Estimate High Quantiles for Stroke Data

Justin Ushize Rutikanga[1,2*], **Aliou Diop**[3], **Charline Uwilingiyimana**[2,4]

[1]African Institute of Mathematical Sciences Rwandan Center, Kigali, Rwanda
[2]College of Agriculture, Animal Sciences and Veterinary Medicine (CAVM), University of Rwanda, Musanze, Rwanda
[3]LERSTAD, Gaston Berger University, Saint Louis, Senegal
[4]Department of Statistics Applied to Economy, INES Ruhengeri Institute of Applied Sciences, Musanze, Rwanda
Email: *ushizerj@gmail.com

How to cite this paper: Rutikanga, J.U., Diop, A. and Uwilingiyimana, C. (2024) Using Extreme Value Theory Approaches to Estimate High Quantiles for Stroke Data. *Open Journal of Statistics*, **14**, 150-162.
https://doi.org/10.4236/ojs.2024.141007

Received: December 4, 2023
Accepted: February 26, 2024
Published: February 29, 2024

Abstract

This paper aims to explore the application of Extreme Value Theory (EVT) in estimating the conditional extreme quantile for time-to-event outcomes by examining the functional relationship between ambulatory blood pressure trajectories and clinical outcomes in stroke patients. The study utilizes EVT to analyze the functional connection between ambulatory blood pressure trajectories and clinical outcomes in a sample of 297 stroke patients. The 24-hour ambulatory blood pressure measurement curves for every 15 minutes are considered, acknowledging a censored rate of 40%. The findings reveal that the sample mean excess function exhibits a positive gradient above a specific threshold, confirming the heavy-tailed distribution of data in stroke patients with a positive extreme value index. Consequently, the estimated conditional extreme quantile indicates that stroke patients with higher blood pressure measurements face an elevated risk of recurrent stroke occurrence at an early stage. This research contributes to the understanding of the relationship between ambulatory blood pressure and recurrent stroke, providing valuable insights for clinical considerations and potential interventions in stroke management.

Keywords

Censored Data, Conditional Extreme Quantile, Kernel Estimator, Weibull Tail Coefficient

1. Introduction

Recurrent stroke is considered one of the leading causes of death and disability worldwide, accounting for approximately 5 million deaths annually, which constitutes 9% of the total. Additionally, another 5 million people suffer from long-term disability.

According to the literature, numerous risk factors for recurrent stroke exist, including age, sex, smoking status, high blood pressure measurement, and lipid metabolism. Several researchers have identified these risk factors, as noted by [1]. Their findings reveal a significant increase in systolic blood pressure among patients with late recurrent stroke. Besides, [2] identified hypertension as the leading cause of recurrent stroke.

Furthermore, [3] discovered a significantly higher stroke recurrence rate in men, older individuals, and those with a prior history of ischemic stroke compared to women, younger individuals, and those with no history of stroke. Moreover, in their results, [4] recommend considering hypertension, diabetes mellitus, atrial fibrillation, and coronary heart disease as factors associated with a high risk of stroke recurrence.

In the context of this paper, the time of stroke occurrence is treated as survival data. Survival or (time-to-event) data analysis problems have arisen in a number of scientific fields. For instance, an event time of interest can be the survival time of a stroke patient in a medical study, the time to high school dropout studied by sociologists, the survival time of a new business addressed in economic research, or a lifetime of a part under stress evaluated in an engineering reliability study.

A common characteristic of survival data is often the presence of incomplete time-to-event information due to censoring or truncation. Here, we consider that the censoring appeared when a time to-event is known to have occurred only within certain intervals. Besides, truncation is defined as a condition which excludes certain subjects from the study population for more details see [5]. However, survival data analysis needs an appropriate statistical approach which takes into account a different form of censoring. Many authors have addressed this issue, we can cite a few among them [6] [7] [8] [9] for more details.

Nowadays, due to the progress in technology, it is possible for some covariate information to be recorded simultaneously with the quantity of interest in some sort of continuum. This continuum may have a link with time, and space or originate from multiple sources. For those kinds of problems, we deal with the statistical unit as a curve, a space or any more complex mathematic object having the concept of some continuum feature. Then such data are called function data by enumerating few of authors who work with the functional data such as [10] [11].

This paper proposes to focus on three statistical aspects in order to derive a methodology for estimating conditional extreme quantiles where the variable of interest has a heavy-tailed distribution under right random censoring in the presence of functional random covariate.

Let us consider $Y_1,\cdots,Y_n$ independent identically distributed copies of random variable of time to the event of interest Y. It has become a challenge in several fields to estimate extreme quantiles of the distribution of Y which has the form $F^{\leftarrow}(1-\alpha)=\inf\{y:F(y)\geq 1-\alpha\}$, with α is small such that α closed to zero as the sample size is large enough. Further, that quantile falls beyond the range of the observed data $Y_1,\cdots,Y_n$. According to the literature, extreme value theory has been proven to be a powerful tool for studying the behavior of extreme event distributions and is widely used in the estimation of the extreme value index (tail) of the distribution of Y. The extreme value index measures the tail heaviness of the distribution of Y and thus has a key role in the analysis of extreme event distribution. One of the known famous results in extreme value theory is the Fisher-Tippett-Gnedenko Theorem [12] [13].

As aforementioned, the estimation of the extreme-value index or tail-index is a cornerstone when we deal with various problems in extreme value analysis such as the estimation of the conditional extreme quantile of a random variable in the presence of covariate. Nevertheless, in this paper, we consider the situation where some covariate information X is available to the investigator, and the distribution of Y depends on X. Our focus centers on the problem of estimating a conditional extreme quantile of a heavy-tailed distribution when there is access to functional covariate information $X \in \mathbf{E}$ is available, where $\mathbf{E}$ is an infinite dimensional space associated with a semi-metric $d(\cdot,\cdot)$.

Recently, many authors have been interested in the estimation of the extreme value index and extreme quantile we can enumerate a few of them such as [14] [15] [16] [17] have considered the cases of the estimation of extreme value index and extreme quantile from censored data when the covariate information is not available. In [14] the authors proposed to estimate the extreme value index by using the modification of Hill's estimator version. In [18] [19] [20] authors proposed the Bayesian extreme value index and extreme quantile for the case of uncensored data. [21] [22] [23] investigated the estimation of extreme value index and extreme quantile where there is no covariate information and censored data are taken into consideration. [7] investigate the estimation of the conditional extreme value-index and conditional extreme quantile under randomly right censored with the presence of covariate for finite dimension.

Motivated by studies that utilize conditional extreme quantiles to assess the probability of survival for AIDS patients across various age groups within heavy-tailed distributions in the presence of finite-dimensional covariates, this study aims to estimate the conditionally extreme quantile of recurrent stroke occurrence time distribution under right random censoring. The ambulatory blood pressure curve will be considered as a functionally random covariate. However, the aim of this study is to estimate the conditionally extreme quantile of the time of occurrence of recurrent stroke distribution under right random censoring, with the ambulatory blood pressure curve as a functionally random covariate.

The remainder of this paper is organized as follows. Section 2 is devoted to the data description and the theoretical framework. A real data application illustrates the use of our estimators in Section 3, while Section 4 presents the discussion of our results. Finally, the conclusion and some perspectives are presented in Section 5.

2. Materials and Methods

2.1. Data Description

The data used in paper obtained by considering $n = 297$ stroke patients, consists of triplet $\left(X_i, \delta_i, Y_i\right)$, where X_i is the 24-hr ambulatory blood pressure curve of i^{th} patient, while δ_i is indicator function equal to one when a patient i is uncensored, otherwise equal to zero. The censoring rate is 40%. Finally, Y_i is an interesting clinical outcome about i^{th} stroke patients. The primary endpoint is the time to the composite stroke recurrent event, including death, disability, or vascular events (see [24] for more details). Each patient's systolic blood pressure (SBP) is measured every 15 min starting from 19:00 for 24 hr. The covariate X_i is thus defined by $X_i = \left(x_{i,1}, \cdots, x_{i,96}\right)$ with $x_{i,j}$ the SBP for each patient for all $i = 1, \cdots, 297$.

The data is available online at https://amstat.tandfonline.com/doi/suppl/10.1080/01621459.2019.1602047/suppl_file/uasa_a_1602047_sm0766.zip. Figure 1 below illustrates some realizations of random curves of the given functional random variable $X(\cdot)$. The covariate X_i is in fact a discretized curve but the fineness of the grid spanning the discretization allows us to consider each subject as a continuous curve as stated in [25]. Hence, the covariate can be considered as belonging to an infinite dimensional space $\mathbf{E}$. Figure 2 shows an estimated density of the time to recurrent stroke.

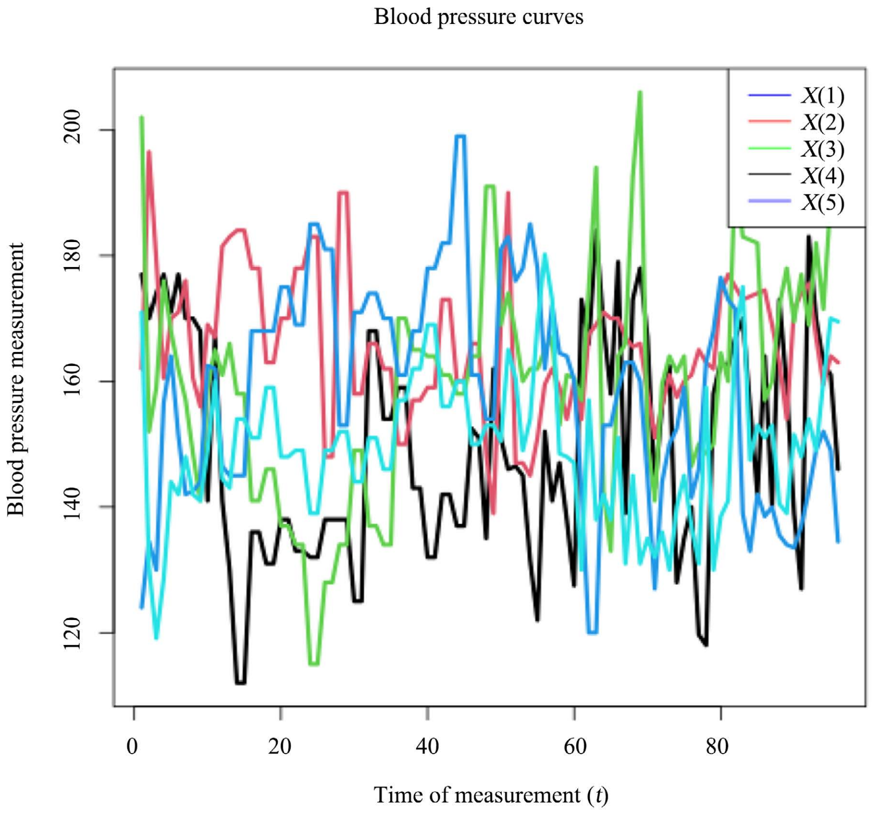

Figure 1. Measurement of blood pressure at $t = 1, \cdots, 96$.

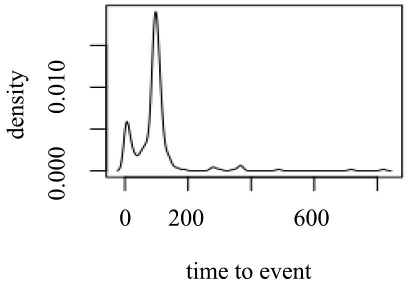

Figure 2. Density function of the time to stroke.

2.2. Extreme Value Theory

Let Y_i be the survival time, C_i be the censoring time and then let X_i be a functional random variable covariate. Let (X_i, Y_i) $i = 1, \cdots, n$ be the independent copies of the random pairs (X, Y), where Y is positive real random variable and $X \in \mathbf{E}$, where $\mathbf{E}$ is an infinite dimensional space associated to a semi-metric $d(.,.)$. Therefore, we really observe independent triplets (X_i, δ_i, Z_i), where $Z_i = \min(Y_i, C_i)$ and $\delta_i = \mathbf{1}_{\{Y_i \le C_i\}}$ for $i = 1, \cdots, n$ where $\mathbf{1}_A$ is the indicator function of the event A.

Let $F(\cdot|x)$ and $G(\cdot|x)$ be the conditional cumulative distribution functions of random variables Y and C given $X = x$ respectively.

Let $\bar{F}(\cdot|x)$ and $\bar{G}(\cdot|x)$ be the conditional survival function of random variable Y and C given $X = x$ respectively.

In this paper, we focus on heavy tails. More specifically, we assume that the conditional survival functions satisfy the following assumption.

(A1).

$$\bar{F}(t|x) = r_1(x)\exp\left\{-\int_1^t \left(\frac{1}{\gamma_1(x)} - \varepsilon_1(\mu|x)\right)\frac{\mathrm{d}\mu}{\mu}\right\} \tag{1}$$

and

$$\bar{G}(t|x) = r_2(x)\exp\left\{-\int_1^t \left(\frac{1}{\gamma_2(x)} - \varepsilon_2(\mu|x)\right)\frac{\mathrm{d}\mu}{\mu}\right\} \tag{2}$$

where $\gamma_1(x), \gamma_2(x)$ are positive unknown functions of the covariate *x*, r_1, r_2 are positive functions and $|\varepsilon_1(\mu|x)|$, $|\varepsilon_2(\mu|x)|$ are continuous and ultimately decreasing to zero. From (1) and (2), we can state that the conditional distribution functions of Y and C given $X = x$ are in Fréchet maximal domain of attraction. Thus, $\gamma_1(x)$ and $\gamma_2(x)$ are taken as the conditional extreme tail index functions. Therefore, for all $t > 0$, $\bar{F}(\cdot|x)$ and $\bar{G}(\cdot|x)$ are regularly varying functions at infinity with index $-\frac{1}{\gamma_1(x)}$ and $-\frac{1}{\gamma_2(x)}$ respectively. Thus,

$$\bar{F}(u|x) = u^{-\frac{1}{\gamma_1(x)}} L_1(u|x) \text{ and } \bar{G}(u|x) = u^{-\frac{1}{\gamma_2(x)}} L_2(u|x) \tag{3}$$

where for x fixed, $L_1(.|x)$ and $L_2(.|x)$ are slowly varying functions at infinity, that is, for all $\lambda > 0$,

$$\lim_{u\to\infty}\frac{L_i(\lambda u\,|\,x)}{L_i(u\,|\,x)}=1,\quad i=1,2.$$

By conditional independence between *Y* and *C*, the conditional survival function $\bar{H}(\cdot\,|\,x)$ of *Z* given $X=x$ is also a regularly varying function at infinity with index $-\frac{1}{\gamma(x)}$ as expressed as follows:

$$\begin{aligned}\bar{H}(\mu\,|\,x)&=1-H(\mu\,|\,x)=\bar{F}(\mu\,|\,x)\bar{G}(\mu\,|\,x)\\&=r(x)\exp\left\{-\int_1^z\left(\frac{1}{\gamma(x)}-\varepsilon(\mu\,|\,x)\right)\frac{\mathrm{d}\mu}{\mu}\right\}\end{aligned}\tag{4}$$

with $\gamma(x)=\gamma_1(x)p(x)$ where $p(x)=\frac{\gamma_2(x)}{\gamma_1(x)+\gamma_2(x)}$ is the ultimate proportion of uncensored observations among $Z_i, i=1,\cdots,n$; (see [15] [26] for more details) and $r(x)=r_1(x)r_2(x)$, $\varepsilon(\mu\,|\,x)=\varepsilon_1(\mu\,|\,x)+\varepsilon_2(\mu\,|\,x)$.

2.3. Estimation of Conditional Extreme Tail Index

Let (X_i,δ_i,Z_i), $i=1,\cdots,n$, be independent realizations of the random vector (X,δ,Z) where $Z_i=\min(Y_i,C_i)$ and $\delta_i=\mathbf{1}_{\{Y_i\le C_i\}}$ for $i=1,\cdots,n$ and $(X,Z)\in\mathbf{E}\times\mathbb{R}_+^*$.

If Z_i were uncensored it means that $Z_i=Y_i$ for all *i*. In this situation, [27] proposed a Hill's version of the conditional extreme value index when the covariate response is in $\mathbb{R}^p$. Following the same idea, we propose a functional Hill-type estimator depending on a semi-metric $d(.,.)$:

$$\hat{\gamma}^H_{Z_{n-k}}(x)=\frac{\sum_{i=1}^n K\left(h^{-1}d(x,X_i)\right)\left(\log(Z_i)-\log(y_n)\right)\mathbf{1}_{\{Z_i>y_n\}}}{\sum_{i=1}^n K\left(h^{-1}d(x,X_i)\right)\mathbf{1}_{\{Z_i>y_n\}}},\tag{5}$$

where $K(.)$ is a real-valued kernel function on $\mathbf{E}$, $h=h_n$ is a positive non-random bandwidth sequence such that $h\to 0$ as $n\to\infty$ and y_n is a local non-random threshold sequence for estimation with $y_n\to\infty$ as $n\to\infty$. Here, as stated in [27], a local threshold means a threshold depending on the point *x* in the covariate space where the estimation takes place, though the threshold is constant in a neighbourhood of *x*.

The estimator (5) is not consistent for $\gamma_1(x)$ if it is directly applied to the censored sample $(X_i,\delta_i,Z_i), i=1,\cdots,n$. Indeed, under appropriate regularity assumptions, estimator (5) will converge to the extreme-value index $\gamma(x)$ of the conditional distribution of *Z* given $X=x$. To accommodate censoring, we suggest, like in [7], to divide (5) by the proportion $\hat{p}_n(x)$ of uncensored observations among the $Z_i, i=1,\cdots,n$ that are larger than y_n, in a neighborhood of *x*:

$$\hat{p}_n(x)=\frac{\bar{H}^1_n(y_n\,|\,x)}{\bar{H}_n(y_n\,|\,x)}\tag{6}$$

where $\bar{H}_n(y_n\,|\,x)=\sum_{i=1}^n B_i(x)\mathbf{1}_{\{Z_i>y_n\}}$, $\bar{H}^1_n(y_n\,|\,x)=\sum_{i=1}^n B_i(x)\mathbf{1}_{\{Z_i>y_n,\delta_i=1\}}$ and

$B_i(x)$ are the well-known Nadaraya-Watson weights defined by

$$B_i(x)=\frac{K\left(h^{-1}d(x,X_i)\right)}{\sum_{j=1}^{n}K\left(h^{-1}d(x,X_j)\right)}. \tag{7}$$

The survival functions $\bar{H}_n(y_n \mid x)$ and $\bar{H}_n^1(y_n \mid x)$ can be rewritten as follow:

$\bar{H}_n^1(y_n \mid x)=\hat{\psi}_n(y_n,x)/\hat{g}_n(x)$ and $\bar{H}_n(y_n \mid x)=\hat{\zeta}_n(y_n,x)/\hat{g}_n(x)$ respectively, where

$$\hat{\psi}_n(y_n,x)=\frac{1}{n\left(\mu_x^{(1)}(h)\right)}\sum_{i=1}^{n}K\left(\frac{d(x,X_i)}{h}\right)\mathbf{1}_{\{Z_i>y_n,\delta_i=1\}};$$

$$\hat{\zeta}_n(y_n,x)=\frac{1}{n\left(\mu_x^{(1)}(h)\right)}\sum_{i=1}^{n}K\left(\frac{d(x,X_i)}{h}\right)\mathbf{1}_{\{Z_i>y_n\}}$$

and

$$\hat{g}_n(x)=\frac{1}{n\left(\mu_x^{(1)}(h)\right)}\sum_{i=1}^{n}K\left(\frac{d(x,X_i)}{h}\right).$$

Therefore we propose to estimate $\gamma_1(.)$ by

$$\hat{\gamma}_{Z_{n-k}}^{c,H}(x)=\frac{\hat{\gamma}_{Z_{n-k}}^{H}(x)}{\hat{p}_n(x)}. \tag{8}$$

This estimator depends on the bandwidth *h*, the threshold y_n and the semi-metric $d(\cdot,\cdot)$. The choice of the semi-metric is a crucial point in nonparametric functional data analysis (see [11]). Once the semi-metric has been chosen, packages are available in the literature (see https://cran.r-project.org/web/packages/fda.usc/index.html) to evaluate proximities between functional data. The semi-metric distance based on the derivative will be used to determine the distance between two curves X_1 and X_2. We consider the semi-metric:

$$d_q^{derive}(X_1,X_2)=\sqrt{\int\left(X_1^{(q)}(t)-X_2^{(q)}(t)\right)^2\mathrm{d}t}, \tag{9}$$

where q is the degree of derivative and where $X^{(q)}$ denotes the q^{th} derivative of *X*. In the following, second, third and fourth derivatives are considered. The impact of the degree of derivatives on the performance of our estimator we will be discussed when semi-metric based on derivatives are considered for smooth curves as covariates.

2.4. Estimation of Conditional Extreme Quantile

We now investigate the estimation of large conditional quantile $q(\alpha_n \mid x)$ of order $1-\alpha_n$ of $F(\cdot \mid x)$ for a variable *Y* given $X=x$ defined by $1-F\left(q(\alpha_n \mid x) \mid x\right)=\alpha_n$ with $\alpha_n \to 0$ as $n \to \infty$. To define our estimator, we have in the first step to define $\hat{q}_n^c(\alpha_n \mid x)$ the functional estimator of a large conditional quantile $q(\alpha_n \mid x)$ within the sample.

Let us consider the Kernel conditional Kaplan-Meier estimator of the conditional survival function $1-F(\cdot \mid x)$, for all $x \in \mathbf{E}$ and $y_n \in (0,\infty)$ defined as follows :

$$\hat{\bar{F}}_n(y_n \mid x) = \prod_{i=1}^{n} \left(1 - \frac{B_{ni}(x)\mathbf{1}_{\{Z_i > y_n, \delta_i = 1\}}}{1 - \sum_{j=1}^{n} B_{nj}(x)\mathbf{1}_{\{Z_j \le Z_i\}}}\right). \tag{10}$$

This function may be rewritten as

$$\hat{\bar{F}}_n(y_n \mid x) = \prod_{i=1}^{n} \left(1 - \frac{B_{ni}(x)}{\sum_{j=1}^{n} \mathbf{1}_{\{Z_j \ge Z_i\}} B_{nj}(x)}\right)^{\mathbf{1}_{\{Z_i > y_n, \delta_i = 1\}}} \quad \text{if } y_n \le Z_{(n)} \tag{11}$$

and zero otherwise where $Z_{(1)} \le \cdots \le Z_{(n)}$ denoted the order statistics of $Z_1, \cdots, Z_n$.

By taking into account the estimator in Equation (11), we propose to estimate conditional quantile $q(\alpha_n \mid x)$ within the sample of observation (*i.e.* for fixed $\alpha_n \in (0,1)$) as a generalized inverse of $\hat{\bar{F}}(\cdot \mid x)$ as

$$\hat{q}_n^c(\alpha_n \mid x) = \hat{\bar{F}}_n^{\leftarrow}(\alpha_n \mid x) = \inf\left\{u : \hat{\bar{F}}_n(u \mid x) \le \alpha_n\right\}, \tag{12}$$

where $\alpha_n \to 0$ as $n \to \infty$, we propose to estimate the conditional extreme quantile $q(\alpha_n \mid x)$ by Weissman-type estimator

$$\hat{q}_n^{c,W}(\alpha_n \mid x) = \hat{q}_n^c\left(\hat{\bar{F}}_n\left(Z_{(n-k)} \mid x\right)\right)\left(\frac{\hat{\bar{F}}_n\left(Z_{(n-k)} \mid x\right)}{\alpha_n}\right)^{\hat{\gamma}_{Z_{n-k}}^{c,H}(x)}. \tag{13}$$

The term $\left(\frac{\hat{\bar{F}}_n\left(Z_{(n-k)} \mid x\right)}{\alpha_n}\right)^{\hat{\gamma}_{Z_{n-k}}^{c,H}(x)}$ is an extrapolation factor allowing to estimate arbitrary large quantiles and $\hat{\gamma}_n^{c,H}(x)$ is the estimator of the censored functional conditional extreme value index $\gamma_1(x)$.

3. Results

In recurrent stroke patients, clinical outcomes were assessed using ambulatory blood pressure measurements from 297 patients to estimate conditional extreme values. This estimation considers that the time of occurrence of the recurrent stroke is randomly right-censored. We examined the distribution of time to recurrent strokes is whether they follow a heavy-tailed distribution. In statistics, a quantile-quantile Q-Q plot is a powerful tool to check whether the sample comes from a specific distribution. In EVT, the QQ plot is plotted against the standard exponential distribution to measure the heaviness of the tail of the distribution.

Besides, another tool to examine whether the sample comes from a specific distribution in extreme value theory is the sample mean excess function (MEF). The MEF is a sum of the excess over a threshold β divided by the number of data points that exceed the threshold β. A positive gradient above a certain threshold β of the empirical MEF, is a sign that the data has a heavy tailed distribution with a positive extreme value index $\gamma_1(x)$ as illustrated in **Figure 3**.

We therefore carry out our analysis of the conditional tail index quantile using the methodology described in [8]. The results, presented in **Table 1**, give an overview of the estimates of conditional extreme value index for different degrees of derivative for semi-metric distance. In addition, the confidence interval is provided using resampling techniques which reveal that the confidence interval becomes narrow as the degree of derivative increases. To get these empirical confidence intervals, we suggest a bootstrap methodology described as follows:

1) Draw $N = 500$ samples of the indexes of our dataset from $1, 2, \cdots, 297$ with replacement.

Table 1. Table of estimation result of $\gamma_1(x)$ and $q(\alpha_n \mid x)$ for the stroke data, $[\cdot]$ Bootstrap 95%-empirical confidence interval for $\gamma_1(x)$ and $q(\alpha_n \mid x)$, $(\cdot)$ empirical width of the confidence interval for $\alpha_n = 0.005$.

		4^{th} derivative	3^{rd} derivative	2^{nd} derivative
mean (x) − sd (x)	$\gamma_1(x)$	0.5170	0.5290	0.5093
	ACI	[0.1966, 0.8374]	[0.1766, 0.8813]	[0.0678, 0.8509]
	LCI	(0.6407)	(0.7047)	(0.7831)
	$q(0.005 \mid x)$	203.5339	217.25	211.1639
	ACI	[203.2135, 203.8543]	[216.8976, 217.6024]	[210.7223, 211.5054]
	LCI	(0.6407)	(0.7047)	(0.7831)
mean (x)	$\gamma_1(x)$	0.5026	0.5545	0.4997
	ACI	[0.1395, 0.8657]	[0.2492, 0.9898]	[0.0459, 0.8342]
	LCI	(0.7261)	(0.7406)	(0.7882)
	$q(0.005 \mid x)$	186.3794	213.3353	170.632
	ACI	[186.0163186.7425]	[213.0300, 213.7706]	[170.1878, 170.9761]
	LCI	(0.7261)	(0.7406)	(0.7882)
mean (x) + sd (x)	$\gamma_1(x)$	0.5186	0.5462	0.4743
	ACI	[0.2104, 0.8248]	[0.2148, 0.8777]	[0.1270, 0.8217]
	LCI	(0.6144)	(0.6628)	(0.6946)
	$q(0.005 \mid x)$	175.718	191.0137	163.3799
	ACI	[175.4108, 176.0252]	[190.6823, 191.3451]	[163.0326, 163.7272]
	LCI	(0.6144)	(0.6628)	(0.6946)

ACI presents the asymptotic confidence interval. LCI presents length of confidence interval.

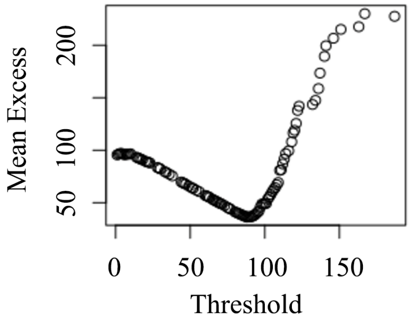

Figure 3. Meplot of time to event.

2) Generate $N = 500$ samples of $(Y_{i,1}, \delta_{i,1}, X_{i,1}), \cdots, (Y_{i,297}, \delta_{i,297}, X_{i,297})$ for corresponding indexes sampled in the first step.

3) Carry out on each of these N samples the estimation of the conditional extreme value index by $\hat{\gamma}^{c,H}_{Z_{(n-k^*)}}(x)$ using the procedure described in [9] (with the same (h^*, k^*)).

4) Also for each of these N samples, we work out the estimation of the conditional extreme quantile $\hat{q}^{c,W}_n(\alpha_n \mid x)$ which corresponding to each $\hat{\gamma}^{c,H}_{Z_{(n-k^*)}}(x)$ for the same (h^*, k^*)).

5) Take the interval bounded by the 2.5% and 97.5% quantile of the conditional extreme value index estimates as a confidence interval. Therefore, the average of low and upper bounds formed the 95%-level asymptotic empirical confidence interval presented in Table 1.

4. Discussion

In this paper, we address the estimation of the tail index and extreme quantiles of a heavy-tailed distribution when some functional covariate information is available and the data are randomly right-censored.

To the best of our knowledge, this is the first study about ischemic stroke and transient ischemic attack patients, with the main objective of studying a functional relationship between ambulatory blood pressure trajectories and clinical outcomes in stroke patients using a concept of conditional extreme value analysis.

To achieve our goal, we are interested in evaluating the conditional extreme quantile of Y the time to recurrent stroke in days given the ambulatory blood pressure trajectory as a functional covariate. In this paper, we assess the quantile $q(5/1000 \mid x)$ of order $1-5/1000$ of the conditional distribution of time to recurrent stroke Y given x, for x has the value $x = \text{mean} - \text{sd}(x_1, \cdots, x_n)$, $x = \text{mean}(x_1, \cdots, x_n)$ and $x = \text{mean} + \text{sd}(x_1, \cdots, x_n)$ with $\text{sd}(x_1, \cdots, x_n)$ denoted the empirical standard deviation of $x_1, \cdots, x_n$. For example, the estimated conditional extreme quantiles of the time to recurrent stroke were 203.5339, 186.3794 and 175.718 days at 95% of confidence interval [203.2135, 203.8543], [186.0163, 186.7425] and [175.4108, 176.0252] for lower, middle and higher the ambulatory blood pressure trajectories respectively at fourth derivative as described above.

Furthermore, the average bootstrap of low and upper bounds formed the 95%-level asymptotic empirical confidence interval for estimate $q(5/1000|x)$ and $\gamma_1(x)$ presented in Table 1 where the confidence interval becomes narrow as the degree of derivative increases.

As illustrated in Table 1 the stroke patients with higher blood pressure measurements had a higher risk of occurrence of recurrent stroke at early time. This result is not a surprise because the hypertension was an independent predictor of recurrent stroke according to the literature for more details [28].

5. Conclusions

We have explored the estimation of the functional Weissman kernel type estimator in the presence of a functional random covariate, valued in an infinite-dimensional space, alongside a right-censored scalar response variable. Our primary application revolves around discerning the potential impact of ambulatory blood pressure trajectories on the time of stroke recurrence.

Our findings suggest that higher blood pressure measurements significantly elevate the risk of stroke recurrence within a short time period, consistently observed across multiple quantiles of the time-to-recurrence distribution, as revealed by the estimated extreme quantiles.

The application of extreme value theory in the medical field, particularly those involving functional covariates, is still in its infancy. Nevertheless, various intriguing topics within this domain warrant further investigation in future research.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Elnady, H.M., Mohammed, G.F., Elhewag, H.K., Mohamed, M.K. and Borai, A. (2020) Risk Factors for Early and Late Recurrent Ischemic Strokes. *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, **56**, Article No. 56. https://doi.org/10.1186/s41983-020-00190-3

[2] Kariasa, I.M., Nurachmah, E., Koestoer, R.A., *et al.* (2019) Analysis of Participant's Characteristics and Risk Factors for Stroke Recurrence. *Enfermeria Clinica*, **29**, 286-290. https://doi.org/10.1016/j.enfcli.2019.04.035

[3] Lee, J.-D., Hu, Y.-H., Lee, M., Huang, Y.-C., Kuo, Y.-W. and Lee, T.-H. (2019) High Risk of One-Year Stroke Recurrence in Patients with Younger Age and Prior History of Ischemic Stroke. *Current Neurovascular Research*, **16**, 250-257. https://doi.org/10.2174/1567202616666190618164528

[4] Zheng, S.B. and Yao, B.D. (2019) Impact of Risk Factors for Recurrence after the First Ischemic Stroke in Adults: A Systematic Review and Meta-Analysis. *Journal of Clinical Neuroscience*, **60**, 24-30. https://doi.org/10.1016/j.jocn.2018.10.026

[5] Peng, L.M. (2021) Quantile Regression for Survival Data. *Annual Review of Statistics and Its Application*, **8**, 413-437. https://doi.org/10.1146/annurev-statistics-042720-020233

[6] Stupfler, G. (2016) Estimating the Conditional Extreme-Value Index under Random Right-Censoring. *Journal of Multivariate Analysis*, **144**, 1-24. https://doi.org/10.1016/j.jmva.2015.10.015

[7] Ndao, P., Diop, A. and Dupuy, J.-F. (2016) Nonparametric Estimation of the Conditional Extreme-Value Index with Random Covariates and Censoring. *Journal of Statistical Planning and Inference*, **168**, 20-37. https://doi.org/10.1016/j.jspi.2015.06.004

[8] Rutikanga, J.U. and Diop, A. (2021) Functional Kernel Estimation of the Conditional Extreme Quantile under Random Right Censoring. *Open Journal of Statistics*, **11**, 162-177. https://doi.org/10.4236/ojs.2021.111009

[9] Rutikanga, J.U. and Diop, A. (2021) Functional Kernel Estimation of the Conditional Extreme Value Index under Random Right Censoring. *Afrika Statistika*, **16**, 2647-2688. https://doi.org/10.16929/as/2021.2647.178

[10] Ferraty, F. and Vieu, P. (2003) Curves Discrimination: A Nonparametric Functional Approach. *Computational Statistics & Data Analysis*, **44**, 161-173. https://doi.org/10.1016/S0167-9473(03)00032-X

[11] Ferraty, F. and Vieu, P. (2006) Nonparametric Functional Data Analysis: Theory and Practice (Springer Series in Statistics). Springer-Verlag, Berlin.

[12] Fisher, R.A. and Tippett, L.H.C. (1928) Limiting Forms of the Frequency Distribution of the Largest or Smallest Member of a Sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, **24**, 180-190. https://doi.org/10.1017/S0305004100015681

[13] Gnedenko, B. (1943) Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, **44**, 423-453. https://doi.org/10.2307/1968974

[14] Beirlant, J., Guillou, A., Dierckx, G. and Fils Villetard, A. (2007) Estimation of the Extreme Value Index and Extreme Quantiles under Random Censoring. *Extremes*, **10**, 151-174. https://doi.org/10.1007/s10687-007-0039-x

[15] Einmahl, J.H.J., Fils-Villetard, A., Guillou, A., *et al.* (2008) Statistics of Extremes under Random Censoring. *Bernoulli*, **14**, 207-227. https://doi.org/10.3150/07-BEJ104

[16] Brahimi, B., Meraghni, D. and Necir, A. (2013) On the Asymptotic Normality of Hill's Estimator of the Tail Index under Random Censoring.

[17] Ivette Gomes, M. and Manuela Neves, M. (2010) A Note on Statistics of Extremes for Censoring Schemes on a Heavy Right Tail. 2010 *IEEE* 32*nd International Conference on Information Technology Interfaces* (*ITI*), Cavtat, 21-24 June 2010, 539-544.

[18] Cabras, S., Eugenia, C., *et al.* (2011) A Bayesian Approach for Estimating Extreme Quantiles under a Semiparametric Mixture Model. *ASTIN Bulletin*, **41**, 87-106.

[19] Coles, S.G. and Powell, E.A. (1996) Bayesian Methods in Extreme Value Modelling: A Review and New Developments. *International Statistical Review*, **64**, 119-136. https://doi.org/10.2307/1403426

[20] Stephenson, A. and Tawn, J. (2004) Bayesian Inference for Extremes: Accounting for the Three Extremal Types. *Extremes*, **7**, 291-307. https://doi.org/10.1007/s10687-004-3479-6

[21] Worms, J. and Worms, R. (2014) New Estimators of the Extreme Value Index under Random Right Censoring, for Heavy-Tailed Distributions. *Extremes*, **17**, 337-358. https://doi.org/10.1007/s10687-014-0189-6

[22] Ivette Gomes, M. and Manuela Neves, M. (2011) Estimation of the Extreme Value

Index for Randomly Censored Data. *Biometrical Letters*, **48**, 1-22.

[23] Matthys, G., Delafosse, E., Guillou, A. and Beirlant, J. (2004) Estimating Catastrophic Quantile Levels for Heavy-Tailed Distributions. *Insurance*: *Mathematics and Economics*, **34**, 517-537. https://doi.org/10.1016/j.insmatheco.2004.03.004

[24] Jiang, F., *et al*. (2020) Functional Censored Quantile Regression. *Journal of the American Statistical Association*, **115**, 931-944. https://doi.org/10.1080/01621459.2019.1602047

[25] Gardes, L. and Girard, S. (2012) Functional Kernel Estimators of Large Conditional Quantiles. *Electronic Journal of Statistics*, **6**, 1715-1744. https://doi.org/10.1214/12-EJS727

[26] Ndao, P. (2015) Modélisation de valeurs extrémes conditionnelles en présence de censure. PhD Thesis, Université Gaston Berger de Saint Louis, Saint Louis.

[27] Goegebeur, Y., Guillou, A. and Schorgen, A. (2014) Nonparametric Regression Estimation of Conditional Tails: The Random Covariate Case. *Statistics*, **48**, 732-755. https://doi.org/10.1080/02331888.2013.800064

[28] Han, J., Mao, W.J., Ni, J.X., *et al*. (2020) Rate and Determinants of Recurrence at 1 Year and 5 Years after Stroke in a Low-Income Population in Rural China. *Frontiers in Neurology*, **11**, Article No. 2. https://doi.org/10.3389/fneur.2020.00002