



Parametric and non-parametric estimation of extreme earthquake event: the joint tail inference for mainshocks and aftershocks

Juan-Juan Cai¹ · Phyllis Wan² · Gamze Ozel³

Received: 15 June 2019 / Revised: 5 November 2020 / Accepted: 9 November 2020 /
Published online: 23 November 2020
© The Author(s) 2020

Abstract

In an earthquake event, the combination of a strong mainshock and damaging aftershocks is often the cause of severe structural damages and/or high death tolls. The objective of this paper is to provide estimation for the probability of such extreme events where the mainshock and the largest aftershocks exceed certain thresholds. Two approaches are illustrated and compared – a parametric approach based on previously observed stochastic laws in earthquake data, and a non-parametric approach based on bivariate extreme value theory. We analyze the earthquake data from the North Anatolian Fault Zone (NAFZ) in Turkey during 1965–2018 and show that the two approaches provide unifying results.

Keywords Bivariate extreme value theory · Earthquake data · Tail probability · Mainshock · Aftershock

Mathematics Subject Classification (2010) 62G32 (60G70; 86A17)

✉ Phyllis Wan
wan@ese.eur.nl

Juan-Juan Cai
j.cai@vu.nl

Gamze Ozel
gamzeozl@hacettepe.edu.tr

¹ Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081HV, Amsterdam, the Netherlands

² Econometric Institute, Erasmus University Rotterdam, Burg. Oudlaan 50, 3062PA, Rotterdam, the Netherlands

³ Department of Statistics, Hacettepe University, 06800, Ankara, Turkey

1 Introduction

In a seismically active area, a strong earthquake, namely the *mainshock*, is often followed by subsequent damaging earthquakes, known as the *aftershocks*. These aftershocks may occur in numerous quantity and with magnitudes equivalent to powerful earthquakes on their own. For instance, in the 1999 İzmit earthquake, a magnitude 7.6 mainshock triggered hundreds of aftershocks with magnitudes greater than or equal to 4 in the first six days, cf. Polat et al. (2002). In the 2008 Sichuan earthquake, a mainshock of magnitude 8.0 induced a series of aftershocks with magnitudes up to 6.0. The results are severe structural damage and loss of life, especially when the area has already been weakened by the mainshock. The İzmit earthquake killed over 17,000 people and left half a million homeless (Marza 1999). The Sichuan earthquake caused over 69,000 deaths and damages of over 150 billion US dollars (Cui et al. 2011).

The goal of this paper is to provide a statistical analysis for the joint event of an extreme mainshock and extreme aftershocks. Throughout the paper, we denote the magnitude of a mainshock with X and that of the largest aftershock with Y . We estimate via two approaches the probability of

$$\mathbb{P}(X > x, Y > y), \quad (1)$$

for large values of x and y . The first approach uses a parametric model based on a series of well-known stochastic laws that describe the empirical relationships of the aftershocks and the mainshock, which we briefly review in Section 3.1. In the second approach, we apply bivariate extreme value theory to estimate the joint tail. Both methods are applied to extreme earthquake events in the North Anatolian Fault Zone (NAFZ) in Turkey, the region where the 1999 İzmit earthquake occurred.

The remainder of the paper is structured as follows. In Section 2, we present the earthquake data in NAFZ and describe the relevant data processing. Section 3 provides the parametric and non-parametric estimation procedures for the joint main-/after-shock distribution. The detailed data analysis and results are presented in Section 4. We conclude in Section 5 with some discussions. The processed data and codes of implementation in this paper can be found at the author's website.¹

2 Data description

We use the North Anatolian Fault Zone (NAFZ) as an area of investigation due to its long and extensive historical record of large earthquakes (Ambraseys 1970; Ambraseys and Finkel 1987). Extending from eastern Turkey to Greece, the 1,500-kilometer-long rip sustained several cycle-like sequences of large-magnitude ($M > 7$) earthquakes over the past centuries (Stein et al. 1997), several resulting in high

¹<https://sites.google.com/view/pwan>

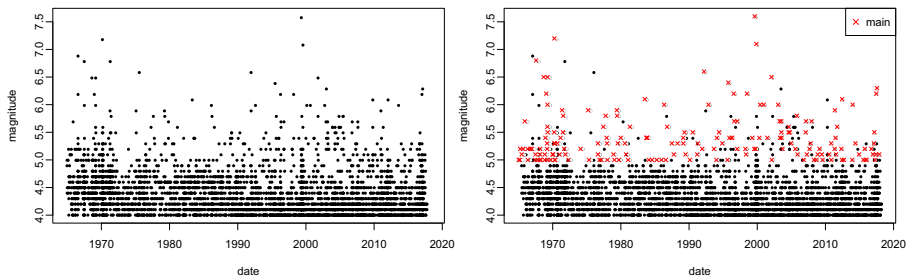


Fig. 1 Shocks and labelled mainshocks in the NAFZ during 1965–2018

death tolls and severe economic losses. The most recent activities include the İzmit (Mw 7.6) and Düzce (Mw 7.1) earthquakes of 1999 (Parsons et al. 2000; Reilinger et al. 2000).²

We obtain data from the Presidential of Earthquake Department database of the Turkish Disaster and Emergency Management Authority (<https://deprem.afad.gov.tr/?lang=en>) and consider all earthquake records between 1965–2018 with magnitudes 4 or higher in the area of latitudes $36.00^{\circ}N - 42.00^{\circ}N$ and longitudes $26.00^{\circ}E - 40.00^{\circ}E$. The left panel of Fig. 1 shows the time series of all earthquakes from 1965 onward.

We now label earthquake events by identifying the mainshocks and their corresponding aftershocks. Being interested in extreme events, we only consider earthquake events with significant mainshocks such that $X \geq 5$. We use the window algorithm proposed in Gardner and Knopoff (1974) as follows. For each shock with magnitude $X \geq 5$, we scan the window within distance $L(X)$ and time $T(X)$. If a larger shock exists, we move on to that shock and perform the same scan. If not, then the shock is labelled as the mainshock and all shocks within the specified window are pronounced as its aftershocks. Table 1 provides the values for $L(X)$ and $T(X)$. For example, for an earthquake of magnitude 6.0, any shock following $T = 510$ days and within $L = 54$ km radius, with a magnitude less than 6, is considered to be its aftershock.

The right panel of Fig. 1 shows the labelled mainshocks in the time series. The algorithm identifies $n = 180$ earthquake events with mainshocks $X \geq 5$ among which 129 have aftershocks with magnitude greater than 4. Note that a few large earthquakes in the early years are not labelled as mainshocks, instead they are identified as aftershocks of mainshocks before the year 1965 from earlier data.

²Regarding the earthquake scale in our data: Before 1977, all earthquakes were recorded in the body-wave magnitude scale (mb) or the surface-wave magnitude scale (Ms), depending on the depth of the earthquake. Following the development of the moment magnitude scale (Mw) by Kanamori (1977) and Hanks and Kanamori (1979), earthquakes with magnitude larger than 5.0 are recorded in the Mw scale whereas smaller earthquakes were still generally measure in the mb or the Ms scale. From 2012 on, all earthquakes are recorded in the Mw scale.

Table 1 Window specification $L(X)$, $T(X)$ for aftershock labelling from Gardner and Knopoff (1974)

X	L (km)	T (days)
5.0–5.4	40	155
5.5–5.9	47	290
6.0–6.4	54	510
6.5–6.9	61	790
7.0–7.4	70	915
7.5–7.9	81	960
8.0–8.4	94	985

3 Methodology

3.1 Parametric approach

It is agreed in the literature that the distribution of aftershocks in space, time and magnitude can be characterized by stochastic laws, see Utsu (1970), Utsu (1971), and Utsu (1972) for a summary with detailed empirical studies. In this section, we propose a simple parametric model for the joint magnitudes of the mainshock and the largest aftershock based on the expert knowledge from these studies.

We assume that each earthquake event consists of a mainshock and a sequence of aftershocks. The following pieces of empirical evidence for mainshocks and aftershocks are noted in prior literature.

1. Each earthquake event can be considered as independent from each other. Therefore the magnitudes of mainshocks can be modelled as independent realizations from a random variable X . By the Gutenberg-Richter's law (Gutenberg and Richter 1944), X can be modelled by an exponential distribution with distribution and density functions

$$\mathbb{P}(X > x) = e^{-\alpha x}, \quad f_X(x) = \alpha e^{-\alpha x}. \quad (2)$$

2. For an earthquake event, let $g(t)$ be the frequency of aftershocks at time t per time unit. Then from the modified Omori's law (Utsu 1970), regardless of the mainshock magnitude, $g(t)$ follows

$$g(t) = \frac{K}{(t + c)^p},$$

where K , c , p are constants.

3. For each earthquake event, the magnitude of the aftershocks follows also Gutenberg-Richter's law (Utsu 1972) such that the number of aftershocks $N(m)$ with magnitude greater or equal to m follows

$$N(m) \propto 10^{-bm}, \quad (3)$$

where b is some constant.

4. The magnitude difference between a mainshock and its largest aftershock is approximately constant, independent of the mainshock magnitude and typically between 1.1 and 1.2. This is known as the Bath's law (Bath 1965).

Based on the above empirical rules, we adopt the framework in Utsu (1970) and for each earthquake event with mainshock magnitude m_0 , model the intensity rate of aftershocks with magnitude m or greater as proportional to

$$\frac{10^{a+b(m_0-m)}}{(t+c)^p}, \quad (4)$$

where a, b, c, p are constants. This modelling is widely used in ensuing literature, cf. Reasenber and Jones (1989), and is the basis of the ETAS (epidemic-type aftershock sequence) simulation model, cf. Ogata (1988). As for the occurrences of the aftershocks, it is common to model them using a Poisson point process. Since by definition of our data, the aftershocks must have magnitude $m \leq m_0$, we define the intensity rate of aftershocks with magnitude m or greater to be

$$\lambda(t, m) = \frac{10^{a+b(m_0-m)} - 10^a}{(t+c)^p}, \quad (5)$$

such that aftershocks with magnitude $m \geq m_0$ have intensity rate 0.

3.1.1 The model

Given any earthquake event, let X be the magnitude of its mainshock and Y be the magnitude of the largest aftershock. We assume that the aftershocks sequence follows a non-homogeneous Poisson process with intensity function (5). Let N_m denote the number of aftershocks with magnitude m or greater. Then N_m follows a Poisson distribution with mean

$$\mathbb{E}[N_m|X = m_0] = \sum_{t=1}^{\infty} \lambda(t, m) := C(e^{\beta(m_0-m)} - 1), \quad m \leq m_0,$$

where $\beta = b \ln 10$ and $C = \frac{1}{\beta} 10^a \sum_{t=1}^{\infty} \frac{1}{(t+c)^p}$.

In order to investigate in the distribution of the largest aftershock, we are interested in the event $\{Y < m\}$, which is equivalent to $\{N_m = 0\}$. Hence, for $m \leq m_0$,

$$\mathbb{P}(Y < m|X = m_0) = \mathbb{P}(N_m = 0|X = m_0) = \exp \left\{ -C \left(e^{\beta(m_0-m)} - 1 \right) \right\}.$$

Let $Z := X - Y$, then Z has distribution function

$$\mathbb{P}(Z < z|X = m_0) = 1 - \exp \left\{ -C \left(e^{\beta z} - 1 \right) \right\} =: F_Z(z), \quad z \geq 0. \quad (6)$$

If we model the earthquake magnitudes on a continuous scale and hence Z as a continuous variable, this suggests that Z follows a Gompertz distribution with parameter set (C, β) , that is, $-Z$ follows a Gumbel distribution conditioned to be negative. By definition, $Z < m_0$, since the largest aftershock cannot be of non-positive magnitude. However, if we impose the convention that $\{Z \geq m_0\}$ represents the event that no aftershock occurs, then we can model Z as independent of X . Note that when m_0 is large the probability of $\{Z > m_0\}$ is negligible.

Combining (2) and (6) yields the joint model for (X, Y) given by

$$\mathbb{P}(X > x, Y > y) = \mathbb{P}(X > x, Z < X - y) = \int_x^\infty f_X(u) \int_0^{u-y} f_Z(z) dz du, \quad (7)$$

where $f_X(x)$ is as defined in Eq. 2 and $f_Z(z)$ is the density function of Z from Eq. 6. Given data, the parameters (α, β, C) can be estimated through maximum likelihood.

Note that by modelling the magnitude difference between the mainshock and the largest aftershock to be an independent variable, this model is in agreement with the empirically observed Båth's Law as mentioned above. It is also in line with the results in Vere-Jones et al. (2006) which were derived from a simple model where aftershock magnitudes are assumed to be independent and identical.

3.2 Bivariate extreme value approach

In this subsection, we describe an alternative nonparametric approach to estimate (1), the tail probability of interest, based on multivariate extreme value theory.

Multivariate extreme statistics has been exhibited to be a powerful tool for inference on multidimensional risk factors. Depending on the context, different types of multivariate extreme events have been studied in the extreme value literature. In the application to the offshore structure safety, de Haan and de Ronde (1998) and Coles and Tawn (1994) both estimated the probability of an event defined by $\{aX + Y > v\}$, where a and v are problem-specified constants. For monitoring airline performance, Einmahl et al. (2009) studied the estimation of $\mathbb{P}(X > s \text{ or } Y > t)$ where s and t are both large and subjected to the constraint that $\mathbb{P}(X > s) = c\mathbb{P}(Y > t)$ for a positive c . For environmental data, Cai et al. (2013) showed a case study on estimating $\mathbb{P}(X > s, Y > t)$ for stationary sequences of X and Y using approaches based on conditional extreme value theory (Heffernan and Tawn 2004) and hidden regular variation (Ledford and Tawn 1996; Resnick 2002). A recent work closely related to our study is Cooley et al. (2019), which presented a method for isolines of equal joint exceedance probability, that is the level curves we showed in Section 4.3.

We assume that the joint distribution of (X, Y) is in the max domain of a bivariate extreme distribution introduced in de Haan and Resnick (1977). This is a most common condition in multivariate tail analysis and includes distributions with various types of copulas. Let F_1 and F_2 denote the marginal distribution functions of X and Y , respectively. The assumption implies that (cf. Corollary 6.1.4 in de Haan and Ferreira (2006)) for any $(x, y) \in [0, \infty]^2 \setminus (\infty, \infty)$, the following limit exists:

$$\lim_{t \rightarrow 0} \frac{1}{t} \mathbb{P}(1 - F_1(X) < tx, 1 - F_2(Y) < ty) =: R(x, y). \quad (8)$$

The function R characterizes the extremal dependence between X and Y and it can be expressed via other extremal dependence measures. For instance, it is linked to the stable tail dependence function L and the Pickands function A :

$$R(x, y) = x + y - L(x, y) = (x + y) \left(1 - A \left(\frac{y}{x + y} \right) \right). \quad (9)$$

For a general review on the multivariate extreme value theory, see for example Chapter 6 in de Haan and Ferreira (2006) and Chapter 8 in Beirlant et al. (2004).

The limit relation in Eq. 8 guarantees the regularity in the right tail of the copula of (X, Y) , which enables us to do the bivariate extrapolation to the range far beyond the historical observations. Let s and t be sufficiently large and denote that $p_1 = 1 - F_1(s)$ and $p_2 = 1 - F_2(t)$. Then, we have

$$\begin{aligned}\mathbb{P}(X > s, Y > t) &= \mathbb{P}(1 - F_1(X) < p_1, 1 - F_2(Y) < p_2) \\ &= p_2 \cdot \frac{1}{p_2} \mathbb{P}\left(1 - F_1(X) < p_2 \cdot \frac{p_1}{p_2}, 1 - F_2(Y) < p_2\right) \\ &\approx p_2 R\left(\frac{p_1}{p_2}, 1\right).\end{aligned}\quad (10)$$

In this paper, we assume that $R(1, 1) > 0$, that is, X and Y are asymptotic dependent. For estimating the bivariate tail probability for an asymptotic independence pair, one can consult (Draisma et al. 2004) and (Cooley et al. 2019).

From Eq. 10, the problem transforms to estimate p_1 , p_2 and $R(x, 1)$. Due to the relation in Eq. 9, the various methods of estimating L or A can be applied to estimate $R(x, 1)$; for instance see Capéraà et al. (1997), Einmahl et al. (2008), Bücher et al. (2011), Fougères et al. (2015), and Beirlant et al. (2016) among many others. Because of the particular features of earthquake data – that they have been rounded to the first decimal and censored below – we use a basic non-parametric estimator of $R(x, 1)$, which requires minimal assumptions on the data and is the basis of other more advanced estimation approaches. Let n be the sample size and $k = k(n)$ be a sequence of integers such that $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$. Let R_i^X and R_i^Y denote the ranks of X_i and Y_i in their respective samples. The estimator of $R(x, 1)$ is given by

$$\hat{R}(x, 1) = \frac{1}{k} \sum_{i=1}^n I(R_i^X > n + 1/2 - kx, R_i^Y > n + 1/2 - k). \quad (11)$$

The asymptotic normality of \hat{R} is studied in Huang (1992); see also Drees and Huang (1998).

As for estimating p_1 and p_2 , we fit exponential distributions to both margins, which is a typical choice for modeling earthquake magnitude justified by the Gutenberg–Richter law (Gutenberg and Richter 1944). A natural alternative is to apply univariate extreme value theory to estimate these tail probabilities. Many studies have been devoted to study the tail distribution or the endpoint of earthquake magnitude; see for instance (Kijko 2004) and (Beirlant et al. 2018). However, due to the small sample size and the rounding issue, we choose to fit parametric margins. Let \hat{p}_i denote the estimator of p_i , for $i = 1, 2$. Then, the final estimator of $p_{st} := \mathbb{P}(X > s, Y > t)$ is given by

$$\hat{p}_{st} := \hat{p}_2 \hat{R}\left(\frac{\hat{p}_1}{\hat{p}_2}, 1\right). \quad (12)$$

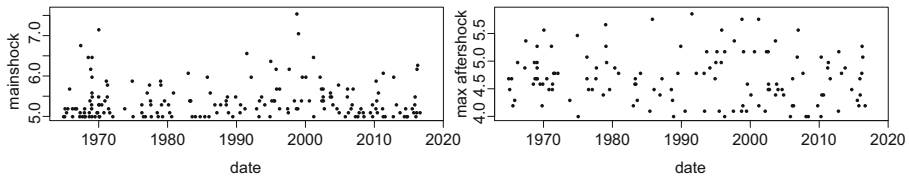


Fig. 2 Time series of mainshocks (left) and largest aftershocks (right)

4 Results

In Section 2 we extract from the NAFZ dataset time series of mainshocks magnitude (x_i) where $x_i \geq 5$. For the time series of the corresponding largest aftershock, we only observe the values that are above 4, that is, we observe $(y_i \mathbf{1}_{\{y_i \geq 4\}})$. The two time series are plotted in Fig. 2 and their joint scattered plot can be seen by the black dots in Fig. 4.

4.1 Parametric approach

In this subsection, we describe the results from the parametric model. The mainshock sequence (x_i) is fitted with an exponential distribution truncated at 4.95 – we take into consideration the continuity correction. Since all observations are discrete by 0.1 increment, from now on whenever we show the fit of a distribution or calculate the goodness-of-fit p -value, we jitter all observations by uniform noises between $(-.05, .05)$. The fit is shown on the left panel of Fig. 3 and the Kolmogorov-Smirnov p -value is 0.83, indicating a good fit.

Next we fit a Gompertz distribution to the difference $x_i - y_i$ by maximizing the following censored likelihood:

$$L(\beta, C | x_i, y_i) = \prod_{y_i \geq 4} f_Z(x_i - y_i; \beta, C) \prod_{y_i < 4} (1 - F_Z(x_i - 4; \beta, C)),$$

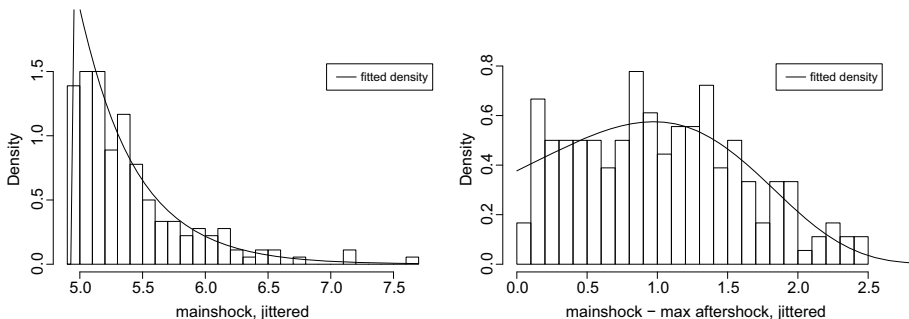


Fig. 3 The histogram and fitted curve of: i) mainshocks with exponential distribution (left); ii) differences between mainshocks and largest aftershocks with Gompertz distribution (right)

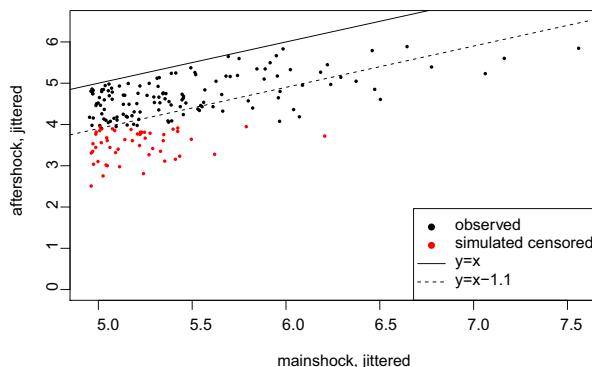


Fig. 4 Scatterplot of mainshocks vs. aftershocks. Black indicates observations and red indicates censored aftershocks with magnitude < 4 and simulated from the fitted model. The solid line indicate $y = x$. Clearly all observations falls below as $Y \leq X$ by definition. The dotted line indicates $y = x - 1.1$ as suggested by the Båth's law

where F_Z is as defined in Eq. 6 and f_Z is the corresponding density. To assess the goodness-of-fit, we first approximate the complete set of maximum aftershock sequence by \tilde{y}_i as follows . When $y_i \geq 4$, set $\tilde{y}_i := y_i$. When $y_i < 4$, simulate z_i from F conditional on $z_i \geq x_i - 4$ and set $\tilde{y}_i := x_i - z_i$. The histogram of jittered \tilde{y}_i is shown on the right panel of Fig. 3 with the fitted density. The Kolmogorov-Smirnov p -value is 0.95.

The scatterplot of the jittered (x_i, \tilde{y}_i) is plotted in Fig. 4, with the red point indicating the simulation for the censored observations. As we can see, the simulation is in agreement with the pattern of the observed pairs. We also note that the Båth's law (Båth 1965) – the empirical evidence that the magnitude difference between the mainshock and the largest aftershock is constant between 1.1 and 1.2 – can be well-justified by the fitted model. The fitted mean of $Y - X$ is 1.1. The dotted line in Fig. 4 corresponds to $x = y + 1.1$.

In Table 2, we display the fitted parameter values for (α, β, C) and their 95% confidence interval approximated through parametric bootstraps. Note the β is referred to as the b -value of aftershocks in seismology and measures the relative likelihood of large shocks to small shocks. This value generally centers around 1 but varies with regions and tectonic settings (Reasenber and Jones 1989).

Table 2 Fitted values of parameters (α, β, C) and their 95% confidence interval approximated through parametric bootstraps

	estimate	95% C.I.
α	2.22	(1.95,2.58)
β	1.11	(0.72,1.53)
C	0.34	(0.18,0.68)

4.2 Extreme value approach

For the extreme value analysis approach, we use the same estimate for the marginal distribution of X as in the parametric approach. As for the marginal distribution of Y , we assume that $(Y - \mu | Y > \mu) \sim \text{Exp}(\lambda)$. For a chosen μ , we obtain the maximum likelihood estimator $\hat{\lambda}$. Then, for $t > \mu$, the estimator of $\mathbb{P}(Y > t)$ is given by

$$\hat{p}_2 = \int_t^\infty \hat{\lambda} \exp\{-\hat{\lambda}(y - \mu)\} dy \cdot \frac{1}{n} \sum_{i=1}^n I(Y_i > \mu).$$

For this data set, we choose $\mu = 4.55$ and obtain $\hat{\lambda} = 2.2$. The fitted density is shown in Fig. 5 with the Kolmogorov-Smirnov p -value 0.11.

When estimating $R(x, 1)$, we note that there are ties in the data as they are rounded to one decimal place. For this, we randomly assign ranks to the tied observations. The data on aftershock magnitude includes only observations that are larger than or equal to 4. In other words, the missing values of Y are all below 4. Let n_1 denotes the number of observed values of Y . The ranks of all the missing values are all less than $n - n_1$, therefore the corresponding indicator functions in Eq. 11 equal to zero as far as we choose $k < n_1$. In this way, the missing values don't effect the estimator in Eq. 11. Note that in our data set, $n = 180$ and $n_1 = 129$.

The left panel of Fig. 6 shows the estimates of $R(x, 1)$ for three different values of x and $k \in [10, 100]$. Also note that $R(1, 1)$ is a commonly used quantity to distinguish asymptotic dependence ($R(1, 1) > 0$) and asymptotic independence ($R(1, 1) = 0$). Roughly, asymptotic dependence says that the extremes of X and Y tend to occur simultaneously while joint extremes rarely occur under asymptotic independence. This plot clearly suggests tail dependence between X and Y because the estimates of $R(1, 1)$ are clearly above zero. Based on these three curves, we choose $k = 40$ because the estimates indicated with three curves are all rather stable for k round and slightly larger than 40.

With the choice of $k = 40$, we obtain the non-parametric estimate of $R(x, 1)$ for $x \in [0.02, 5]$ plotted in the black curve in right panel of Fig. 6. The wiggly behaviour of this estimator motivates us to consider a smoothing method. We adopt

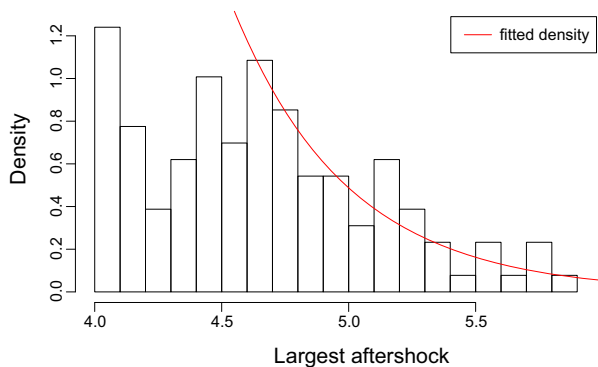


Fig. 5 The histogram of the largest aftershocks and fitted exponential density

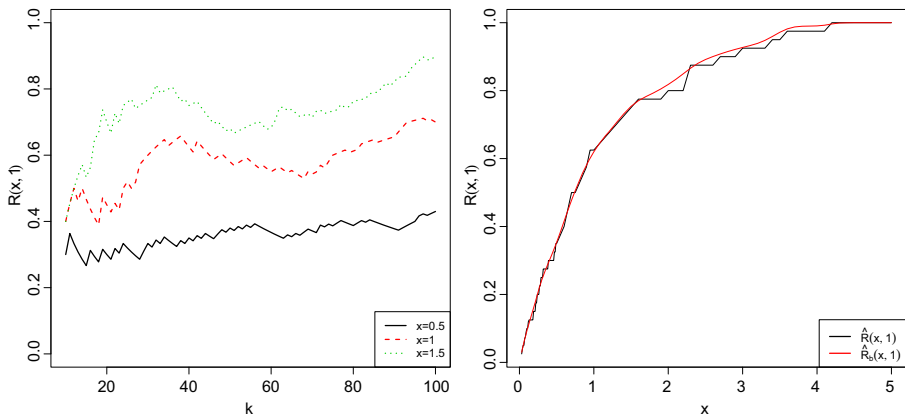


Fig. 6 (Left panel) The choices of k v.s. the corresponding estimates $\hat{R}(x, 1)$ given in Eq. 11 for $x = 0.5, 1, 1.5$. (Right panel) The estimates $\hat{R}(x, 1)$ given in Eq. 11 and the beta smoothed $\hat{R}_b(x, 1)$ from Kiriliouk et al. (2018), both with $k = 40$

the smoothing method introduced in Kiriliouk et al. (2018), which makes use of the beta copula. This smoothed estimator, denoted as $\hat{R}_b(x, 1)$, respects the pointwise upper bounds of the function, that is $R(x, 1) \leq \max(x, 1)$ and it does not require smoothing parameter such as bandwidth. The resulted estimates are represented by the red curve in right panel of Fig. 6. The two estimators are coherent with each other. $\hat{R}_b(x, 1)$ is only used in obtaining the level curves in Fig. 7.

4.3 Results and comparison

We are ready to estimate probabilities for the joint tail of (X, Y) . First, we estimate the tail probability³ defined in Eq. 10 for the ten largest earthquakes (mainshocks) in the NAFZ since 1965. As shown in the fifth and sixth column of Table 3, the estimates by two approaches are surprisingly close to each other, which supports the reasonability of the results. We emphasize that the two approaches only share one common assumption, that is, the marginal distribution of X . The distribution of Y and the dependence between X and Y are modelled separately.

Next we obtain the level curves of (X, Y) for the tail probability sequence

$$(10^{-3}, 5 \cdot 10^{-4}, 10^{-4}, 5 \cdot 10^{-5}, 10^{-5}, 5 \cdot 10^{-6}, 10^{-6}),$$

as shown in Fig. 7. The points (x, y) on each curve represents such that $\mathbb{P}(X > x, Y > y) = p$ for the given probability level. Albeit based on different theories, the two approaches provide coinciding prediction results. The two dot lines in Fig. 7 correspond to $x = y$ and $x = y + 1.1$. The horizontal shape of the curves between these two lines indicates that the results respect the B  th's law. This is particularly

³We remark that as our data set consists of only mainshocks with magnitude ($X \geq 5$), the probability in this section has to be interpreted as a conditional probability that given a significant mainshock $X \geq 5$ occurs.

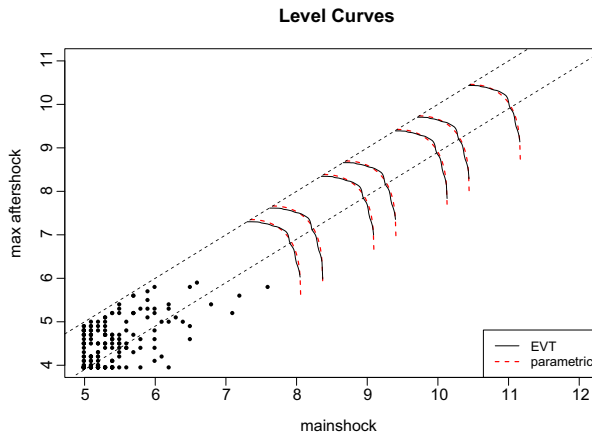


Fig. 7 Predicted level curve for (X, Y) based on the parametric model (red dotted) and extreme value analysis (black solid), with existing observations

remarkable for the non-parametric approach, which does not impose any dependence structure for (X, Y) .

The surprisingly close results by two different approaches might be explained by the following two points. First, the derived parametric model is a good fit for this earthquake data set (that is why it gives very similar result to the non-parametric method). Second, the parametric model of (X, Y) defined in Eq. 7 implies that X and Y are asymptotic dependence because $Y = X - Z$, where Z has a lighter tail than X . The parametric model falls in the framework of our non-parametric approach.

Table 3 Tail probability estimation for the ten largest earthquakes in the NAFZ since 1965

	date	mainshock	largest aftershock	parametric probability	non-parametric probability	location
1	1999-08-17	7.6	5.8	0.00265	0.00257	İzmit
2	1970-03-28	7.2	5.6	0.00618	0.00580	Gediz
3	1999-11-12	7.1	5.2	0.00815	0.00805	Düzce
4	1967-07-22	6.8	5.4	0.01413	0.01356	Mudurnu
5	1992-03-13	6.6	5.9	0.01429	0.01464	Erzincan
6	2002-02-03	6.5	5.8	0.01785	0.01827	Afyon
7	1969-03-28	6.5	4.9	0.02927	0.02745	Alaşehir
8	1968-09-03	6.5	4.6	0.03092	0.03051	Bartın
9	1995-10-01	6.4	5.0	0.03437	0.03296	Dinar
10	2017-07-20	6.3	5.1	0.03938	0.03747	Mugla

4.4 Uncertainty quantification via a simulation study

We carry out a small simulation study to investigate if the extreme value approach does give a coherent result for data generated from the parametric model, and to quantify the uncertainties of the level curves obtained by the extreme value approach for this type of data sets. We simulate data from the model in Eq. 7 with parameters $(\alpha, \beta, C) = (2.22, 1.11, 0.34)$, which are the parameter estimates based on our real data set, and we consider the same censoring for Y . Let (X_i, \tilde{Y}_i) be a pair of random observations from Eq. 7, $i = 1, \dots, n$. Then $Y_i = \tilde{Y}_i$, if $\tilde{Y}_i \geq 4$; Y_i is otherwise set to a missing value.

We first consider $n = 200$, a sample size similar to that of our real data set. The tuning parameters are chose the same as in Section 4.2: $\mu = 4.55$ and $k = 40$. The upper left panel in Fig. 8 shows estimated level curves for $p = 10^{-3}$ and

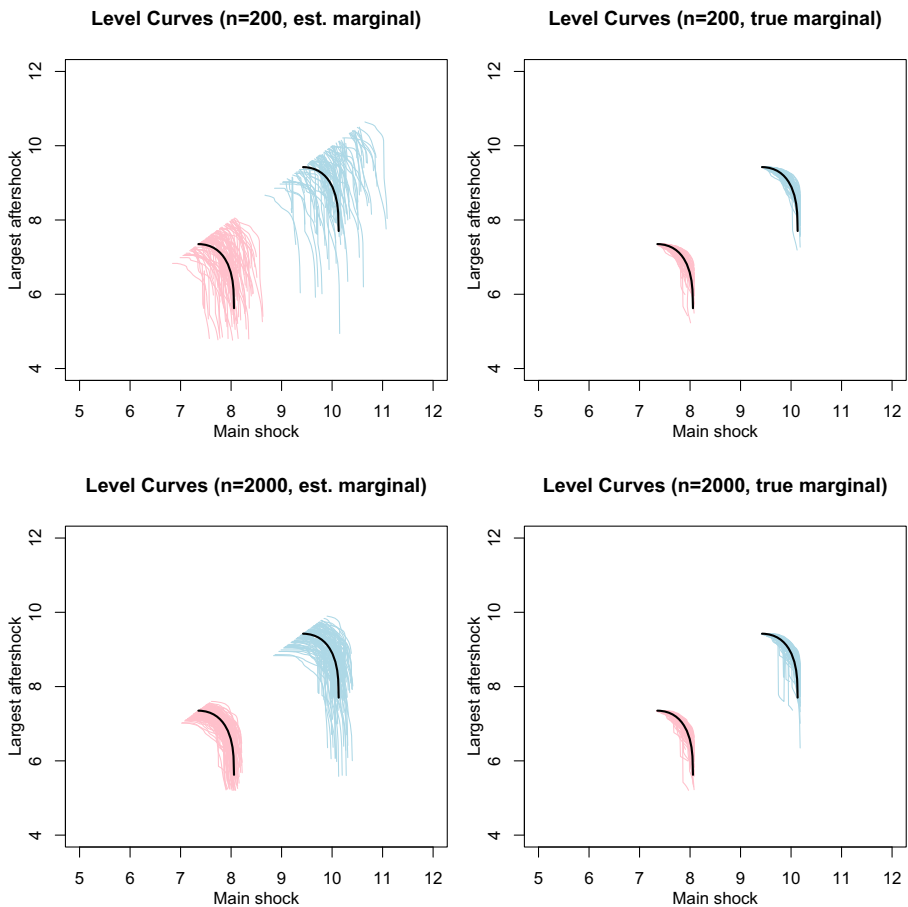


Fig. 8 Level curves for $p = 10^{-3}$ (pink) and $p = 10^{-5}$ (light blue): the left panel uses the estimator given in Eq. 12 with estimated marginal exceedance probabilities and right panel uses the estimator given in Eq. 13 with true values of p_1 and p_2 . Sample size $n = 200$ (upper panels) and $n = 2000$ (lower panels)

$p = 10^{-5}$ based on 100 random samples. The theoretical level curves indicated by the black lines do lie in the center of the estimated level curves, however the spread of the level curves are large. The level curves are estimated using Eq. 12. The uncertainty consists of two parts: the marginal distributions estimations namely \hat{p}_1 and \hat{p}_2 , and the estimation of extremal dependence, $\hat{R}(x, 1)$. We further consider a pseudo estimator of p_{st} by using the true value of p_1 and p_2 :

$$\tilde{p}_{st} := p_2 \hat{R}\left(\frac{p_1}{p_2}, 1\right). \quad (13)$$

The uncertainty of the estimated level curve based on \tilde{p}_{st} is largely reduced, as shown in the upper right panel of Fig. 8.

Further, we increase the sample size to 2000. For this sample size, we choose $\mu = 5$ and $k = 50$. The resulting level curves are displayed in the lower panels of Fig. 8. As the sample size increases, the uncertainty for the estimated level curve decreases, especially for the case with estimated marginal distribution.

Note that the level curves stay below the line $y = x$ because by definition, with probability one, $X > Y$, that is, there is zero probability mass for the event $(X < Y)$. In theory, the true level curve, from left to right, starts from the point (y_p, y_p) and ends at $(x_p, 0)$ ⁴, where y_p and x_p are such that $\mathbb{P}(Y > y_p) = \mathbb{P}(X > x_p) = p$. Therefore, the starting and end positions of a level curve depend only on the marginal distributions of X and Y . However, the shape of the level curves depends on the extremal dependence of X and Y . This explains why using the true marginal distribution leads to a much smaller spread of level curves, compared to the ones with estimated margins.

5 Discussion

In this paper we consider estimating the tail probability of an extreme earthquake event where the mainshock magnitude X and the largest aftershock magnitude Y both exceed certain thresholds. We approach the problems from two directions. On one hand, based on the well-known stochastic rules for aftershocks, we propose a joint parametric model for (X, Y) , estimate the model using (censored) maximum likelihood, and from the model, calculate the desired probabilities. On the other hand, we use non-parametric methods from bivariate extreme value analysis to extrapolate tail probabilities. We illustrates both methods using the earthquake data in the North Anatolian Fault Zone (NAFZ) in Turkey from 1965 to 2018. The two approaches produce surprisingly agreeing results.

This is an exploratory effort in applying multivariate extreme value analysis to seismology problems and much extension is possible. For example, the occurrences of the earthquake events can be modelled in time and return level of extreme events can be estimated. Further information, such as distance between shocks and other

⁴The plotted level curves stop earlier at (x_p, \tilde{y}_p) , due to the fact that $\mathbb{P}(X > x_p, Y \leq \tilde{y}_p)$ is much smaller than p .

geological covariates, can be incorporated into the analysis to provide more accurate or customized results.

This paper serves as a confirmation that simple techniques from multivariate extreme value analysis, though with little expert knowledge behind the data, is able to provide useful information in the analysis of extreme events.

Acknowledgments The authors would like to thank Anna Kiriliouk for a discussion on estimators of the stable tail dependence function and the two referees for helpful comments.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ambraseys, N.N.: Some characteristic features of the Anatolian fault zone. *Tectonophysics* **9**(2-3), 143–165 (1970)
- Ambraseys, N.N., Finkel, C.F.: Seismicity of Turkey and neighbouring regions, 1899-1915. In: *Annales geophysicae Series B. Terrestrial and Planetary Physics*, vol. 5, pp. 701–725 (1987)
- Båth, M.: Lateral inhomogeneities of the upper mantle. *Tectonophysics* **2**(6), 483–514 (1965)
- Beirlant, J., Escobar-Bach, M., Goegebeur, Y., Guillou, A.: Bias-corrected estimation of stable tail dependence function. *J. Multivar. Anal.* **143**, 453–466 (2016)
- Beirlant, J., Goegebeur, Y., Segers, J., Teugels, J.: *Statistics of extremes theory and applications*. Wiley, Chichester (2004)
- Beirlant, J., Kijko, A., Reynkens, T., Einmahl, J.H.J.: Estimating the maximum possible earthquake magnitude using extreme value methodology: the Groningen case. *Natural Hazards* 1–23 (2018)
- Bücher, A., Dette, H., Volgushev, S.: New estimators of the pickands dependence function and a test for extreme-value dependence. *Ann. Stat.* **39**(4), 1963–2006 (2011)
- Cai, J.J., Fougères, A.L., Mercadier, C.: Environmental data: multivariate extreme value theory in practice. *Journal de la Société Française de Statistique & revue de statistique appliquée* **154**(2), 178–199 (2013)
- Capéraà, P., Fougères, A.-L., Genest, C.: A nonparametric estimation procedure for bivariate extreme value copulas. *Biometrika* **84**(3), 567–577 (1997)
- Coles, S.G., Tawn, J.A.: Statistical methods for multivariate extremes: an application to structural design. *J R Stat Soc Ser C (Appl Stat)* **43**(1), 1–31 (1994)
- Cooley, D., Thibaud, E., Castillo, F., Wehner, M.F.: A nonparametric method for producing isolines of bivariate exceedance probabilities. *Extremes* **22**(3), 373–390 (2019)
- Cui, P., Chen, X.-Q., Zhu, Y.-Y., Su, F.-H., Wei, F.-Q., Han, Y.-S., Liu, H.-J., Zhuang, J.-Q.: The Wenchuan earthquake (May 12, 2008), Sichuan province, China, and resulting geohazards. *Nat. Hazards* **56**(1), 19–36 (2011)
- de Haan, L., de Ronde, J.: Sea and wind: multivariate extremes at work. *Extremes* **1**(1), 7 (1998)
- de Haan, L., Ferreira, A.: *Extreme value theory: an introduction*. Springer, Berlin (2006)
- de Haan, L., Resnick, S.I.: Limit theory for multivariate sample extremes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **40**(4), 317–337 (1977)
- Draisma, G., Drees, H., Ferreira, A., De Haan, L.: Bivariate tail estimation: dependence in asymptotic independence. *Bernoulli* **10**(2), 251–280 (2004)

- Drees, H., Huang, X.: Best attainable rates of convergence for estimators of the stable tail dependence function. *J. Multivar. Anal.* **64**(1), 25–46 (1998)
- Einmahl, J.H.J., Krajina, A., Segers, J.: A method of moments estimator of tail dependence. *Bernoulli* **14**(4), 1003–1026 (2008)
- Einmahl, J.H.J., Li, J., Liu, R.Y.: Thresholding events of extreme in simultaneous monitoring of multiple risks. *J. Am. Stat. Assoc.* **104**(487), 982–992 (2009)
- Fougères, A.-L., De Haan, L., Mercadier, C.: Bias correction in multivariate extremes. *Ann. Stat.* **43**(2), 903–934 (2015)
- Gardner, J.K., Knopoff, L.: Is the sequence of earthquakes in Southern California, with aftershocks removed Poissonian? *Bull. Seismol. Soc. Am.* **64**(5), 1363–1367 (1974)
- Gutenberg, B., Richter, C.F.: Frequency of earthquakes in California. *Bull. Seismol. Soc. Am.* **34**(4), 185–188 (1944)
- Hanks, T.C., Kanamori, H.: A moment magnitude scale. *J. Geophys. Res.: Solid Earth* **84**(B5), 2348–2350 (1979)
- Heffernan, J.E., Tawn, J.A.: A conditional approach for multivariate extreme values. *J. R. Stat. Soc. Series B Stat. (Methodol.)* **66**(3), 497–546 (2004)
- Huang, X.: Statistics of bivariate extreme values. Thesis Publishers, Amsterdam (1992)
- Kanamori, H.: The energy release in great earthquakes. *J. Geophys. Res.* **82**(20), 2981–2987 (1977)
- Kijko, A.: Estimation of the maximum earthquake magnitude, m_{\max} . *Pure Appl. Geophys.* **161**(8), 1655–1681 (2004)
- Kiriliouk, A., Segers, J., Tafakori, L.: An estimator of the stable tail dependence function based on the empirical beta copula. *Extremes* **21**(4), 581–600 (2018)
- Ledford, A.W., Tawn, J.A.: Statistics for near independence in multivariate extreme values. *Biometrika* **83**(1), 169–187 (1996)
- Marza, V.I.: On the death toll of the Izmit (Turkey) major earthquake. ESC General Assembly Papers, European Seismological Commission, Potsdam, 2004 (1999)
- Ogata, Y.: Statistical models for earthquake occurrences and residual analysis for point processes. *J. Am. Stat. Assoc.* **83**(401), 9–27 (1988)
- Parsons, T., Toda, S., Stein, R.S., Barka, A., Dieterich, J.H.: Heightened odds of large earthquakes near Istanbul: an interaction-based probability calculation. *Science* **288**(5466), 661–665 (2000)
- Polat, O., Eyidogan, H., Haessler, H., Cisternas, A., Philip, H.: Analysis and interpretation of the aftershock sequence of the August 17, 1999, Izmit (Turkey) earthquake. *J. Seismol.* **6**(3), 287–306 (2002)
- Reasenber, P.A., Jones, L.M.: Earthquake hazard after a mainshock in California. *Science* **243**(4895), 1173–1176 (1989)
- Reilinger, R.E., Ergintav, S., Bürgmann, R., McClusky, S., Lenk, O., Barka, A., Gurkan, O., Hearn, L., Feigl, K.L., Cakmak, R.: Coseismic and postseismic fault slip for the 17 August 1999, $M=7.5$, Izmit, Turkey earthquake. *Science* **289**(5484), 1519–1524 (2000)
- Resnick, S.: Hidden regular variation, second order regular variation and asymptotic independence. *Extremes* **5**(4), 303–336 (2002)
- Stein, R.S., Barka, A., Dieterich, J.H.: Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering. *Geophys. J. Int.* **128**(3), 594–604 (1997)
- Utsu, T.: Aftershocks and earthquake statistics (1): Some parameters which characterize an aftershock sequence and their interrelations. *J. Fac. Sci., Hokkaido University. Series 7 Geophysics* **3**(3), 129–195 (1970)
- Utsu, T.: Aftershocks and earthquake statistics (2): Further investigation of aftershocks and other earthquake sequences based on a new classification of earthquake sequences. *J. Fac. Sci., Hokkaido University. Series 7 Geophysics* **3**(4), 197–266 (1971)
- Utsu, T.: Aftershocks and earthquake statistics (3): Analyses of the distribution of earthquakes in magnitude, time and space with special consideration to clustering characteristics of earthquake occurrence (1). *J. Fac. Sci., Hokkaido University Series 7 Geophysics* **3**(5), 379–441 (1972)
- Vere-Jones, D., Murakami, J., Christophersen, A.: A further note on Bath's law. In: *The 4th International Workshop on Statistical Seismology (Statsei4)*, pp. 611–0011 (2006)