



INVESTIGATION OF THE DEPENDENCE STRUCTURE IN SEISMIC HAZARD ANALYSIS: AN APPLICATION FOR TURKEY

SERPİL ÜNAL KARAÇAM

ABSTRACT. In this study, using the earthquake occurrence data (Richter magnitude is equal to 4 or greater than 4 in the years 1901-2014) of the areas limited by $39.5^\circ - 42^\circ$ N latitudes and $26^\circ - 45^\circ$ E longitudes of North Anatolia and $36^\circ - 39.5^\circ$ N latitudes and $26^\circ - 31^\circ$ E longitudes of West Anatolia, it is aimed to model the dependence structure of Semi-Markov model via conditional copulas in which the copula is parametric and its parameter varies as the covariate, based on the assumption that the successive earthquakes in the same structural discontinuity should not be independent events and the occurrence of the earthquakes should be influenced by the elapsed time between them. From the results obtained for these regions with high seismicity, it is seen that the variation in the strength of dependence between the time elapsed from the previous seismic event and the magnitude of the next seismic event at different magnitudes of previous seismic event is highly significant and a usage of the parametric linear form in the copula parameter will be adequately characterized.

1. INTRODUCTION

Turkey takes part on the Alpine-Himalayan (Mediterranean) seismic belt, one of the important seismic belts of the world and its tectonic framework shows abundant evidence of past and continuing mobility. The tectonic inequilibrium of Turkey is reflected by numerous active faults. The north Anatolian fault (NAF) is the most important and active of these. The NAF was reactivated on 17 August ($M_w = 7.4$) and 12 November ($M_w = 7.2$) 1999 with two destructive earthquakes in the eastern Marmara region as a result of the westward migrating large earthquake series in the 20th century. West Anatolia also takes part among the significant mechanism of active tectonics in Turkey. The area was exposed to continuous earthquakes due

Received by the editors: December 19, 2017; Accepted: September 06, 2018.

Key words and phrases. Unconditional and conditional copulas, semi-Markov model, seismic gap theory, seismic hazard, seismicity of North and West Anatolia.

2010 Mathematics Subject Classifications: 62H20, 60K15, 86A15.

to its highly mixed tectonic appearance and it also creates a region with a high potential for earthquakes in the future.

Nowadays it is accepted that it is impossible either to know where the earthquakes occur or when they occur, their magnitudes, and to prevent these devastating natural events. However, the statistical studies existing in the fields of geophysical, geological and earthquake engineering show that the parameters of possible earthquakes and the severity of ground motions they created can only be estimated probabilistically. In other words, since earthquakes demonstrate randomness according to parameters and there are various uncertainties (such as some deficiencies in the earthquake records), seismic hazard estimation with probabilistic methods is seen as the most appropriate method. Semi-markov model, one of the most commonly used probabilistic model, is based on the assumption that although the earthquakes are dependent on the space dimension, they are independent from the unit-time on the time dimension. According to the model, the magnitude of an earthquake depends on the magnitude of the previous earthquake and the time interval between them. This may indicate that a long period seismic quiescence may end with an earthquake of large magnitude [6, 7, 8, 33].

Long-term or short-term prediction of the earthquakes has been a real challenge in science for decades and several papers have appeared for and/or against the ability to do so. The contributed papers vary with respect to the time horizon of their prediction, the magnitude range that can be predicted, the area covered and of course the method upon which prediction is based. Parallel to this, there are also some theories that try to provide information towards the prediction challenge. One, rather controversial, such theory is the so-called 'seismic gap' theory that relates the time elapsed from the previous seismic event to the magnitude of the next seismic event. According to this theory, there is a positive relation, which implies that an area without seismicity for a time period increases its probability to have a major earthquake. If such a relationship exists then by modeling these two parameters we can predict (in statistical sense) the size of a future earthquake [27].

Whenever we are dealing with the issue of modeling dependencies among random variables, copula models come into play. Copulas are functions which link multivariate distribution functions to their one-dimensional marginal distribution functions. Their use in several scientific fields has a long history dating back to, in economics [13]; in survival analysis [14, 15, 20, 29, 34]; in finance [10, 12]; in insurance [18, 24] and in geology [25, 27]. Although copulas have been used in the applied statistical literature for long years, the covariate adjustment for copulas has been considered only recently. In [31], Patton proposed the concept of conditional copula in which the correlation is affected by covariate(s). To our knowledge, so far conditional copula have been applied only in finance [9, 23, 30, 31, 32], in survival analysis [2, 3, 4, 19].

In the present paper, based on the seismic gap theory, we aim to estimate the seismic hazard by modeling the dependence structure of Semi-Markov model with conditional copula, based on the assumption that the successive earthquakes in the same structural discontinuity should not be independent events and the occurrence of the earthquakes should be influenced by the elapsed time between them.

2. UNCONDITIONAL AND CONDITIONAL COPULAS

A copula is a function which joins or couples a multivariate distribution function to its one-dimensional marginal distribution functions. More formally, the following definition can be given.

Definition 1. *A two-dimensional copula is a mapping $C : [0, 1]^2 \rightarrow [0, 1]$ satisfying the conditions:*

1) For $\forall u, v \in [0, 1]$,

$$C(u, 0) = C(0, v) = 0 \text{ and } C(u, 1) = u, C(1, v) = v$$

2) For $\forall u_1, u_2, v_1, v_2 \in [0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

The informal and formal definitions are connected by the following theorem which also elucidates the role that copulas play in the relationship between multivariate distribution functions and their univariate margins.

Theorem 2. *(Sklar's theorem) Let H be a joint distribution function with margins F_1 and F_2 . Then there exists a copula C such that for all x_1, x_2 in \overline{IR} ,*

$$H(x_1, x_2) = C(F_1(x_1), F_2(x_2)). \quad (1)$$

If F_1 and F_2 are continuous, then C is unique; otherwise, C is uniquely determined on $\text{Ran}F_1 \times \text{Ran}F_2$. Conversely, if C is a copula and F_1 and F_2 are distribution functions, then the function H defined by (1) is a joint distribution function with margins F_1 and F_2 [26].

Table 1. Some of one parametric families of copulas

Family	Distribution function ($C_\theta(u, v)$)	Range of θ
Clayton	$(u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}}$	$(0, \infty)$
Frank	$-\theta^{-1} \ln \left(\frac{[(e^{-\theta} - 1) + (e^{-\theta u} - 1)(e^{-\theta v} - 1)]}{(e^{-\theta} - 1)} \right)$	$(-\infty, \infty) - \{0\}$
Gumbel	$\exp \left(- \left((-\ln u)^\theta + (-\ln v)^\theta \right)^{\frac{1}{\theta}} \right)$	$[1, \infty)$
Çelebioğlu-Cuadras	$uv \exp(\theta(1-u)(1-v))$	$[-1, 1]$
FGM	$uv + \theta uv(1-u)(1-v)$	$[-1, 1]$
Galambos	$uv \exp \left\{ \left[(-\ln u)^{-\theta} + (-\ln v)^{-\theta} \right]^{\frac{-1}{\theta}} \right\}$	$[0, \infty)$
Tawn	$uv \exp \left(-\theta \frac{\ln u \ln v}{\ln(uv)} \right)$	$(0, 1)$
Joe	$1 - \left[\frac{(1-u)^\theta + (1-v)^\theta}{-(1-u)^\theta(1-v)^\theta} \right]^{\frac{1}{\theta}}$	$[1, \infty)$

2.1. Estimation (Inference functions for margins method). The log-likelihood function is given by

$$l(\hat{\vartheta}) = \sum_{t=1}^T \ln c(F_1(x_{1t}; \vartheta_1), F_2(x_{2t}; \vartheta_2); \theta) + \sum_{t=1}^T \sum_{i=1}^2 \ln f_i(x_{it}; \vartheta_i). \tag{2}$$

In the first stage of the method, the parameters of the marginal distributions F_i are estimated and in the second stage, the copula parameters conditioned to the previous estimates of marginal distributions are estimated. In each stage the maximum likelihood method is used [30].

2.2. Goodness of fit test for copulas. Given that we have followed IFM approach to estimate the parameters of a set of copulas, if the number of estimated parameters is the same across all maximum likelihood functions and in the estimation process the same data is used for each model specification, an obvious criterion is to compare the maximized value of the log likelihood function $l(\hat{\vartheta})$ [17].

2.3. Model selection. In [5], Allcroft and Glasbey proposed a bootstrap method for testing the hypothesis, for $k = 1, \dots, m$,

$$H_0 : C = C_k \text{ versus } H_1 : C \neq C_k. \tag{3}$$

Since the joint distributions of the log-likelihoods are approximately multivariate normal based on the central limit theorem, Mahalanobis squared distance is appropriate for making the above comparisons. This distance between the vector of log-likelihoods λ at the original data and the vector of average log-likelihoods $\bar{\lambda}_k$ at the simulated data from the k th copula is obtained as, for $k = 1, \dots, m$,

$$D_k^2 = (\lambda - \bar{\lambda}_k) S^{-1} (\lambda - \bar{\lambda}_k), \quad (4)$$

where S is the sample covariance matrix for these log-likelihoods. Based on the normality of the log-likelihoods, the quantity D_k^2 follows an $mF_{m, B-1}$, distribution under the null hypotheses that k th copula distribution is correct [27, 28].

When the covariate is added to the model in the issue of modeling dependence between two or more random variables, you need to model the conditional copulas whose copula parameter varies according to the values of a measured covariate.

Definition 3. *The conditional copula of $(Y_1, Y_2) | X = x$, where $Y_1 | X = x \approx F_{1|X}(\cdot|x)$ and $Y_2 | X = x \approx F_{2|X}(\cdot|x)$, is the joint distribution function of $U_1 \equiv F_{1|X}(\cdot|x)$ and $U_2 \equiv F_{2|X}(\cdot|x)$ given $X = x$.*

Proposition 4. *A two-dimensional conditional copula is a function $C : [0, 1] \times [0, 1] \times \chi \rightarrow [0, 1]$ with the following properties:*

1) For $\forall u, v \in [0, 1]$ and $\forall x \in \chi$,

$$C(u, 0|x) = C(0, v|x) = 0 \text{ and } C(u, 1|x) = u, \quad C(1, v|x) = v$$

2) For $\forall u_1, u_2, v_1, v_2 \in [0, 1]$ such that $u_1 \leq u_2$ and $v_1 \leq v_2$ and $\forall x \in \chi$,

$$C(u_2, v_2|x) - C(u_1, v_2|x) - C(u_2, v_1|x) + C(u_1, v_1|x) \geq 0.$$

Theorem 5. *Let $F_{1|X}(\cdot|x)$ be the conditional distribution of $Y_1 | X = x$, $F_{2|X}(\cdot|x)$ be the conditional distribution of $Y_2 | X = x$, $H_X(\cdot, \cdot|x)$ be the joint distribution of $(Y_1, Y_2) | X = x$ and χ be the support of X . Then, there exists a unique conditional copula $C(\cdot, \cdot|x)$, whenever $F_{1|X}(\cdot|x)$ and $F_{2|X}(\cdot|x)$ are continuous in y_1 and y_2 , for all $x \in \chi$, such that*

$$H_X(y_1, y_2|x) = C(F_{1|X}(y_1|x), F_{2|X}(y_2|x) | x). \quad (5)$$

Conversely, if we let $F_{1|X}(\cdot|x)$ be the conditional distribution of $Y_1 | X = x$, $F_{2|X}(\cdot|x)$ be the conditional distribution of $Y_2 | X = x$ and $\{C(\cdot, \cdot|x)\}$ be a family of conditional copulas measurable in x , then the function $H_X(\cdot, \cdot|x)$ defined in (5) is a conditional bivariate distribution function with conditional marginal distributions $F_{1|X}(\cdot|x)$ and $F_{2|X}(\cdot|x)$ [31, 32].

2.4. Local copula-likelihood estimation. Since the main focus is on the dependence structure, it is assumed that the conditional marginal distributions $F_{1|X}(\cdot|x)$ and $F_{2|X}(\cdot|x)$ are known and consider the following model:

$$(U_{1i}, U_{2i}) | X_i \approx C(u_{1i}, u_{2i} | \theta(x)), \quad (6)$$

for $i = 1, 2, \dots, n$, where $\theta(x) = g^{-1}(\eta(x))$ and $g^{-1} : IR \rightarrow \Theta$ is the known inverse link function, which ensures that the copula parameter has the correct range, and η is the unknown calibration function to be estimated. As inverse link functions we use $g^{-1}(t) = e^t$, $\theta \in (0, \infty)$ for Clayton and Galambos families, $g^{-1}(t) = t$, $\theta \in (-\infty, \infty) - \{0\}$ for Frank family, $g^{-1}(t) = e^t + 1$, $\theta \in [1, \infty)$ for

Gumbel and Joe families, $g^{-1}(t) = \sin(t)$, $\theta \in [-1, 1]$ for Çelebioğlu-Cuadras and FGM families, $g^{-1}(t) = \arctan\left(\frac{t + (\frac{\pi}{2})}{\pi}\right)$, $\theta \in (0, 1)$ for Tawn family, respectively.

Since, for most copula families, the form of the calibration function $\eta(\cdot)$ characterizing the underlying dependence structure is difficult to discern by inspection, for a nonparametric approach in estimating the target function see [1, 2, 3].

2.5. Choice of bandwidth (Leave-one-out cross-validation). Choice of bandwidth parameter is an important issue in local estimation. A too small bandwidth parameter will yield an estimator with a smaller bias but a greater variance, and the calibration function will be undersmoothed. A too large bandwidth will produce a less variance but a larger bias, and an oversmoothed calibration function. For choosing of optimum bandwidth see [2, 3].

2.6. Conditional copula selection. Suppose we have a (finite) set of candidate families $C = \{C_q : q = 1, \dots, Q\}$ with copula parameter function $\theta_q(\cdot)$, $q = 1, \dots, Q$ from which we want to choose the one that best represents the data at hand. Since the estimation depends on the bandwidth parameter h , these estimates are denoted by $\hat{\theta}_{h_q}(X)$, $q = 1, \dots, Q$. Hence, from each family there is a candidate model and we face the task of choosing the family whose representative best fits the data. For selecting to copula family see [2, 3].

2.7. Generalized likelihood ratio test for copula functions. Suppose that $\{(U_{11}, U_{21}, X_1), \dots, (U_{1n}, U_{2n}, X_n)\}$ is a random sample from the conditional copula model (6). The null hypothesis of interest restricts the space of calibration functions to a subspace f that is fully specified parametrically, such as the set of all linear functions on χ . Then we are interested in testing

$$H_0 : \eta(\cdot) \in f \text{ versus } H_1 : \eta(\cdot) \notin f. \tag{7}$$

For testing a parametric null hypothesis versus a nonparametric alternative hypothesis, e.g. (7), in [16], Fan et al. developed exploration of the asymptotic distribution of the ratio test falls within the scope of the generalized likelihood ratio test (GLRT). According to the test, the log-likelihood function under the null hypothesis is given by

$$L_n(H_0, \tilde{\eta}) = \sum_{i=1}^n \ln c\{U_{1i}, U_{2i} | g^{-1}(\tilde{\eta}(X_i))\}, \tag{8}$$

and the alternative hypothesis is given by

$$L_n(H_1, \hat{\eta}_h) = \sum_{i=1}^n \ln c\{U_{1i}, U_{2i} | g^{-1}(\hat{\eta}_h(X_i))\}. \tag{9}$$

The difference between the two log-likelihoods allows us to evaluate the evidence in the data in favor of (or against) the null model. Hence, GLRT statistic is given by

$$\lambda_n(h) = L_n(H_1, \hat{\eta}_h) - L_n(H_0, \tilde{\eta}). \tag{10}$$

Since nonparametric maximum likelihood estimators are difficult to obtain and may not even exist, in [16], Fan et al. suggested using any reasonable nonparametric estimator under the alternative model. In particular, using a local polynomial estimator to specify the alternative model of a number of hypothesis testing problems, in [16], Fan et al. showed that the null distribution of the GLRT statistic follows asymptotically a chi-square distribution with the number of degrees of freedom independent of the nuisance parameters. Namely, let $|\chi|$ be the range of the covariate X and define

$$r_K = \frac{(K - \frac{1}{2}K * K)(0)}{\int (K - \frac{1}{2}K * K)^2(t) dt}, c_K = \left(K - \frac{1}{2}K * K\right)(0). \tag{11}$$

Then, $r_K \lambda_n(h) \cong \chi^2_{r_K c_K \frac{|\chi|}{h}}$ [2, 4]. Also, for simplicity, according to the Epanechnikov kernel, some values in identifying the degrees of freedom are given in the following table [37]:

Table 2. Kernel-Dependent Constants from the p th Degree Local Polynomial Fit

p	$K * K(0)$	$K(0)$	$(2K - K * K)(0)$	r_K
0,1	0.60	0.75	0.90	2.1153
2,3	1.25	1.4062	1.5625	1.9755
4,5	1.8930	2.0508	2.2085	1.9336

3. APPLICATION TO EARTHQUAKE DATA

In this study, it is investigated the following two regions of Turkey by the considerations of earthquake zones map in geographic information system, the seismic activity maps to Turkey, its vicinity in the Integrated Homogeneous Earthquake Catalog and Turkey’s fault lines (North Anatolia, Eastern Anatolia, Western Anatolia), as follows [35, 36]:



FIGURE 1. Separation of regions of Turkey.

Region 1, if latitude $\geq 39.5^\circ$; Region 2, if latitude $< 39.5^\circ$ and longitude $\leq 31^\circ$; and using the earthquake occurrence data (Richter magnitude is equal to 4 or greater than 4 in the years 1901-2014) of these regions, it is aimed to model the dependence structure of Semi-Markov model via conditional copulas in which the copula is parametric and its parameter varies as the covariate, based on the assumption that the successive earthquakes in the same structural discontinuity should not be independent events and the occurrence of the earthquakes should be influenced by the elapsed time between them.

In the modeling process with conditional copula, firstly, conditional marginals are estimated by the following equations:

$$P(Y_{i-1} \leq y | X_{t_{i-1}} = x) = \frac{\partial C(F(x), G(y))}{\partial F(x)}, \tag{12}$$

$$P(X_{t_i} \leq x_1 | X_{t_{i-1}} = x) = \frac{\partial C(F(x), F(x_1))}{\partial F(x)}, \tag{13}$$

where $X_{t_{i-1}}$, X_{t_i} and $Y_{i-1} = t_i - t_{i-1}$ are respectively the earthquake magnitude at time t_{i-1} , the earthquake magnitude at time t_i and the elapsed time between successive earthquakes.

The first step in specifying the conditional marginal distributions given by equations (12) and (13) is to find the marginal distributions of the elapsed time between successive earthquakes and the earthquake magnitude at time t_i . Accordingly, the distributions determined are as follows:

Table 3. The marginal distributions of the elapsed time between successive earthquakes and the earthquake magnitudes

Region 1	Region 2
$Y_{i-1} \approx Weibull \begin{pmatrix} \alpha = 0.3910 \\ \beta = 0.0396 \\ \gamma = 0 \end{pmatrix}$	$Y_{i-1} \approx Weibull \begin{pmatrix} \alpha = 0.4237 \\ \beta = 0.0232 \\ \gamma = 0 \end{pmatrix}$
$X_{t_i} \approx Gen - Pareto \begin{pmatrix} k = -0.2371 \\ \sigma = 0.7564 \\ \mu = 3.9296 \end{pmatrix}$	$X_{t_i} \approx Gen - Pareto \begin{pmatrix} k = -0.1455 \\ \sigma = 0.5741 \\ \mu = 3.9454 \end{pmatrix}$

The next and final step in specifying the conditional marginal distributions is the selection of the most appropriate copula which describe the dependence between variables $X_{t_{i-1}}$ and Y_{i-1} and $X_{t_{i-1}}$ and X_{t_i} . Accordingly, copula parameter estimation, the log-likelihood value $l(\hat{\vartheta})$ and simulation results regarding each region are as follows:

Table 4. Copula parameter estimation, the log-likelihood value and simulation results for Region 1

Copula	$\hat{\theta}(X_{t_{i-1}}, Y_{i-1})$	$l(\hat{\vartheta})(X_{t_{i-1}}, Y_{i-1})$	$\hat{\theta}(X_{t_{i-1}}, X_{t_i})$	$l(\hat{\vartheta})(X_{t_{i-1}}, X_{t_i})$
$C_{\theta}^{Clayton}$	0.0228	-123.240	0.8511	-1170.04
C_{θ}^{Frank}	0.2912	-121.998	2.9303	-1171.55
C_{θ}^{Gumbel}	1.0311	-122.007	1.2671	-1240.14
$C_{\theta}^{Celebioğlu}$	0.1330	-122.099	0.9990	-1183.40
C_{θ}^{FGM}	0.1390	-122.065	0.9990	-1189.90
$C_{\theta}^{Galambos}$	0.0009	-123.508	0.0009	-1308.78
C_{θ}^{Tawn}	0.1750	-116.974	0.6240	-1225.25
C_{θ}^{Joe}	1.8810	-330.186	2.0080	-1421.09

Regarding Region 1, goodness of fit measure hint towards the conclusion that the Tawn copula with $\hat{\theta} = 0.1750$ best represents the dependence structure between $X_{t_{i-1}}$ and Y_{i-1} ($p = 0.2768$). Also, the Clayton copula with $\hat{\theta} = 0.8511$ best represents the dependence structure between $X_{t_{i-1}}$ and X_{t_i} ($p = 0.0413$).

Table 5. Copula parameter estimation, the log-likelihood value and simulation results for Region 2

Copula	$\hat{\theta}(X_{t_{i-1}}, Y_{i-1})$	$l(\hat{\vartheta})(X_{t_{i-1}}, Y_{i-1})$	$\hat{\theta}(X_{t_{i-1}}, X_{t_i})$	$l(\hat{\vartheta})(X_{t_{i-1}}, X_{t_i})$
$C_{\theta}^{Clayton}$	1.4509E-6	1492.62	0.6674	-1036.31
C_{θ}^{Frank}	-0.2770	1494.64	2.5237	-1007.22
C_{θ}^{Gumbel}	1.0362	1498.33	1.3156	-1023.04
$C_{\theta}^{Celebioğlu}$	-0.1100	1494.21	0.9020	-1029.01
C_{θ}^{FGM}	-0.1220	1494.41	1.000	-1025.98
$C_{\theta}^{Galambos}$	0.0009	1492.62	0.0009	-1163.93
C_{θ}^{Tawn}	0.0920	1497.60	0.6400	-1017.04
C_{θ}^{Joe}	1.7810	1243.94	2.1350	-1221.94

Regarding Region 2, goodness of fit measures hint towards the conclusion that the Gumbel copula with $\hat{\theta} = 1.0362$ best represents the dependence structure between $X_{t_{i-1}}$ and Y_{i-1} ($p = 0.1296$). Also, the Frank copula with $\hat{\theta} = 2.5237$ best represents the dependence structure between $X_{t_{i-1}}$ and X_{t_i} ($p = 0.0828$).

After estimating conditional marginals, it is needed to estimate the calibration function in the local copula-likelihood method for estimating the functional relationship between the copula parameter and the covariate. For this purpose, firstly, in the local constant estimation ($p = 0$), under the considered copula families, the optimum bandwidths regarding each region are chosen as follows:

Table 6. The optimum bandwidths in the local constant estimation

Region	Clayton	Frank	Gumbel	Çelebioğlu	FGM	Galambos	Tawn	Joe
1	0.20	1.19	0.92	0.33	0.20	1.19	0.55	3.30
2	0.40	0.58	0.40	0.40	0.40	0.40	0.48	0.40

In the local copula-likelihood method, another stage is the selection of the most appropriate conditional copula regarding each region under the optimum bandwidths. Our copula selection method chooses the FGM family for Region 1, having the minimum cross-validated prediction error with value 209.4950, the Galambos family with value 312.5947 for Region 2.

For comparison, we also perform the global estimations regarding each region with constant forms to be $\tilde{a}_{01} = 0.449375$, $\tilde{a}_{02} = -1,164563$. Then, under the chosen copula, for each region, we perform the generalized likelihood ratio test to check whether the earthquake magnitude at time t_{i-1} has a significant effect on the strength of dependence. According to this, results regarding each region are as follows:

For Region 1, under FGM copula, using optimum bandwidth, we obtain the degrees of freedom of chi-square distribution as 15.71. The test statistic 35.197 yields a p-value to be 0.003735. Thus, we conclude that the effect of $X_{t_{i-1}}$, the earthquake magnitude at time t_{i-1} , on the strength of dependence between X_{t_i} , the earthquake magnitude at time t_i , and Y_{i-1} , the elapsed time between successive earthquakes, is statistically significant.

For Region 2, under Galambos copula, using optimum bandwidth, we obtain the degrees of freedom of chi-square distribution as 7.139. The test statistic 27.6269 yields a p-value to be 0.000257. Thus, we conclude that the effect of $X_{t_{i-1}}$, the earthquake magnitude at time t_{i-1} , on the strength of dependence between X_{t_i} , the earthquake magnitude at time t_i , and Y_{i-1} , the elapsed time between successive earthquakes, is statistically significant.

Since, for regions 1 and 2, we conclude that the effect of $X_{t_{i-1}}$ on the strength of dependence between X_{t_i} and Y_{i-1} is statistically significant, for this regions, under the chosen copulas, we perform the generalized likelihood ratio test to check whether the functional relationship between the covariate and the calibration function is linear when approached the calibration function with local linear estimates. According to this, in the local linear estimation ($p = 1$), the optimum bandwidths are chosen as follows:

Table 7. The optimum bandwidths in the local linear estimation

Region	Clayton	Frank	Gumbel	Çelebioğlu	FGM	Galambos	Tawn	Joe
1	1.98	2.56	1.54	0.43	0.33	1.19	0.55	1.19
2	0.58	2.50	0.83	0.40	0.48	0.48	1.20	1.00

Our copula selection method chooses the FGM family for Region 1, having the minimum cross-validated prediction error with value 209.1614, the Galambos family

with value 312.6570 for Region 2. For comparison, we also perform the parametric estimations regarding Region 1 and Region 2 with linear forms to be $\tilde{a}_{01} = 0.666436$ and $\tilde{a}_{11} = -0.047099$, $\tilde{a}_{02} = -2.676213$ and $\tilde{a}_{12} = 0.325290$, respectively. According to the generalized likelihood ratio test, obtained results regarding Region 1 and Region 2 are as follows:

For Region 1, under the FGM copula, using optimum bandwidth, we obtain the degrees of freedom of chi-square distribution as 9.4331. The test statistic 20.31946 yields a p-value to be 0.016043. Thus, we conclude that the linear effect of $X_{t_{i-1}}$, the earthquake magnitude at time t_{i-1} , on the strength of dependence between X_{t_i} , the earthquake magnitude at time t_i , and Y_{i-1} , the elapsed time between successive earthquakes, is statistically significant. So, for Region 1, the copula parameter is calculated to be $\hat{\theta}_1(x_{t_{i-1}}) = \sin(0.666436 - 0.047099x_{t_{i-1}})$.

For Region 2, under the Galambos copula, using optimum bandwidth, we obtain the degrees of freedom of chi-square distribution as 5.9442. The test statistic 16.81059 yields a p-value to be 0.010004. Thus, we conclude that the linear effect of $X_{t_{i-1}}$, the earthquake magnitude at time t_{i-1} , on the strength of dependence between X_{t_i} , the earthquake magnitude at time t_i , and Y_{i-1} , the elapsed time between successive earthquakes, is statistically significant. So, for Region 2, the copula parameter is calculated to be $\hat{\theta}_2(x_{t_{i-1}}) = \exp(-2.676213 + 0.325290x_{t_{i-1}})$.

According to the results obtained regarding regions 1 and 2, since the effect of $X_{t_{i-1}}$ on the strength of dependence between X_{t_i} and Y_{i-1} is statistically significant, the dependence structure of Semi-Markov model will be modeled with conditional copula. According to this, for Region 1, to be among a number of copula families the one that best represents dependence structure between $X_{t_{i-1}}$ and X_{t_i} is Clayton (unconditional) copula with $\hat{\theta} = 0.8511$ and $X_{t_i} \approx \text{Gen} - \text{Pareto}(k = -0.2371, \sigma = 0.7564, \mu = 3.9296)$ the conditional marginal distribution can be written as follows:

$$P(X_{t_i} \leq x_1 | X_{t_{i-1}} = x) = (F(x))^{-1.8511} A, \quad (14)$$

where $A = \left((F(x))^{-0.8511} + (F(x_1))^{-0.8511} - 1 \right)^{\left(\frac{-1}{0.8511} \right)^{-1}}$. Additionally, for Region 1, to be among a number of copula families the one that best represents dependence structure between $X_{t_{i-1}}$ and Y_{i-1} is Tawn (unconditional) copula with $\hat{\theta} = 0.175$, $X_{t_i} \approx \text{Gen} - \text{Pareto}(k = -0.2371, \sigma = 0.7564, \mu = 3.9296)$ and $Y_{i-1} \approx \text{Weibull}(\alpha = 0.3910, \beta = 0.0396, \gamma = 0)$, the conditional marginal distribution can be written as follows:

$$P(Y_{i-1} \leq y | X_{t_{i-1}} = x) = G(y) B [1 - 0.175 \ln(G(y)) M], \quad (15)$$

where $B = \exp \left[\frac{-0.175(\ln(F(x)))(\ln(G(y)))}{(\ln(F(x)G(y)))} \right]$ and $M = \left[\frac{(\ln(F(x)G(y)) - \ln(F(x)))}{(\ln(F(x)G(y)))^2} \right]$. To be the conditional marginal distributions given by the equations (14) and (15), $P(Y_{i-1} \leq y | X_{t_{i-1}} = x) = U_1$ and $P(X_{t_i} \leq x_1 | X_{t_{i-1}} = x) = U_2$ and among a number of copula families the one that best represents dependence structure between

X_{t_i} and Y_{i-1} is FGM (conditional) copula with $\hat{\theta}_1(x) = \sin(0.666436 - 0.047099x)$, the conditional joint distribution can be written as follows:

$$P\{X_{t_i} \leq x_1, Y_{i-1} \leq y | X_{t_{i-1}} = x\} = u_1 u_2 + \hat{\theta}_1(x) u_1 u_2 (1 - u_1) (1 - u_2). \quad (16)$$

Similarly, for Region 2, to be among a number of copula families the one that best represents dependence structure between $X_{t_{i-1}}$ and X_{t_i} is Frank (unconditional) copula with $\hat{\theta} = 2.5237$ and $X_{t_i} \approx Gen-Pareto(k = -0.1455, \sigma = 0.5741, \mu = 3.9454)$, the conditional marginal distribution can be written as follows:

$$P(X_{t_i} \leq x_1 | X_{t_{i-1}} = x) = \frac{N(E - 1)}{(\exp(-2.5237) - 1) + (N - 1)(E - 1)}, \quad (17)$$

where $N = \exp(-2.5237F(x))$ and $E = \exp(-2.5237F(x_1))$. Additionally, for Region 2, to be among a number of copula families the one that best represents dependence structure between $X_{t_{i-1}}$ and Y_{i-1} is Gumbel (unconditional) copula with $\hat{\theta} = 1.0362$, $X_{t_i} \approx Gen - Pareto(k = -0.1455, \sigma = 0.5741, \mu = 3.9454)$ and $Y_{i-1} \approx Weibull(\alpha = 0.4237, \beta = 0.0232, \gamma = 0)$, the conditional marginal distribution can be written as follows:

$$P(Y_{i-1} \leq y | X_{t_{i-1}} = x) = \exp\left(-J^{\frac{1}{1.0362}}\right) \left(\frac{(-\ln(F(x)))^{0.0362}}{F(x)}\right) \left(J^{\left(\frac{1}{1.0362}\right)-1}\right), \quad (18)$$

where $J = \left((-\ln(F(x)))^{1.0362} + (-\ln(G(y)))^{1.0362}\right)$. To be the conditional marginal distributions given by the equations (17) and (18), $P(Y_{i-1} \leq y | X_{t_{i-1}} = x) = U_1$ and $P(X_{t_i} \leq x_1 | X_{t_{i-1}} = x) = U_2$ and among a number of copula families the one that best represents dependence structure between X_{t_i} and Y_{i-1} is Galambos (conditional) copula with $\hat{\theta}_2(x) = \exp(-2.676213 + 0.325290x)$, the conditional joint distribution can be written as follows:

$$P\{X_{t_i} \leq x_1, Y_{i-1} \leq y | X_{t_{i-1}} = x\} = u_1 u_2 I, \quad (19)$$

where $I = \exp\left(\left((-\ln(u_1))^{-\hat{\theta}_2(x)} + (-\ln(u_2))^{-\hat{\theta}_2(x)}\right)^{\left(\frac{-1}{\hat{\theta}_2(x)}\right)}\right)$.

As we have already mentioned, it is accepted that it is impossible either to know where the earthquakes occur or when they occur, their magnitudes. However, the equations obtained above show that the parameters of possible earthquakes and the severity of ground motions they created can be estimated probabilistically.

4. RESULT

Turkey is a country in which earthquake hazard is extremely high in terms of geological, historical and instrumental earthquake activities since it takes part on the Alpine-Himalayan (Mediterranean) seismic belt, one of the important seismic belts of the world. In two destructive earthquakes Marmara, August 17, 1999 and Düzce, November 12, 1999, thousands of people died and tens of thousands were

wounded, hundreds of thousands of buildings were destroyed. The experiences we had in the past indicate that we will face with these type destructive earthquakes in the future. For this purpose, with some statistical analysis and predictions done in this study, we tried to show that the casualties and damage occurred in the results of the earthquakes in Turkey, an earthquake zone, could be prevented to some extent.

In this study on the basis of seismic gap theory, it is aimed to model the dependence structure of Semi-Markov model via conditional copulas. According to results obtained, for regions 1 and 2 with high seismicity, we conclude that the variation in the strength of dependence between the time elapsed from the previous seismic event and the magnitude of the next seismic event at different magnitudes of previous seismic event is highly significant and using the parametric linear form in the copula parameter will be adequately characterized, namely FGM (conditional) copula parameter $\hat{\theta}_1(x_{t_{i-1}}) = \sin(0.666436 - 0.047099x_{t_{i-1}})$ for Region 1 and Galambos (conditional) copula parameter $\hat{\theta}_2(x_{t_{i-1}}) = \exp(-2.676213 + 0.325290x_{t_{i-1}})$ for Region 2. According to these results, when 1999 Marmara earthquake with a magnitude 7.4 that occurred in region 1 is considered, the probability of an earthquake with a magnitude greater than 7 in region 1 in the next 20 years is insignificant. Similarly, when 2017 Bodrum earthquake with a magnitude 6.5 that has occurred in region 2 recently is also considered, the probability that an earthquake with a magnitude greater than 7 in region 2 in the next 10 years is insignificant. Furthermore, with these conclusions, we have emphasized the rightfulness of the use of the Semi-Markov model in the previous studies [6, 7, 8] to obtain the earthquake occurrence probabilities in the regions 1 and 2.

Acknowledgement. I thank to the reviewers for their comments and suggestions for the manuscript.

REFERENCES

- [1] Abegaz, F., Gijbels, I. and Veraverbeke, N., Semiparametric estimation of conditional copulas, *Journal of Multivariate Analysis*, 110, (2012), 43-73.
- [2] Acar, E.F., Nonparametric estimation and inference for the copula parameter in conditional copulas, Dissertation, University of Toronto, 2010.
- [3] Acar, E.F., Craiu, R.U. and Yao, F., Dependence calibration in conditional copulas: A nonparametric approach, *Biometrics*, 67(2), (2011), 445-453.
- [4] Acar, E.F., Craiu, R.U. and Yao, F., Statistical testing of covariate effects in conditional copula models, *Electronic Journal of Statistics*, 7, (2013), 2822-2850.
- [5] Allcroft, D.J. and Glasbey, C.A., A simulation-based method for model evaluation, *Statistical Modelling*, 3(1), (2003), 1-13.
- [6] Altınok, Y., Seismic risk estimation of the North Anatolian Fault Zone Using Semi-Markov model, *Jeofizik*, 2, (1988), 44-58 (article in Turkish with English abstract).
- [7] Altınok, Y., Evaluation of earthquake risk in West Anatolia by Semi-Markov model, *Jeofizik*, 5, (1991), 135-140 (article in Turkish with English abstract).
- [8] Altınok, Y. and Kolçak, D., An application of the Semi-Markov model for earthquake occurrences in North Anatolia, Turkey, *Journal of Balkan Geophysical Society*, 2, (1999), 90-99.

- [9] Bartram, S., Taylor, S. and Wang, Y., The Euro and European financial market dependence, *Journal of Banking and Finance*, 31, (2007), 1461–1481.
- [10] Bee, M., Modelling credit default swap spreads by means normal mixtures and copulas, *Applied Mathematical Finance*, 11, (2004), 125-146.
- [11] Bhatia, S.C., Kumar, M.R. and Gupta, H.K., A Probabilistic Seismic Hazard Map of India and Adjoining Regions, *Annali Di Geofisica*, 2(6), (1999), 1153-1164.
- [12] Breymann, W., Dias, A. and Embrechts, P., Dependence structures for multivariate high-frequency data in finance, *Quantitative Finance*, 3, (2003), 1-14.
- [13] Cameron, A.C., Li, T., Trivedi, P.K. and Zimmer, D.M., Modelling the differences in counted outcomes using bivariate copula models with application to mismeasured counts, *The Econometrics Journal*, 7, (2004), 566-584.
- [14] Chen, X., Fan, Y., Pouzo, D. and Ying, Z., Estimation and model selection of semiparametric multivariate survival functions under general censorship, *Journal of Econometrics*, 157, (2010), 129-142.
- [15] Clayton, D.G., A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, 65, (1978), 141-151.
- [16] Fan, J., Zhang, C. and Zhang, J., Generalized Likelihood Ratio Tests and Wilks Phenomenon, *Annals of Statistics*, 29(1), (2001), 153-193.
- [17] Fischbach, P., Copula-Models in the Electric Power Industry, Dissertation, University of St. Gallen, 2010.
- [18] Frees, E.W. and Valdez, E.A., Understanding relationships using copulas, *North American Actuarial Journal*, 2, (1998), 1-25.
- [19] Gijbels, I., Veraverbeke, N. and Omelka, M., Conditional copulas, association measures and their application, *Computational Statistics and Data Analysis*, 55(5), (2011), 1919-1932.
- [20] Hougaard, P., A class of multivariate failure time distributions, *Biometrika*, 73, (1986), 671-678.
- [21] Internet: Earthquake Katalog, Bogazici University, Kandilli Observatory and Earthquake Research Institute, National Earthquake Monitoring Center. (1901- 2008). URL:<http://www.webcitation.org/query?url=http%3A%2F%2Fwww.koeri.boun.edu.tr%2Fsisimo%2FMudim%2Fkatalog.asp&date=2015-12-15>, (2015).
- [22] Internet: 1900-20xx Earthquake Katalog, AFAD. (2008-2014). URL:<http://www.webcitation.org/query?url=http%3A%2F%2Fwww.deprem.gov.tr%2Ftr%2Fdepremkatalogu&date=2015-12-16>, (2015).
- [23] Jondeau, E. and Rockinger, M., The Copula-GARCH model of conditional dependencies: an international stock market application, *Journal of International Money and Finance*, 25, (2006), 827–853.
- [24] Klugman, S.A. and Parsa, R., Fitting bivariate loss distributions with copulas, *Insurance: Mathematics and Economics*, 24, (1999), 139-148.
- [25] Li, Y., Shi, B. and Zhang, J., Copula joint function and its application in probability seismic hazard analysis, *Acta Seismologica Sinica*, 21(3), (2008), 296-305.
- [26] Nelsen, R.B., An introduction to copulas, Springer, New York, 2006.
- [27] Nikoloulopoulos, A.K. and Karlis, D., Fitting copulas to bivariate earthquake data: The Seismic Gap Hypothesis revisited, *Environmetrics*, 19, (2008a), 251-269.
- [28] Nikoloulopoulos, A.K. and Karlis, D., Copula model evaluation based on Parametric Bootstrap, *Computational Statistics and Data Analysis*, 52, (2008b), 3342-3353.
- [29] Oakes, D., A model for association in bivariate survival data, *Journal of Royal Statistical Society Series B*, 44, (1982), 414-422.
- [30] Palaro, H.P. and Hotta, L.K., Using conditional copula to estimate value at risk, *Journal of Data Science*, 4, (2006), 93-115.

- [31] Patton, A.J., Modelling Asymmetric Exchange Rate Dependence, Working paper, University of California, San Diego, 2003.
- [32] Patton, A.J., Modelling Asymmetric Exchange Rate Dependence, *International Economic Review*, 47, (2006), 527-556.
- [33] Sadeghian, R., Forecasting time and place of earthquakes using a Semi-Markov model (with case study in Tehran province), *Journal of Industrial Engineering International*, 8(20), (2012), 1-7.
- [34] Shih, J.H. and Louis, T.A., Inferences on the association parameter in copula models for bivariate survival data, *Biometrics*, 51, (1995), 1384-1399.
- [35] Ünal, S., Modelling the earthquakes occurring on Turkey by Markov Chains, Dissertation, Gazi University, Ankara, 2010.
- [36] Ünal, S. and Çelebioğlu, S., A Markov Chain modelling of the earthquakes occurring in Turkey, *Gazi University Journal of Science*, 24, (2011), 263-274.
- [37] Zhang, C., Calibrating the degrees of freedom for automatic data-smoothing and effective curve checking, *Journal of the American Statistical Association*, 98(463), (2003), 609-628.

Current address: Serpil ÜNAL KARAÇAM: Usak University, Faculty of Arts and Sciences, Department of Statistics, Uşak, Turkey

E-mail address: serpil.unal@usak.edu.tr

ORCID Address: <http://orcid.org/0000-0002-4043-6832>