TÜBİTAK

Original Article

Turk J Med Sci
2011; 41 (4): 725-734
© TÜBİTAK
E-mail: medsci@tubitak.gov.tr
doi:10.3906/sag-1007-959

Analyzing dependence structure of thyroid hormones: a copula approach

Burcu ÜÇER

Aim: To model the level and structure of the dependence between thyroid hormones by using the copula approach.

Materials and methods: Dependence models were constructed with the help of copula functions explaining the relationships between thyroid hormones for the data supplied by Dokuz Eylül University's Faculty of Medicine's endocrine laboratory. For this purpose, 684 patients aged 0-85 were examined.

Results: Results indicated that none of the pairs of thyroid hormones exhibited a tail dependence structure; however, valid models exhibited a symmetric dependence structure. The findings implied that both the T3 and T4 levels had a significant dependence structure with TSH levels. Furthermore, Gaussian and t-copula structures were appropriate for the pairs.

Conclusion: Findings were compared with the results of conventional scalar measures to establish the importance of using copula models in analyzing thyroid hormone levels. Findings showed that the copula models revealed better indicators for health scientists for more accurate dependence modeling.

Key words: Thyroid hormones, copula, IFM, correlation, dependence

Tiroid hormonlarının bağımlılık yapısının analizi: bir copula yaklaşımı

Amaç: Bu çalışmada, copula yaklaşımı ile tiroid hormonları arasındaki bağımlılığın derecesini ve yapısını modellemek amaçlanmıştır.

Yöntem ve gereç: DEÜ, Tıp Fakültesi, endokrin laboratuarından alınan veriler için, tiroid hormonları arasında ilişkiyi açıklayan copula fonksiyonları yardımıyla bağımlılık modelleri oluşturulmuştur. Bu amaçla 0-85 yaşları arasında 684 hasta incelenmiştir.

Bulgular: Sonuçlar, bu hormonlar arasındaki bağımlılık yapısında kuyruk bağımlılığı olmadığını, bununla birlikte geçerli modellerin simetrik bir bağımlılık yapısına sahip olduğunu göstermektedir. Aynı zamanda, bulgular TSH düzeyi ile T3 ve T4 düzeyleri arasında anlamlı bir bağımlılık yapısı olduğunu göstermektedir. Ayrıca, Gaussian ve t-copula yapısının bu değişken çiftleri arasındaki bağımlılık yapısı için uygun olduğu saptanmıştır.

Sonuç: Bulgular, daha sonra klasik yöntemlerle karşılaştırılmış, tiroid hormon düzeylerinin analizinde copula yaklaşımının önemi vurgulanmıştır. Sağlık bilimciler için çalışmadaki bulgular, copula modelinin bağımlılık yapısını doğru modellemede daha iyi göstergeler sunduğunu ortaya çıkarmıştır.

Anahtar sözcükler: Tiroid hormonları, copula, IFM, korelasyon, bağımlılık

Received: 20.07.2010 – Accepted: 25.09.2010
Department of Statistics, Faculty of Sciences, Dokuz Eylül University, İzmir - TURKEY
Correspondence: *Burcu ÜÇER, Department of Statistics, Faculty of Sciences, Dokuz Eylül University, İzmir- TURKEY*
E-mail: burcu.hudaverdi@deu.edu.tr

Introduction

Thyroid hormones, produced by the thyroid gland, which rests in the middle of the lower neck, are essential hormones as they help control the metabolism of the body. Two active thyroid hormones, thyroxine (T4) and triiodothyronine (T3), are produced by the thyroid gland. These hormones play an important role in the production of proteins, in the regulation of body temperature, and in overall energy production and regulation (1). Thyroid hormones directly affect most of the organs in the body. If the thyroid is not operating properly, problems will occur in other parts of the body.

Iodine, which is a chemical element, helps the thyroid do its job. The body absorbs iodine from food and water. The body contains about 50 mg of iodine, and about one-fifth to one-third of this is stored in the thyroid. The iodine is combined with tyrosine to make important hormones.

When the thyroid functions properly, it will ensure the right amount of hormones to keep metabolism at a satisfactory rate. The pituitary gland controls the quantity of thyroid hormones in the bloodstream. When the pituitary gland senses that the thyroid hormone levels are too low or too high, it produces thyroid-stimulating hormone (TSH) and sends it to the thyroid to tell it what to do. The relation between the thyroid and thyrotropes is like a negative feedback loop. When the T4 level increases, TSH production is suppressed, and vice versa. The hypothalamus produces thyrotropin-releasing hormone (TRH), which is used to modulate TSH production. This production is blunted by somatostatin (SRIH), rising levels of glucocorticoids and sex hormones (estrogen and testosterone), and excessively high blood iodide concentrations.

When the thyroid gland does not work well, thyroid disease occurs. For example, when the thyroid is overactive, hyperthyroidism is the result, which means that the thyroid releases higher le vels of thyroid hormone. In the case of hyperthyroidism, the body uses up energy more quickly. When lower levels of thyroid hormone are produced by the thyroid, hypothyroidism occurs. In this case, the body uses up energy more slowly. Both of these levels affect the metabolism of the body in inappropriate ways.

To understand the relationships between TSH and T3, and TSH and T4, many scalar measures can be used. Unfortunately, modeling the dependent variable has often been based on a set of simplified assumptions. Dependence structures are usually too complex for these types of dependence criteria to identify. To this end, copula models have started to be used in recent empirical health literature. A convenient way to express joint distribution and dependence structures of 2 or more random variables is provided by copulas.

Copulas have become very important tools for describing the dependence structures between random variables, with different copulas representing different dependencies. The dependence concept is critically important, especially in biostatistics, in order to carry out accurate analyses. Because of this fact, understanding and applying results from copulas can be very beneficial.

Copula models, different from the conventional methods, consider the excess comovements in different type of hormones, and hence provide information about both the level and the structure of the dependence.

This study provides evidence that the copula models reveal more comprehensive information than the conventional methods for analyzing and comparing the dependence level and structure of thyroid hormones. It is hoped that these results will provide valuable and extra information for health scientists and researchers about the benefits of copulas.

The rest of the study is organized as follows: methodological issues are provided in Section 2, data and numerical results are given in Section 3, and Section 4 is devoted to conclusions.

Materials and methods

A copula is a function that models the dependence structure between random variables. The concept of a copula was first introduced by Sklar in 1959, gaining practical use in statistics in the late 1980s. The application of copulas in health sciences is, however, a relatively recent phenomenon. In this section, we first give some basic properties of copulas and then introduce our model.

Definition

A copula can be defined as a function that relates one-dimensional, uniformly distributed marginal distribution functions to their joint distribution functions (2). According to Sklar's theorem, when *F* is a joint distribution function of continuous random variables (X,Y) on R^2 with marginal distribution functions *F* and *G*, there exists a bivariate copula *C* such that:

$$F(x,y) = C(F_X(x), F_Y(y)). \quad (1)$$

By using a copula, the dependence structure between the random variables can be constructed through the cumulative distribution functions instead of actual observed values.

There are many copula functions (3). In this paper, those most preferred in the health sciences were used: the Gaussian copula, the t-copula, the Gumbel copula, the Frank copula, and the Clayton copula. The Gaussian copula enables us to establish a normality assumption. However, the marginal distributions and/or the dependence structure between the random variables may not be normal. Tail dependence may dominate the relationship between the random variables under consideration. We used the Gumbel and Clayton survival copulas to consider this possibility. The dependence structure may also exhibit a symmetric shape, in which case we used the Frank copula or the t-copula.

In this study, we investigated the dependence structure between 2 random variables, *X* and *Y*, which are hormone levels. After the cumulative distribution function transformation with U = F(X) and V = F(Y), the density of the bivariate copula C(u,ν) is defined as:

$$c(u,\nu) = \frac{\partial C(u,\nu)}{\partial u \partial \nu} \quad (2)$$

The density of the bivariate distribution F(x,y) can then be obtained by using marginal densities $f_X(x)$ and $f_Y(y)$ as follows:

$$f(x,y) = c(F_X(x), F_Y(y)) f_X(x) f_Y(y). \quad (3)$$

Schweizer and Wolff combined copulas with measures of dependence (4). They showed that Kendall's tau and Spearman's rho can be written in terms of copulas.

Model

The choice of copula is crucial in analyzing dependence across random variables. However, there may not be a unique copula that fits all situations. The most commonly used copula models in the health sciences are given in Table 1.

For the Gaussian copula, higher values of the dependence parameter, θ, indicate a stronger dependence between 2 random variables. Here, we see that a Gaussian copula is the dependence function of a joint Gaussian distribution. The Gaussian copula is symmetric, and for the hormone data, it implies that hormone levels are equally likely to increase or decrease together.

The t-copula is symmetric and exhibits tail dependence. With a t-copula, 2 random variables can be asymptotically tail-dependent even in the extreme case in which they are uncorrelated.

For the Gumbel copula, $\alpha = 1$ indicates independence, and the limit of the Gumbel distribution for $\alpha \to 0+$ shows perfect dependence. The Gumbel copula is asymmetric and indicates positive right-tail dependence. For the hormone data, it implies that 2 kinds of hormone are more likely to increase together than decrease together, indicating stronger dependence.

The Clayton copula is asymmetric and indicates negative left-tail dependence. For the hormone data, negative left-tail dependence implies that 2 kinds of hormones are more likely to decrease together than increase together.

The Frank copula is characterized by upper and lower tail independence. It is symmetric, assigning zero probability to events that are deep in the tails.

The association parameters provide the degree of dependence between the hormone levels. The first step, in order to calibrate the copula, is to transform the data using the probability integral transform. We find a parametric distribution that fits the univariate marginal data. This method requires an appropriate distribution for each margin. As a general approach, we estimate the parameters of margins by the maximum likelihood (ML) method for the sample. The log-likelihood function for $\{X_i\}_{i=1}^{N}$ is:

$$L(\alpha; x) = \sum_{i=1}^{N} \ln f(x_i;\alpha).$$

Table 1. Some copula functions.

Name	Model	Dependence parameter θ
Gaussian	$C(u,v;\theta)=\int_{-\infty}^{\phi^{-1}(u)}\int_{-\infty}^{\phi^{-1}(v)}\frac{1}{2\pi(1-\theta^2)^{1/2}}\exp\left(\frac{-(s^2-2\theta st+t^2}{2(1-\theta^2)}\right)dsdt$	$-1<\theta<1$
t	$C(u,v;\theta)=\int_{-\infty}^{t_\nu^{-1}(u)}\int_{-\infty}^{\phi_\nu^{-1}(v)}\frac{1}{2\pi(1-\theta^2)^{1/2}}\exp\left(-(1+\frac{s^2-2\theta st+t^2}{\nu(1-\theta^2)}\right)dsdt$	$-1<\theta<1$ $\nu>2$
Gumbel	$C_{gu}(u,v;\theta)=\exp\{-[(-\log u)^{1/\theta}+(-\log v)^{1/\theta}]^{\theta}\}$	$0<\theta\leq 1$
Clayton	$C(u,v;\theta)=\{u^{-\theta}+\nu^{-\theta}-1\}^{-1/\theta}$	$\theta\geq 1$
Frank	$C(u,v;\theta)=-\frac{1}{\theta}\ln\left(1+\frac{(\exp(-\theta u)-1)(\exp(-\theta\nu)-1)}{(\exp(-\theta)-1)}\right)$	$\theta\in IR\backslash\{0\}$

The maximum likelihood estimator (MLE) is then the value of $\alpha\in\theta$, maximizing $L(\alpha;x)$:

$$\hat{\alpha}_{ML}=\operatorname{argmax} L(\hat{\alpha};x).$$

Correct selection of the marginals is as important as a correct selection of the copula. Once the margins have been transformed, they are embedded into the copula in order to estimate the copula parameters. Given a random sample X, we calibrate the copula parameters by the ML method. Assuming known margins, the maximization of the function is:

$$L(\theta)=\sum_{i=1}^{N}\ln c(F_X(x,\hat{\alpha}_1),F_Y(y,\hat{\alpha}_2);\theta),$$

where $c(u,\nu)$ is the copula density, which can be obtained from Eq. (2).

The 2-step procedure explained above is called an inference function for the margins (IFM). This method separates the margins and the dependence structure (5). Some other estimation methods for copula models have also been given (6,7).

Once the model is estimated, we need to verify its goodness of fit to explain the dependence structure between the random variables. Graphical methods offer many possibilities. The easiest and natural way of checking the adequacy of a copula model is to plot empirical copula C_n and compare it with an artificial data set of the same size generated from C_θ.

Data and numerical results

The aim of this study was to find the degree and the structure of the dependence between thyroid hormones. For this, laboratory values of free T3 (FT3), free T4 (FT4), and TSH were taken into account. Dependence models were constructed with the help of copula functions to explain the relationship between these hormones for the data supplied by Dokuz Eylül University's Faculty of Medicine's endocrine laboratory, and 684 patients aged 0-85 were examined. These data were used before in another study (8), in which the dependence relation was investigated with a different approach. Computational results were obtained with MATLAB (9).

The testing methodology can be summarized as follows:

1. Fitting the univariate marginal distribution of each series of hormone data by maximum likelihood estimation. Among the different distributions, we select the one that best fits the data. The selection criterion is the higher log-likelihood value.
2. Transforming each data vector into uniform variates using the probability integral transform.
3. Plugging the data into copula functions for each pair of transformed data vectors and estimating the copula parameters by the ML method.
4. Applying graphical techniques for choosing the appropriate copula model.

In Table 2, Spearman's rho and Kendall's tau between the variables are given. These values explain only the degrees of relations between the variables, which are both negative, but not so strong. However, these measures do not contain any information about the shape of the dependence structure between the variables.

Table 2. Spearman's rho and Kendall's tau for the hormone data.

	Spearman's rho	Kendall's tau
TSH and FT3	–0.032 (0.411*)	–0.020 (0.429*)
TSH and FT4	–0.269 (0.000*)	–0.186 (0.000*)

(*P-value in parentheses)

In Tables 3, 4, and 5, univariate marginal fittings are given for TSH, FT3, and FT4, respectively. The log-likelihood values are compared to select the best fitting marginal distribution. The selection criterion between the distributions is the minimum of the negative log-likelihood values. The selected marginal distributions for the copula models, which are indicated in bold, are the generalized extreme value

Table 3. Negative log-likelihood values of the marginal distributions of TSH.

Distribution	Parameter 1	Parameter 2	Parameter 3	Likelihood
Lognormal	0.0703	1.3732		1.2332e + 003
Gamma	0.7783	2.9635		1.2382e + 003
Exponential	2.3063			1.2538e + 003
Weibull	2.0084	0.8157		1.2195e + 003
GEV	**0.5556**	**0.8676**	**0.8496**	**1.1955e + 003**
Normal	2.3063	4.9394		2.0594e + 003

Table 4. Negative log-likelihood values of the marginal distributions of FT3.

Distribution	Parameter 1	Parameter 2	Parameter 3	Likelihood
Lognormal	**0.8754**	**0.2830**		**704.4294**
Gamma	12.7150	0.1964		707.6011
Exponential	2.4973			1.3081e + 003
Weibull	2.7599	3.3186		778.5479
GEV	–0.0446	0.6040	2.1834	705.4278
Normal	2.4973	0.7285		752.2889

Table 5. Negative log-likelihood values of marginal distributions of FT4.

Distribution	Parameter 1	Parameter 2	Parameter 3	Likelihood
Lognormal	0.3855	0.1750		41.5332
Gamma	**33.3606**	**0.0447**		**38.0417**
Exponential	1.4926			956.5510
Weibull	1.6016	5.0509		135.2436
GEV	–0.1001	0.2482	1.3883	52.9701
Normal	1.4926	0.2621		54.0282

distribution for TSH, the lognormal distribution for FT3, and the gamma distribution for FT4.

After the selection of the most convenient marginal distributions, each pair was fitted with all of the copula models given in Tables 6 and 7, in which estimated parameter values are shown.

In the last step, we determined which copula model was appropriate for the data. Graphical techniques were applied to all copula models and the best-fitting copula model was chosen visually (Figures 1 and 2). The distance between the theoretical and empirical copula was the decision criterion. The lower the distance, the better the model. The overall results suggested that the Gaussian copula offers good accuracy for the bivariate case for TSH and FT3 data, and the t-copula gives good accuracy for the bivariate case for TSH and FT4 data. Since both of the copula models are symmetric, the dependence structures are also symmetric for both pairs. This result implies that the hormone levels are equally likely to have negative relationships. TSH and FT4 exhibited tail dependence. With the t-copula, the dependence can be asymptotically tail-dependent in extreme cases. The Gaussian and t-copula models with estimated parameters for TSH and FT3, and TSH and FT4, are also given in Figures 3 and 4. These Figures help to understand the shape of the dependence structure. It is seen that the Gaussian copula type of dependence is symmetric, while the t-copula type of dependence is symmetric but also has some dependence in the tails.

Table 6. MLE estimates for copula models for TSH and FT3.

Copula	Parameter
Gaussian	**–0.0627**
t	–0.0472, (2.1)
Gumbel	1.001
Frank	–0.1940
Clayton	1.4509e – 006

Table 7. MLE estimates for copula models for TSH and FT4.

Copula	Parameter
Gaussian	–0.3508
t	**–0.3382**
Gumbel	1.00001
Frank	–2.01310
Clayton	1.4509e – 006

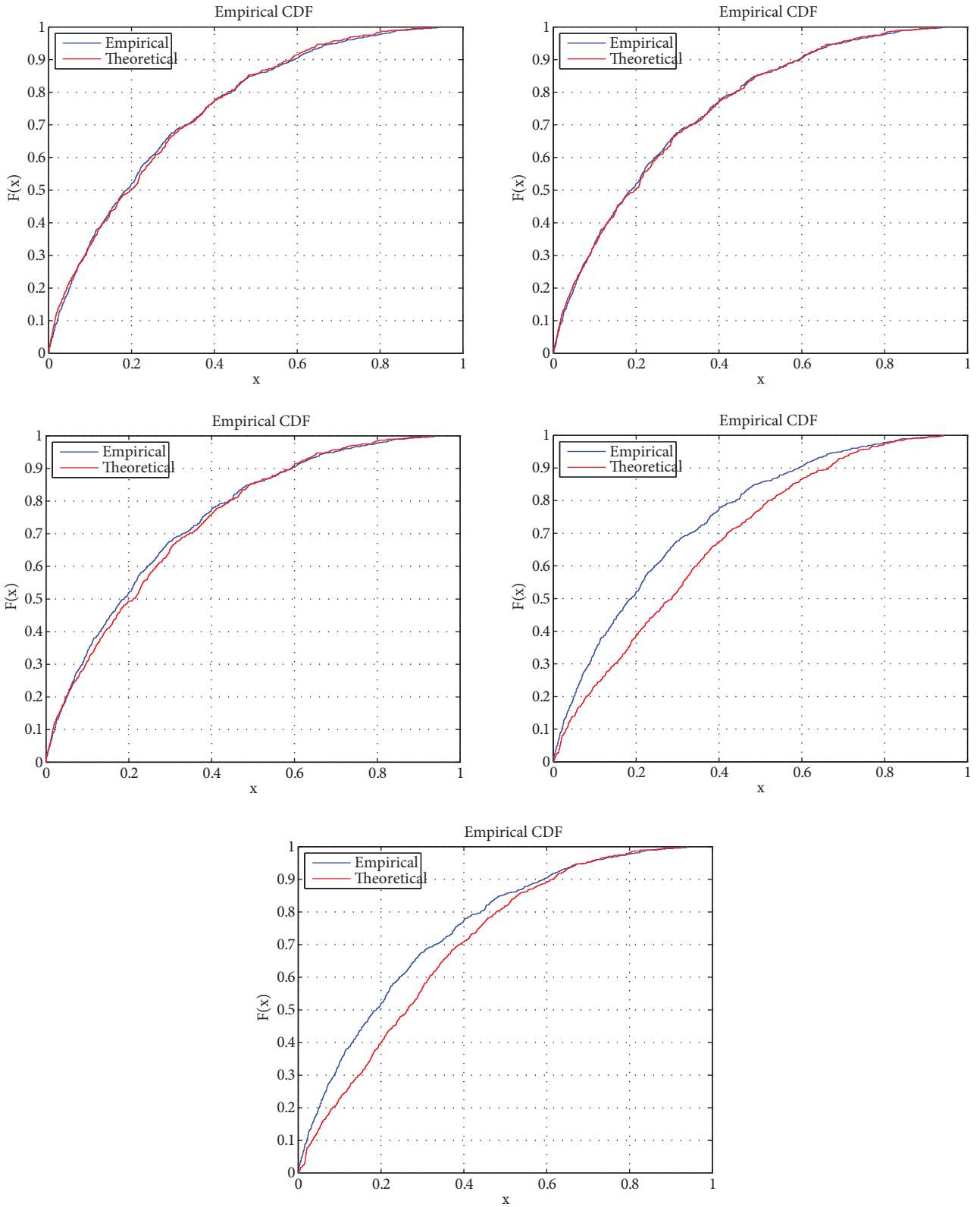

Figure 1. Empirical and theoretical comparison of copula models for TSH and FT3.

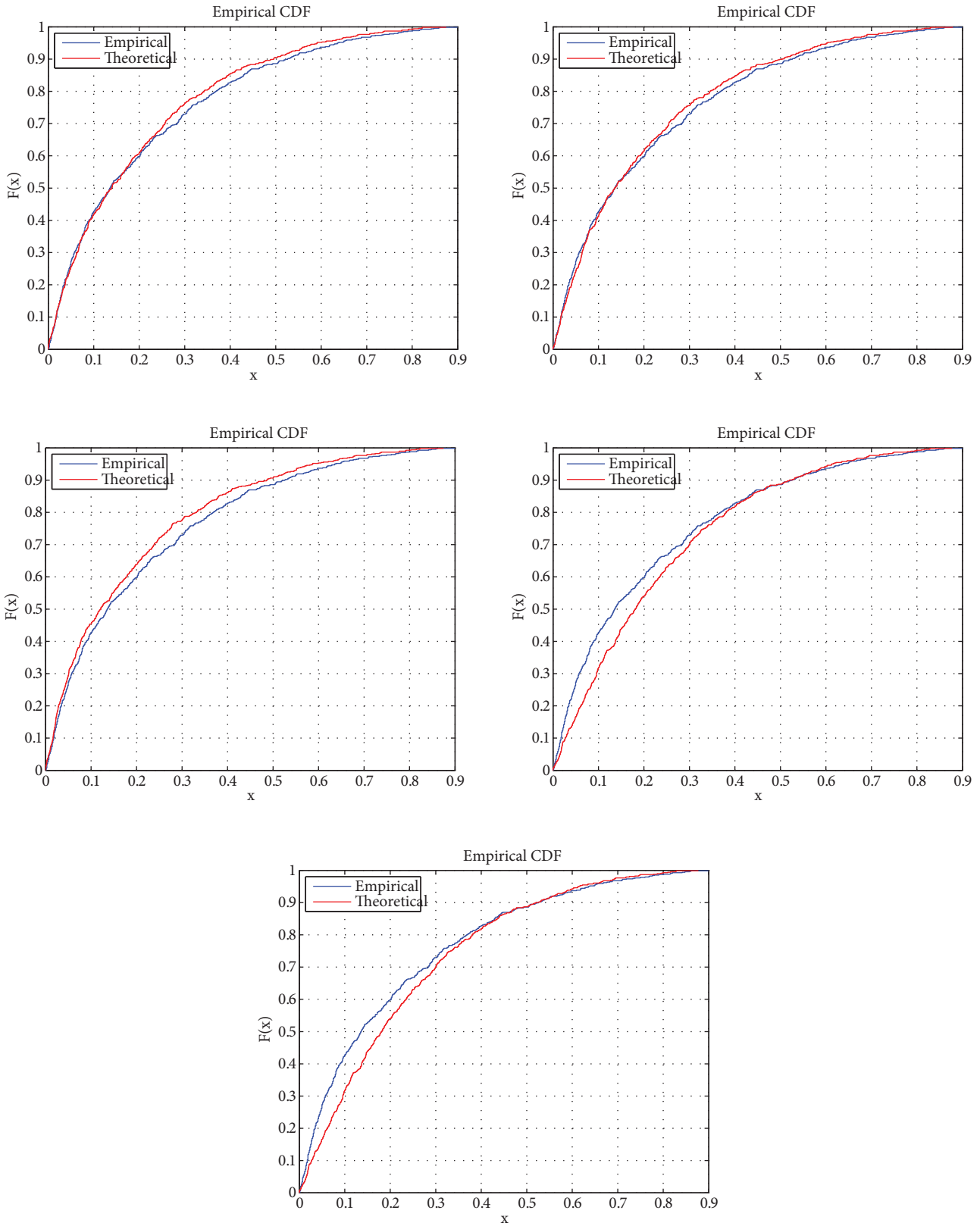

Figure 2. Empirical and theoretical comparison of copula models for TSH and FT4.

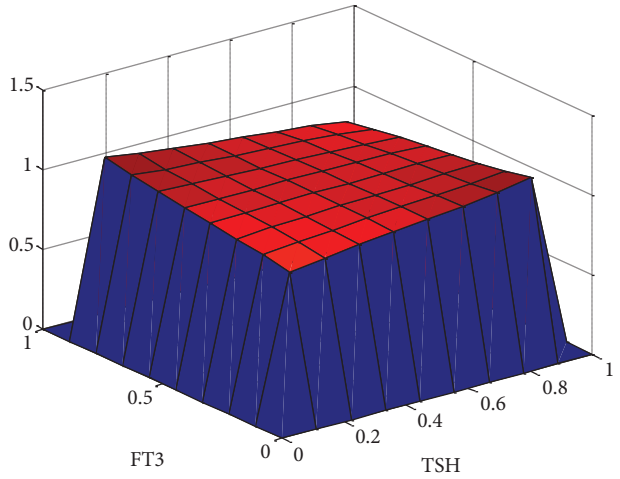

Figure 3. The Gaussian copula with parameter θ = –0.0627 for TSH and FT3.

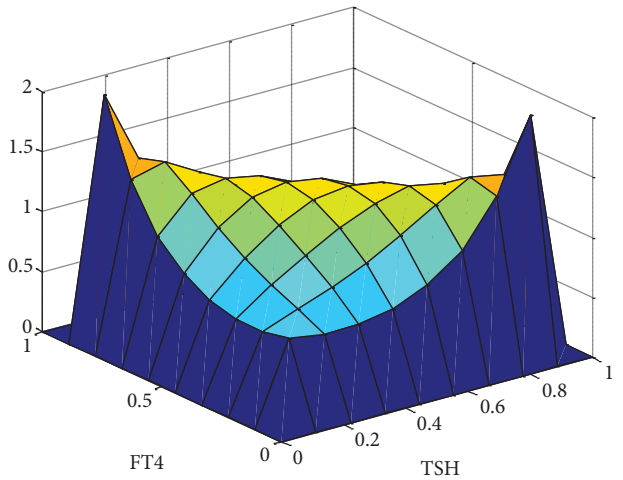

Figure 4. The t-copula with parameter θ = –0.3882 for TSH and FT4.

Conclusions

Simple dependence measures such as Pearson's correlation coefficient, Kendall's tau, and Spearman's rho can be an essential part of medical science studies. Unfortunately, these measures are not adequate for summarizing complex dependence structures. The dependences between pairs of variates can be various. For bivariate data sets, in understanding the relationship between the variables, not only the level of dependence but also the structure of dependence is important.

By using the copula approach, it is easy to see the model of the dependence between variables visually, whereas some scalar measures like Kendall's tau and Spearman's rho only explain the level of dependence.

The analysis suggests that the Gaussian copula gives good accuracy for the bivariate case of TSH and FT3 data, and the t-copula gives good accuracy for the bivariate case of TSH and FT4 data. A symmetric dependence structure is convenient for both of the pairs. Clinically, this means that hormone levels are equally likely to have a negative relationship. For example, when the TSH level is decreasing (or increasing), the FT3 level is increasing (or decreasing) at a similar rate, and vice versa. This interpretation is also valid for the relationship of TSH and FT4. However, the difference with TSH and FT4 is that their relation exhibits tail dependency. With the t-copula, the dependence can be asymptotically tail-dependent in the extreme case. In other words, for very high or very low levels of TSH or FT4, there is some some negative dependency, which the Gaussian type of dependence of TSH and FT3 lacks.

Acknowledgement

The authors would like to thank Dr. Dilek Çımrın from Dokuz Eylül University's Faculty of Medicine's endocrine laboratory for supplying data and the referees for helpful comments.

References

1. Ulutagay G, Nasibov E. A fuzzy inference system for thyroid disease diagnosis. Proceedings of 1st International Symposium on Computing in Science & Engineering, Gediz Üniversitesi, İzmir, 2010.
2. Mari D, Kotz S. Correlation and dependence. Imperial College Press; 2001.
3. Nelsen RB. An introduction to copulas. Springer; 1998.
4. Schweizer B, Wolff E. On parametric measures of dependence for random variables. Annals of Statistics 1981; 9: 879-85.
5. Joe H. Multivariate models and dependence concepts. London: Chapman and Hall; 1997.

6. Genest C, Rivest L. Statistical inference procedures for bivariate Archimedean copulas. Journal of the American Statistical Association 1993; 88: 1034-43.

7. Genest C, Ghoudi K, Rivest L. A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 1995; 82: 543-552.

8. Oruç Ege Ö, Üçer B. A new method for local dependence map and its applications. Türkiye Klinikleri J Biosta 2009; 1: 1-8.

9. MathWorks Inc. MATLAB: the language of technical computing. Natick (MA): MathWorks; 2007.