Investigating tail dependence for the Bivariate Generalized Gamma

D. J. de Waal, A. Verster, R. Fraser and C. Marais

Introduction

In this study we investigate the tail dependence of the two variables of the bivariate Generalized Gamma distribution. Tail dependence plays an important role in for example the estimation of tail probabilities in multivariate models. The dependence structure in the tails are popularly describes through copulas as discussed in various literature, see for example Beirlant *et al*. (2004). In this paper we start by discussing the univariate Generalized Gamma distribution and then we extend the univariate distribution to the multivariate Generalized Gamma distribution. Tail dependency is then investigated for the bivariate Generalized Gamma distribution, by following the literature of Ledford and Tawn (1997), for different parameter values.

Generalized Gamma distribution

The Gamma class of distributions covers a number of well-known distributions such as the Gamma, Weibull and Exponential distributions. In this section we define and discuss the Generalized Gamma (GGAM) distribution. The distribution function of $X \sim GGAM(k,\mu,\sigma)$ is given in [Eq (1)]

$$F(x|k,\mu,\sigma) = P(X < x) = \frac{1}{\Gamma(k)}\int_0^{\nu} e^{-u}u^{k-1}du \ , x > 0 \tag{1}$$

where

$$\nu = (xe^{\alpha})^{\frac{1}{\beta}} \tag{2}$$

or

$$\nu = \exp\left(\psi(k) - \frac{y-\mu}{\sigma}\sqrt{\psi'(k)}\right) \tag{3}$$

and

$$Y = -\log(X). \tag{4}$$

$\psi(k) = \frac{d}{dk}\log(\Gamma(k))$ and $\psi'(k) = \frac{d}{dk}\psi(k)$ are the digamma and trigamma functions respectfully where $-\psi(1) = 0.5772$ is the well known Euler's constant and $\psi'(1) = \frac{\pi^2}{6}$. The parameters μ and σ can be expressed in terms of α and β as $\mu = \alpha - \beta\psi(k)$ and $\sigma = \beta\sqrt{\psi'(k)}$. $\beta = \frac{\sigma}{\sqrt{\psi'(k)}}$ is known as the tail index. From [Eq (1)] V are Gamma distributed denoted by $V \sim Gam(k,1)$. It is also known that $E(\log(V)) = \psi(k)$, the distribution [Eq (1)] is constructed such that $E(log(X)) = -\mu$ and $Var(log(X)) = \sigma^2$ (Beirlant *et al*., 2002). The parameter space is $\Omega = \{-\infty < \mu < \infty, \sigma > 0, k > 0\}$.

Multivariate Generalized Gamma distribution

The Multivariate Generalized Gamma (MGGAM) distribution is in accordance with the GGAM distribution. The probability density function of $\underline{X} = (X_1, X_2, \dots, X_p) \sim MGGAM(\underline{k}, \underline{\mu}, \underline{\Sigma})$ is given in [Eq (5)]

$$f(\underline{x}) = |\Sigma|^{-\frac{1}{2}} \prod_{i=1}^{p} \left\{ \frac{\sqrt{\psi'(k_i)}}{\Gamma(k_i) x_i} e^{-v_i} v_i^{k_i} \right\}, \underline{X} > \underline{0} \tag{5}$$

where

$$V_i = e^{\psi(k_i) - \sqrt{\psi'(k_i)} \Sigma_{(i)}^{-\frac{1}{2}} (\underline{Y} - \underline{\mu})}, \quad i = 1, \dots, p,, \tag{6}$$

$$\underline{\mu} = (\mu_1, \mu_2, \dots, \mu_p),$$

$$\psi(k_i) = \frac{d}{dk_i} log\Gamma(k_i)$$

and

$$\psi'(k_i) = \frac{d}{dk_i} \psi(k_i).$$

$\underline{V}$ can be written in matrix form as $\underline{V} = exp\{\psi(\underline{k}) - D_\psi \Sigma^{-\frac{1}{2}} (\underline{Y} - \underline{\mu})\}$ where $Y_i = -log(X_i)$ and $\underline{Y} = (Y_1, Y_2, \dots, Y_p)$,$D_\psi = diag\left(\sqrt{\psi'(k_1)}, \sqrt{\psi'(k_2)}, \dots, \sqrt{\psi'(k_p)}\right)$ and $\Sigma^{-\frac{1}{2}}$ is the symmetric square root of the inverse of the covariance matrix Σ. Further $\underline{\mu} = E(\underline{Y})$ and $\Sigma = Cov(\underline{Y}, \underline{Y}')$. the parameter space is: $\underline{-\infty} < \underline{\mu} < \underline{\infty}, \underline{k} > 0$ and $\Sigma > 0$. A random variable $\underline{X}$ is $MGGAM(\underline{k}, \underline{\mu}, \underline{\Sigma})$ if the elements of $\underline{V}$ are distributed Gamma($k_i, 1$), $i = 1, \dots, p$, independent of each other. The matrix $B^{-1} = D_\psi \Sigma^{-\frac{1}{2}}$ can be considered as the tail index of the distribution in line with the univariate tail index of the GGAM, namely $\beta = \frac{\sigma}{\sqrt{\psi'(k)}}$. Marginally $\beta_i = \frac{\sigma_{ii}}{\sqrt{\psi'(k_i)}}$ will be the tail index for the i[th] variable. The bivariate case of the MGGAM is illustrated through the following example.

<u>Example 1</u>

Let $p = 2$, $\Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$, $\underline{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $\underline{k} = \begin{pmatrix} 0.8 \\ 5 \end{pmatrix}$. Figure 1 shows a simulated dataset of $n = 500$ observations distributed GMGAM with the given parameters. An interesting observation from the scatter plot is the presence of tail dependency between the two variables in the extreme right hand corner. The effect of tail dependency is discussed in the next section.

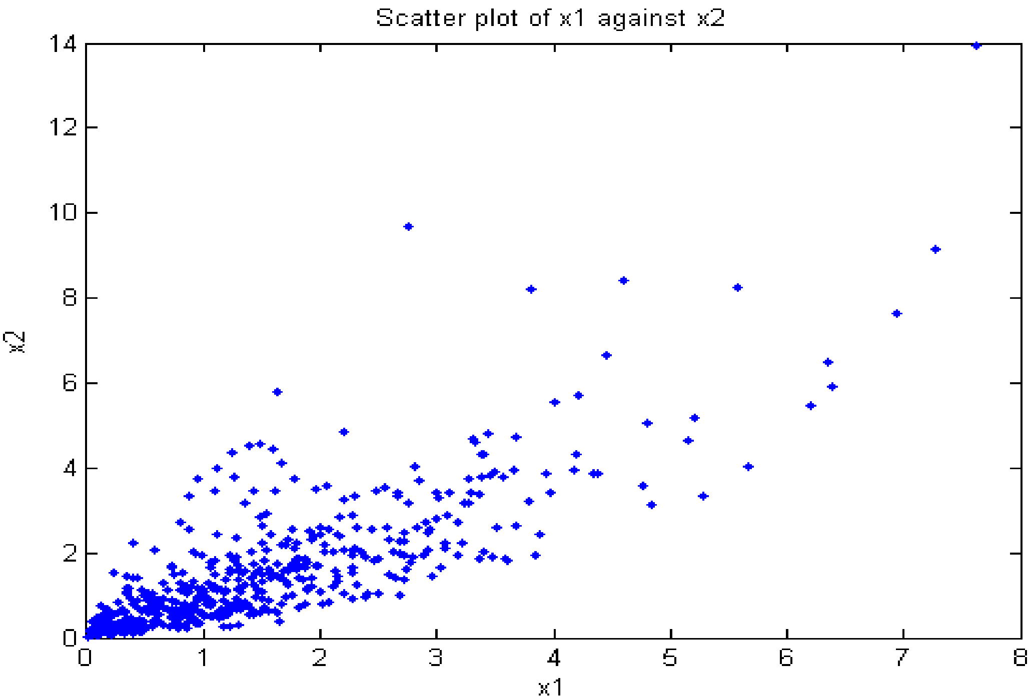

Figure 1: Scatter plot of 500 simulated MGGAM observations

Tail dependence

In the modelling of multivariate data tail dependence plays a critical role. Copulas are popular methods for modelling multivariate data especially to describe the dependence structure in the tails. To read more about copulas and tail dependence see Beirlant *et al*., 2004. The multivariate normal is a classical case where asymptotic tail dependence breaks down although there is a high correlation between the variables. The estimation of the coefficient of tail dependence between two variables X_1 and X_2 has been discussed by various authors and Beirlant *et al*., (2004) also provides a good review on this. We will follow the method discussed by these authors as proposed by Ledford and Tawn (1996), namely defining η as the coefficient of tail dependence between variables Z_1 and Z_2 if the joint survival function $P(Z_1 > z, Z_2 > z)$ can be written as

$$P(Z_1 > z, Z_2 > z) = \mathcal{L}(z) z^{-\frac{1}{\eta}},\ z > 0. \tag{7}$$

$\mathcal{L}$ is a slowly varying function as $z \to \infty$ and $0 < \eta \leq 1$. $\eta = 1$ implies dependence and $\eta < 1$ independence (Zellner, 1977). $Z_i, i = 1, 2$ are Fréchet variables defined as $Z_i = -\frac{1}{\log(F_i)}$ where F_i is the cdf of the marginal of X_i. If F_i is not known, it can be estimated through the empirical cdf $\hat{F}_i(\mathrm{x_{j,n}}) = \frac{j}{n+1}, \mathrm{j} = 1, .., \mathrm{n}$ based on the n sorted observations $x_{j,n}$ on X_i. Any other estimates of F_i can also be considered. The assumption in [Eq (7)] implies that $T = \min(Z_1, Z_2)$ has a regular varying tail with index $-\frac{1}{\eta}$ and any of the peaks over threshold (POT) distributions may be candidates to estimate η. The question now arises; for what values of (k_1, k_2, ρ) will η be equal to one. Thus for what values of (k_1, k_2, ρ) will there be tail dependence between the two variables. This question is investigated in the following example.

<u>Example 2</u>

In this example we investigate for what values (k_1, k_2, ρ) the tail dependency between the two variables X_1 and X_2 exist. The marginals are not available explicitly, but we approximate the marginals to be Gamma distributed. A Gamma distribution is fitted on each marginal using the MATLAB command gamfit(X_1) and gamfit(X_2) respectively. After doing the Fréchet transformation on the marginals of $X_i, i = 1{,}2$ the joint survival function $P(Z_1 > z, Z_2 > z)$ can be transformed to $P(T > z)$ where $T = \min(Z_1, Z_2)$. Since a tail probability is calculated we consider a POT distribution for estimating η. POT distributions only considers observations greater than a chosen threshold τ. The POT distribution considered here is the generalized single Pareto (GSP) distribution discussed in Verster and De Waal (2010). This distribution has the advantage of only one parameter if the threshold is known. The GSP is shown in [Eq (8)]

$$P(T > z|\tau) = \left\{1 + \frac{\eta}{1+\tau\eta}(z - \tau)\right\}^{-\frac{1}{\eta}} \qquad , t > \tau. \tag{8}$$

The optimum threshold is chosen by minimizing the mean squared error between the predictive quantiles of T and the observed quantiles. See Verster and De Waal (2010) for a discussion of this method for choosing an optimum threshold. To test for tail dependence we consider the posterior distribution of η. The $P(\eta \geq 1)$ is obtained from the posterior. If $P(\eta \geq 1) \geq 5\%$ we accept the hypothesis at $\alpha = 0.05$ and say that η might be equal to 1.This is illustrated in Figure 2.

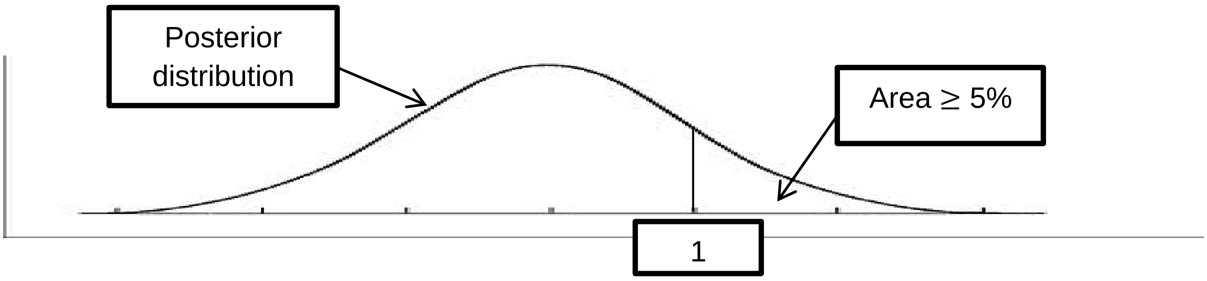

Figure 2: Illustration of the posterior of η

Consider again Example 1 where $p = 2$, $\Sigma = \begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}$, $\underline{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $\underline{k} = \begin{pmatrix} 0.8 \\ 5 \end{pmatrix}$ and $\rho = 0.8$. We will now test for tail dependence which were evident from Figure 1. For different values of η the posterior distribution of η under the GSP distribution is obtained. The posterior distribution is given by

$$\pi(\eta|\boldsymbol{v}) \propto \prod_{i=1}^{N_t} \frac{1}{1+\eta t_v}\left[1+\frac{\eta(v_i-t_v)}{1+\eta t_v}\right]^{\frac{-1}{\eta}-1} \pi(\eta) \tag{9}$$

where

$$\pi(\eta) \propto \frac{e^{-(\eta)}}{1+\eta t_v}. \tag{10}$$

is the maximal data information (MDI) prior (Verster and De Waal, 2010). The mode of the posterior distribution is an estimate of η. The optimum threshold is obtained through first estimating η from the posterior at different values of τ and then calculating the minimum mean squared error (MSE) between the predictive quantiles of T and the observed quantiles [Eq (8)] at the different threshold values. Figure 3 shows the different values of τ plotted against the MSE. A minimum MSE is obtained at $\tau = 17$.

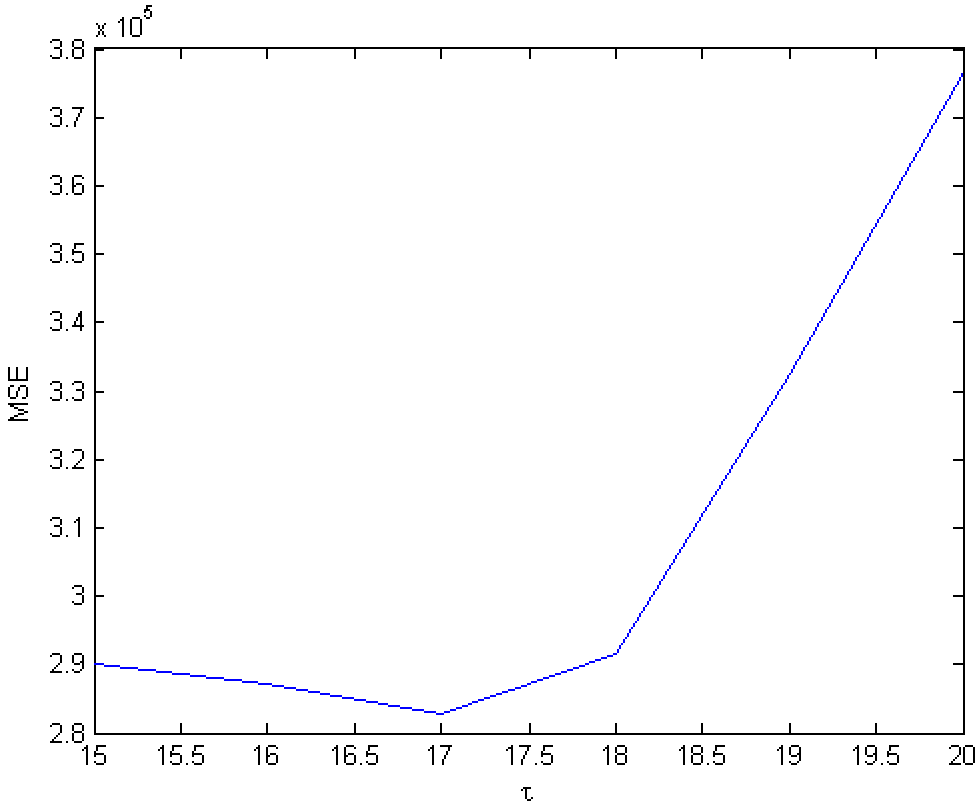

Figure 3: Various threshold values plotted against the respective MSE values

After obtaining the optimum threshold values η is again estimated as the mode of the posterior (Figure 4) given τ. From Figure 4 it can be seen that $\hat{\eta} = 1.05$. The $P(\eta \geq 1)$ is then calculated from the posterior as 0.5115, indicating a tail dependence.

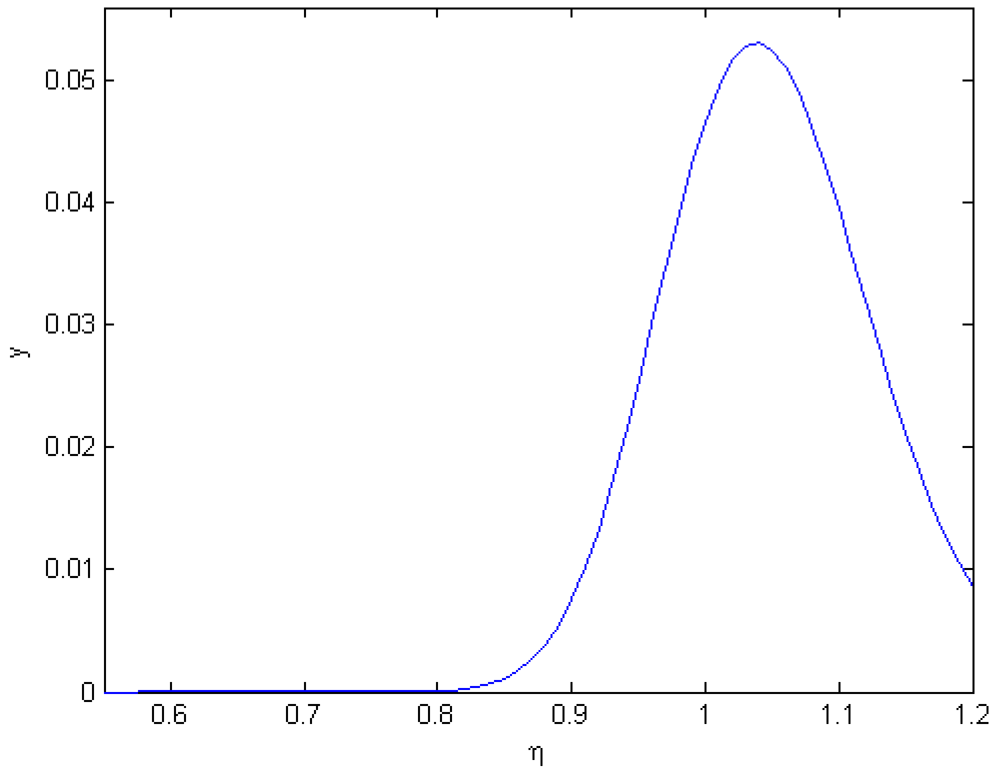

Figure 4: Posterior of η at $\tau = 17$

Various simulation studies were conducted for different values of k_1, k_2 and ρ to test for tail dependence. Each simulation was repeated several times. The outcomes are given in Figure 5 in the Appendix. The dots indicates that the hypothesis $\eta = 1$ is accepted, thus indicating tail dependence. From Figure 5 we can roughly conclude that for $\rho \geq 0.6$ and $k_1, k_2 > 1$ tail dependence between the variables exist.

Conclusion

When modelling multivariate data tail dependence plays an critical role for example in estimated tail probabilities. By following the literature of Ledford and Tawn (1997), we have shown in this paper that tail dependence exist in the bivariate Generalized Gamma distribution for $\rho \geq 0.6$.

References

Beirlant J, Goedgebeur Y, Segers J, Teugels J (2004) Statistics of extremes. Theory and applications. Wiley, Chichester.

Verster A, De Waal DJ (2009) Approximating the Generalized Burr-Gamma with a Generalized Pareto-type of distribution. Technical Report 2/97, University of the Free State, Bloemfontein. http://natagri.ufs.ac.za/dl/userfiles/Documents/00001/1718_eng.pdf

Ledford AW, Tawn JA (1997) Modelling dependence within joint tail regions. Journal of the Royal Statistical Society B 59, 475 – 499.

Appendix

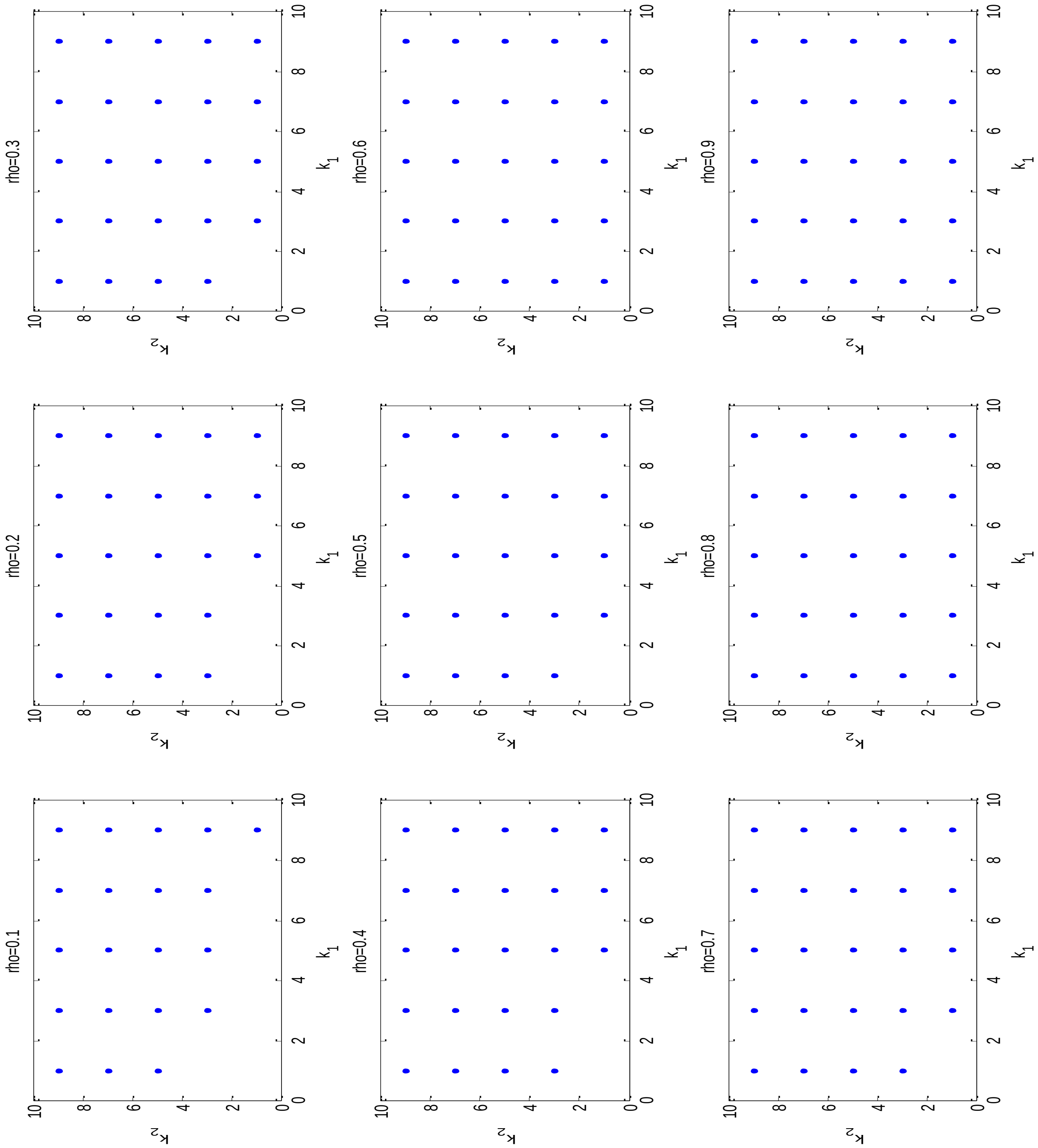

Figure 5: Indicating tail dependence for various values of k_1, k_2 and ρ