Abstract

There is often interest in understanding how the extremely high/low values of different processes are related to each other. One possible way to tackle this problem is via an asymptotic approach, which involves fitting the multivariate generalized Pareto distribution (MGPD) to data that exceed a suitably high threshold. There are two possible definitions of the MGPD. The first, which has classically been used, consider values that jointly exceed the thresholds for all components. The second definition considers those values which exceed a threshold for at least one component. The first type of definition is widely investigated, so here we focus on the second type of definition (MGPD type II). One aim of this paper is
to investigate the applicability of classical parametric dependence models within MGPD type II. Since in this framework the set of applicable asymmetric dependence models is more restricted, a general transformation is proposed for creating asymmetric models from the well-known symmetric ones. We apply the proposed approach to the exceedances of bivariate wind speed data and outline methods for calculating prediction regions as well as evaluating goodness-of-fit.

**Key words:** Asymmetric dependence models, goodness-of-fit, multivariate threshold exceedances, prediction regions, wind speed data.

1 Introduction

Univariate extreme value theory provides limit results for the distribution of extremes of a single stationary process. Thus block maxima (monthly, yearly etc.) of stationary observations usually can be modeled using the generalized extreme value distribution (GEV), which has a distribution function (df.) of the form

\[ G_{\xi, \mu, \sigma}(x) = \exp \left\{ - \left( 1 + \xi \frac{x - \mu}{\sigma} \right)^{-\frac{1}{\xi}} \right\}, \quad (1) \]

where \( 1 + \xi \frac{x - \mu}{\sigma} > 0 \). \( \mu \in \mathbb{R} \) is called the location parameter, \( \sigma > 0 \) the scale parameter and \( \xi \in \mathbb{R} \) the shape parameter. Similarly, threshold exceedances (e.g. over a high quantile of the observations) can be approximated by the gen-
eralized Pareto distribution (GPD), which has a df. of the form

\[ h_{\xi, \sigma}(x) = 1 - \left(1 + \frac{x}{\sigma\xi}\right)^{-\frac{1}{\xi}}, \]  

(2)

where \(1 + \frac{x}{\sigma\xi} > 0\) and \(\sigma > 0\). The above distributions are strongly linked, since the (limit) distribution of the exceedances is GPD if and only if the (limit) distribution of the maxima is GEV. Moreover the shape parameter is the same for both distributions.

Multivariate extremes are often modeled by applying block maxima methods (BMM) to the componentwise maxima of the series being studied. An important drawback of this approach is that it ignores whether the extremes of the different processes occurred simultaneously or not. The annual maxima may of the processes may, for example, all have arisen on the same day, but it is also possible that they all occurred in completely different methods. The BMM approach does not distinguish between these situations. This problem can be avoided by modeling the sizes of all observations that exceed a given high threshold, rather than modeling the highest value within a particular block of time. There are several methods for analyzing multivariate threshold exceedances depending on which observations are considered as exceedances. In a classical framework only those observations which exceed the threshold in all components are involved in modeling. The methods using this definition (which we call type I) are widely investigated in the literature, see e.g. Smith (1994) for one of its first applications or the PhD thesis of Michel (2006) for a review. However the type I definition can be
very restrictive in some cases, as it still ignores many potentially important observations: those having at least one component below the threshold. For illustration see Fig 2 (simulated data) or the upper panels in Fig 3 (wind data).

Here we consider an alternative, which was originally developed in the PhD thesis of Tajvidi (1996), and has been further developed for more general cases in Rootzén and Tajvidi (2006). The main advantage of this approach is that it includes all observations that are extreme in at least one component (later: type II). The above two works have demonstrated that the multivariate exceedances of a random vector can be approximated by a multivariate generalized Pareto distribution (MGPD). We subsequently use the abbreviation ‘MGPD’ to refer to the type II MGPD distribution. Although the mathematical theory for the MGPD model has been developed, the statistical properties of this model are not yet fully understood. An initial investigation about the performance of a basic bivariate model (BGPD) in this setting can be found in Rakonczai and Tajvidi (2010). Here we consider a wider range of BGPD models, which assume various dependence structure. We investigate their applicability and consider further extensions where possible.

Standard models for multivariate extremes are presented in Section 2, along with their role and properties in MGPD models. A new general method for constructing asymmetric versions of known symmetric models is also presented. In Section 3 we discuss a practical application of the proposed models to German wind speed data. Here we compare the performance of different BGPD models fitted to the real data and demonstrate the improvements due to using the model
construction algorithm of the previous section. Some useful conclusions and further directions are summarized in Section 4.

2 Models

In this section we present the currently used models for maxima of random vectors and discuss their applicability for threshold exceedances. Analogously to the univariate case it can be shown that if the componentwise maxima has a limit distribution, it is necessarily a multivariate extreme value distribution (MEVD). Its univariate margins are GEV distributions and there are several different (but equivalent) ways of characterizing the underlying dependence structure.

2.1 Dependence measures

As it is shown by Resnick (1987), the MEVD assuming unit Fréchet margins can be written as

\[ G_{\text{Fréchet}}(t_1, \ldots, t_d) = \exp \left( -V(t_1, \ldots, t_d) \right), \]  

(3)

with

\[ V(t_1, \ldots, t_d) = \nu((0, t_1] \times \cdots \times [0, t_d])^c) = \int_{S_d} \bigvee_{i=1}^d \left( \frac{w_i}{t_i} \right) S(dw). \]  

(4)
Here $S$ is a finite measure on the $d$-dimensional simplex $S_d = \{ w \in \mathbb{R}^d : |w| = 1 \}$ satisfying the equations

$$\int_{S_d} w_i S(dw) = 1 \quad \text{for } i = 1, ..., d,$$

where $V$ and $S$ are called exponent measure and spectral measure respectively. The symbol $\lor$ stands for the maximum (analogously, $a \land b = \min(a, b)$). In particular, the total mass of $S$ is always $S(S_d) = d$. In this context $S$ may have a density not only on the interior of $S_d$ but also on each of the lower-dimensional subspaces of $S_d$. Remark 1. Note that even if $G$ is absolutely continuous in some cases the spectral measure $S$ can put positive mass on the vertices; for instance, when the margins of $G$ are independent then $S(\{e_i\}) = 1$ for all $e_i$ vertices of the simplex. Moreover for some absolutely continuous subclasses of the MEVD $S$ may put positive mass to the vertices at arbitrary level of association (not only for independence), e.g. asymmetric logistic and asymmetric negative logistic models. In such cases there are further difficulties arising in threshold exceedance models, as we will see later in Section 2.2.

Another characterization of the MEVD, due to Pickands (1981), uses the so-called dependence function. In the bivariate setting the dependence function $A(t)$ must satisfy the following three properties: $A(t)$ is convex, $(1 - t) \lor t \leq A(t) \leq 1$ and $A(0) = A(1) = 1$. Finally, the exponent measure can be formulated by $A(t)$
Recalling the definition proposed by Rootzén and Tajvidi (2006) let $Y = (Y_1, ..., Y_d)$ denote a random vector, $u = (u_1, ..., u_d)$ be a suitably high threshold vector and $X = Y - u = (Y_1 - u_1, ..., Y_d - u_d)$ be the vector of exceedances. Then the multivariate generalized Pareto distribution (MGPD) for the $X$ exceedances can be written by a MEVD $G$ with non-degenerate margins as

$$H(x_1, \ldots, x_d) = \frac{-1}{\log G(0, \ldots, 0)} \log \frac{G(x_1, \ldots, x_d)}{G(x_1 \wedge 0, \ldots, x_d \wedge 0)}, \quad (5)$$

where $0 < G(0, \ldots, 0) < 1$. By this definition, (5) provides a model for observations that are extreme in at least one component. Similarly to (3) we may switch to unit Fréchet margins by the transformation

$$t_i = t_i(x_i) = \frac{-1}{\log G_{\xi_i, \mu_i, \sigma_i}(x_i)} = (1 + \xi_i(x_i - \mu_i)/\sigma_i)^{1/\xi_i}, \quad (6)$$
with $1 + \xi_i(x_i - \mu_i)/\sigma_i > 0$ and $\sigma_i > 0$ as in (1) for $i = 1, \ldots, d$. The MGPD density, if exists, is the form of

$$h(x) = \frac{\partial H}{\partial x_1 \cdots \partial x_d}(x) = \frac{\partial}{\partial x_1 \cdots \partial x_d} \left( 1 - \frac{\log G(x)}{\log G(0)} \right).$$

(7)

Remark 2. Note that $H_1(x) = H(x, \infty)$ is not a one dimension GPD as in (2), only the conditional distribution of $X_1 | X_1 > 0$ is GPD.

Remark 3. The interpretation of the parameters is non-standard since the parameters $\mu_i$, $\sigma_i$ and $\xi_i$ in (5) after performing (6) are not only related to the characteristics of the marginal distribution of the $i$-th component of $X$, but influence the $j$-th $(j \neq i)$ margins as well, due to the constant $\log G(0, \ldots, 0)$ appearing in the distribution function.

2.3 Construction of absolutely continuous MGPD

It is reasonable to require that in the MGPD model there is no positive probability mass on the boundary of the distribution, because otherwise the model would not remain absolutely continuous. A model, being not absolutely continuous, is hardly realistic and causes further complications for maximum likelihood estimation. The following lemma is a useful construction of an absolutely continuous MGPD model (as in (7)) from a known absolutely continuous MEVD.

Lemma 1. Let $H$ be a MGPD represented by an absolutely continuous MEVD $G$
with spectral measure $S$. $H$ is absolutely continuous if and only if $S(\text{int}(S_d)) = d$ holds, i.e. all mass is put on the interior of the simplex.

See Appendix for the proof. By Remark 1. in Section 2.1 the class of absolutely continuous MGPD type II models does not include the case of independent margins.

Since MGPD models are defined by the underlying MEVD model, and since MEVD models are defined by their dependence structures, the most popular parametric families for describing the dependence structures of MEVD models may also be considered for the MGPD models. The characteristics of these models are summarized in Table 1. The list is not exhaustive, but covers a rather wide range of families. The first and second column together show which specific parametric cases include an absolutely continuous asymmetric model, based on Lemma 1. In the next section we propose a general approach to constructing asymmetric models from symmetric ones.

### 2.4 Construction of new asymmetric models

From the third column of Table 1 we can see that there are a lack of easily computable asymmetric models, especially if all probability mass is required to be put on the interior of $S_2$, ensuring the absolute continuity of the model. The bilogistic and negative bilogistic models do not have an explicit formula for their exponent measure. The C-T model only have an explicit formula for the spectral density. In order to solve this problem we propose a methodology, which allows for the
construction of new dependence models with extra asymmetry parameter(s) from any valid models (denoted by * in Table 1). As the result of this method we may obtain more flexible asymmetric models defining a new class of absolutely continuous MGPD. Because of its mathematical simplicity we illustrate the method in the bivariate setting using the dependence function, but the same methodology can be eliminated to the higher dimensional cases as well. The algorithm is fairly simple:

i. take an absolutely continuous **baseline dependence model** from Table 1;

ii. take a strictly monotonic **transformation** \( \Psi(x) : [0, 1] \rightarrow [0, 1] \), such that \( \Psi(0) = 0, \Psi(1) = 1 \);

iii. construct a **new dependence model** from the baseline model \( A_{\Psi}(t) = A(\Psi(t)) \);

iv. **check the constraints**: \( A'_{\Psi}(0) = -1, A'_{\Psi}(1) = 1 \) and \( A''_{\Psi} \geq 0 \) (convexity).

For the construction of a feasible transformation it is natural to assume that it has a form of

\[
\Psi(t) = t + f(t)
\]  

hence

\[
\begin{align*}
\left(A(\Psi(t))\right)' &= A'(\Psi(t))\Psi'(t) = A'(\Psi(t)) \times [1 + f'(t)] \\
\left(A(\Psi(t))\right)'' &= A''(t + f(t))(1 + 2f'(t)) + A'(t + f(t))f''(t)
\end{align*}
\]
Obviously by choosing \( f \) such that \( f'(0) = f'(1) = 0 \), the \( A'_\Psi(0) = -1 \) and \( A'_\Psi(1) = 1 \) constraints are automatically fulfilled. Only the convexity needs to be checked. In general we found that the following functional form leads to a wide class of valid models

\[
 f_{\psi_1, \psi_2}(t) = \psi_1(t(1-t))^{\psi_2}, \quad \text{for } t \in [0,1],
\]

(10)

where \( \psi_1 \in \mathbb{R} \) and \( \psi_2 \geq 1 \) are asymmetry parameters. If \( \psi_1 = 0 \) we get back the baseline model. Solving the convexity inequation provides the valid range for the asymmetry parameters, see Fig 1 for illustration.

Remark 4. Of course either \( f \) can be chosen differently from (10) or the baseline dependence functions having explicit dependence functions \(^1\) can differ from logistic or negative logistic as long as the above constraints remained fulfilled. More complicated baselines as C-T model (where numerical integral is needed to obtain dependence function) or Tajvidi’s generalized logistic model (having two dependence parameters instead of one) haven’t been applied yet.

Remark 5. Note that this transformation has particular impact beyond the BGPD models as it can be universally used also for BEVD, BGPD of type I or generally in the copula of any bivariate distributions.

Remark 6. Although the transformations \( \Psi \) was defined now for the bivariate case only, it can be extended for the higher dimensional cases as well as e.g. in the

\(^1\) Therefore the bilogistic and negative bilogistic models are excluded from the list of applicable models.
trivariate case it can have a form of

$$
\Psi(t_1, t_2) = (t_1 + \psi_{1,1}(t_1([1 - t_2] - t_1))^{\psi_{2,1}}, t_2 + \psi_{2,1}(t_2([1 - t_1] - t_2))^{\psi_{2,2}}).
$$

The investigation of the properties and application of this model to data from 3 sites will be the subject of planned future work. We plan to come back to this issue in another paper soon.

When using these transformations we put the prefix "$\Psi$"- before the name of the baseline dependence model, as e.g. $\Psi$-logistic model. As a further illustration of the above method there has been a small simulation study performed for the $\Psi$-negative logistic BGPD model. 200 samples ($n = 2000$) have been generated by different known $\alpha$ and $\psi_1$ parameters (with fixed marginal parameters and $\psi_2 = 2$) and then maximum likelihood (re-)estimation of the parameters was performed. The results are summarized in the left side of Table 2 where the small deviations from the known parameter values and the relatively small standard errors verify the capability of the model construction. We also found that maximum likelihood methods perform well in general. However there is strong pairwise correlation between the parameters $\mu_i, \sigma_i, i = 1, 2$, which can be extinguished in practice by keeping e.g. one location parameter fixed during the optimization process. Estimation within this subclass gives stable convergence results.
3 Statistical inference on wind speed data

We illustrate the practical application of the proposed methods using daily wind-speed maxima at two locations in northern Germany (Bremerhaven and Schleswig) over the past five decades (1957-2007). Originally, average hourly windspeeds were recorded, but in order to reduce the serial correlation within the series, we calculated the maxima of the original observations for each day. The data for the above cities are rather strongly correlated, due to the relatively short distance between them (Kendall’s correlation is 0.57). From Fig 3 we can see that the limit of $\chi(q)$ is significantly above zero, and it appears that $\bar{\chi}(q)$ tends towards one as $q$ approaches one, so that the wind speeds of the two sites may be considered to be asymptotically dependent. (The coefficients $\chi(q)$ and $\bar{\chi}(q)$ can be interpreted as quantile dependent measures\(^2\) of dependence, and asymptotic dependence can be measured by their limits at one.) Previous studies suggest that this level of association is strong enough to ensure reasonable performance for BGPD models (Rakonczai and Tajvidi, 2010). After choosing a suitably high quantile (95%) as a threshold level for fitting the univariate GPD ($13.8$ m/s for Bremerhaven, $10.3$ m/s for Schleswig) the main aim is modeling threshold exceedances occurring at any measuring station. The bivariate data and the threshold levels are displayed in Fig 3. Though observations are considered as being stationary, non-stationarity can also be handled e.g. by choosing time-dependent model parameters as in Rakonczai et al. (2010).

\(^2\)For formulas and exact interpretation see Section 8.4 of Coles (2001).
3.1 Results for bivariate observations

The large number of model parameters (7-9), which are difficult to interpret (location and scale parameters of the underlying BEVD of the BGPD are not location and scale parameters for the BGPD margins any more) requires some transparent illustration. Hence we used bivariate prediction regions as an alternative for visualizing the estimates. A prediction region (Hall and Tajvidi, 2000) is the smallest region containing an observation with a given (usually high) probability $\gamma$. The uncertainty in the estimation can also be visualized using these prediction regions. Fig 2 shows the contours of similar regions for the BGPD type I and type II models, illustrating that indeed in the case of BGPD type II the estimators are more stable. The prediction regions of some fitted models are displayed in Fig 4. The maximum likelihood optimization has been performed by fixing the location parameter $\mu_x = 0$, as it has been suggested in the last paragraph of Section 2.4. (It turned out that for other fixed values of $\mu_x$ the algorithm has found the same maximum and resulted in indistinguishable prediction regions). In addition, in the new ($\Psi$) models we have chosen fixed $\psi_2 = 2$ values for simplicity. In these models the asymmetry parameters ($\psi_1$s) are significantly greater than zero ($\psi_1 = 0$ gives back the baseline model), providing some evidence of asymmetry in the data, whereas the dependence parameters are very close to those estimated for the baseline models. The effect of asymmetry parameters can clearly be seen on the top right panel of Fig 4, where the $\Psi$-negative logistic model has a relevant torsion in its regions to the left. Similar torsion occurs for the regions calculated from the Coles-Tawn dependence model. It is still difficult to make comparisons
based on visually very similar regions, hence a formal test would be very useful. The test we propose in this case can be easily performed as a simple byproduct of the prediction region method. The regions can be viewed as a partition of the plane. As the expected number of observations in each of these partitions is known, we can compare the theoretical frequencies with the realization e.g. by $\chi^2$ statistics. The results for Bremerhaven and Schleswig can be found in Table 3. It seems that the C-T ($\chi^2_{CT} = 9.1$) and $\Psi$-Neglog models ($\chi^2_{\Psi\text{-NegLog}} = 10.7$) perform the best. In general the new asymmetric ($\Psi$) models are significantly better than their symmetric baseline versions, $\chi^2_{Log} = 21.5$ reduces to $\chi^2_{\Psi\text{-Log}} = 16.9$ and $\chi^2_{NegLog} = 14$ reduces to $\chi^2_{\Psi\text{-NegLog}} = 10.7$. Tajvidi’s generalized symmetric logistic model ($\chi^2_{Tajvidi’s} = 11.5$) is the best among the symmetric ones.

In order to investigate the uncertainty of the estimated models a bootstrap study has been performed. Since the original distribution of the data is unknown, 100 samples have been generated and then the BGPD distribution has been re-estimated in every case. The bootstrapped dependence parameters for the $\Psi$-NegLog model are presented in the right side of Table 2. The standard errors are about the same as the ones we got by a parametric simulation study (on the left side of the same table) where known parameters have been re-estimated. The average difference between the log-likelihood values of the Neglog and $\Psi$-NegLog models (fitted to the bootstrap samples) is around 10, which shows a significant improvement due to the 2 new parameters in the likelihood ratio as well.
4 Discussion and further extensions

We have presented several models for bivariate exceedances and illustrated the practical application of these models using a real dataset on wind speed. An \texttt{R} library for fitting the proposed models, \texttt{mgpd}, has been developed as is available on \texttt{CRAN}. For a visual (and empirical) model evaluation we suggested comparing the cover rates of different prediction regions. While this is a valuable tool, the development of more formal tests, like the ones known for copula modeling (Rakonczai and Zempléni, 2008) would definitely be a further improvement. Alternative methods for comparing models, assessing the statistical significance of individual terms in the model, and assessing predictive performance should also be explored. The suggested asymmetric models are promising, as with their help one becomes able to fit asymmetric families without heavy computational burden. Our results have shown that they are real competitors for the previously used models in terms of coverage probabilities. Especially, both $\Psi$—models perform significantly better than their baseline counterparts. Probably the approach can be successfully extended into higher dimensions as well. In this paper we have focused on the new MGPD type II distributions, including the definition of new asymmetric models. Thus it was beyond the scope of this approach to evaluate the differences between the two types of MGPD. However, it turned out that the model we proposed can be estimated and effectively used for real data.
Acknowledgements

The European Union and the European Social Fund have provided financial support to the project under the grant agreement no. TÁMOP 4.2.1/B-09/KMR-2010-0003. We wish to express particular thanks to Adam Butler (Biomathematics and Statistics Scotland) as well as to the anonymous referees of Environmetrics for their very helpful suggestions and comments.

References


Model | Asym. | Density | Complications |
--- | --- | --- | --- |
Sym. Logistic | – | ✓ | – |
Asym. Logistic (Tawn, 1988) | ✓ | – | – |
Ψ-logistic* | ✓ | ✓ | convexity constraints |
Sym. Neg. Logistic | – | ✓ | – |
Asym. Neg. Logistic (Joe, 1990) | ✓ | – | – |
Ψ-negative logistic* | ✓ | ✓ | convexity constraints |
Bilogistic (Smith, 1990) | ✓ | ✓ | not explicit |
Neg. Bilogistic (C. and T., 1994) | ✓ | ✓ | not explicit |
Tajvidi’s (Tajvidi, 1996) | – | ✓ | – |
C-T (Coles and Tawn, 1991) | ✓ | ✓ | only $s(w)$ is given |
Mixed (Tawn, 1988) | – | ✓ | – |
Asym. Mixed | ✓ | – | – |

Table 1: Summary of bivariate dependence models. The asymmetric logistic and negative logistic are not absolutely continuous. Asymmetric models with valid density are the bilogistic, negative bilogistic and Dirichlet models, but all of them have further complications in calculation. New models denoted by * are proposed in Section 2.4.

<table>
<thead>
<tr>
<th>Simulation study for Ψ-negative logistic BGPD</th>
<th>Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>α = 1.3</td>
</tr>
<tr>
<td></td>
<td>( \hat{\alpha} )</td>
</tr>
<tr>
<td>( \psi_1 = 0.0 )</td>
<td>1.30 (0.04)</td>
</tr>
<tr>
<td>( \psi_1 = 0.1 )</td>
<td>1.30 (0.04)</td>
</tr>
<tr>
<td>( \psi_1 = 0.2 )</td>
<td>1.32 (0.04)</td>
</tr>
<tr>
<td>( \psi_1 = 0.3 )</td>
<td>1.31 (0.05)</td>
</tr>
</tbody>
</table>

Table 2: Parameter estimation and standard errors for simulated samples (\( n=2000 \)), number of repetitions is 200 (columns on the left side); bootstrap estimates from the wind speed data (rightmost columns).
Table 3: Number of observations between the $\gamma = 0.99, 0.95, 0.9, 0.75, 0.5$-prediction regions (graphical illustration in Fig 4.) at Bremerhaven and Schleswig. According to the $\chi^2$ test statistics (last column) the C-T and $\Psi$-Neglog models are the closest to the observations.
Figure 1: Upper panels: second derivatives of $\Psi$-logistic and $\Psi$-negative logistic dependence functions with $\psi_2 = 2$ and different parameters $\psi_1$. Lower panels: valid parameters $\psi_1$ and $\alpha$ are between the identically colored curves for a given $\psi_2$.

Figure 2: Prediction regions for simulated data from a logistic model. Left: BGPD type I, right: BGPD type II
Figure 3: Upper left: Wind speed data, with thresholds. Upper right: Density contours for BGPD type I, fitted by the evd package, no (parametric) inference on the sides. (See Stephenson (2002) and reference manual at CRAN.) Lower panels: $\chi$ and $\bar{\chi}$-plots, respectively.
Figure 4: Prediction regions for the 4 best models (having the smallest $\chi^2$-statistics in Table 3).
Appendix

Proof of Lemma 1. Suppose that \((X, Y) \sim H\) is a BGPD with lower endpoints \(l_x, l_y\) for its margins. The marginal transformations mapping \(X\) and \(Y\) into the unit Fréchet scale are \(t_1, t_2\), as in (6). Let the value of the first component of \((X, Y)\) converge to \(l_x\) assuming fixed \(y > 0\) value for the second one, then

\[
\lim_{x \to l_x} H(x, y) = \lim_{x \to l_x} \frac{\log G(x, y) - \log G(x \land 0, y \land 0)}{G(0, 0)} = \frac{\log G_s(t_1(x), t_2(y)) - \log G_s(t_1(x), t_2(0))}{G_s(t_1(0), t_2(0))} = G_1^{-1}(t_1(0), t_2(0)) \times \lim_{x \to l_x} \left[ V(t_1(x), t_2(0)) - V(t_1(x), t_2(y)) \right] = G_1^{-1}(t_1(0), t_2(0)) \times \lim_{x \to l_x} \left\{ \int_{S_2} \max \left( \frac{w}{t_1(x)}, \frac{1 - w}{t_2(y)} \right) S(dw) - \int_{S_2} \max \left( \frac{w}{t_1(x)}, \frac{1 - w}{t_2(y)} \right) S(dw) \right\} = G_1^{-1}(t_1(0), t_2(0)) \times \left\{ \frac{1}{t_2(0)} S(\{0, 1\}) + \lim_{x \to l_x} \int_{S_2 \setminus \{0, 1\}} \frac{w}{t_1(x)} S(dw) - \frac{1}{t_2(y)} S(\{0, 1\}) - \lim_{x \to l_x} \int_{S_2 \setminus \{0, 1\}} \frac{w}{t_1(x)} S(dw) \right\} = G_1^{-1}(t_1(0), t_2(0)) \times \left\{ \frac{1}{t_2(0)} - \frac{1}{t_2(y)} \right\} \times S(\{0, 1\}) = 0 \iff S(\{0, 1\}) = 0 \square
\]

Consequently, independence among the margins leads to a degenerate MGPD model without having density. This case is discussed in detail in the Section 3 of Rootzén and Tajvidi (2006).