Multivariate Stochastic Prioritization of Dependent Actuarial Risks in Agricultural Insurance

Prepared by Ezgi Nevruz¹, Kasirga Yildirak¹, Ashis SenGupta²

Presented to
ASTIN and AFIR/ERM Colloquia
20-24 August 20m-1
Panama

This paper has been prepared for the 20m-1 ASTIN and AFIR/ERM Colloquia. The Organizers wish it to be understood that opinions put forward herein are not those of the Organizers and the event Organizing and Scientific Committees are not responsible for those opinions.

©¹Hacettepe University, Department of Actuarial Sciences
²Indian Statistical Institute, Applied Statistics Unit

The Organizers will ensure that all reproductions of the paper acknowledge the author(s) and include the above copyright statement.
Abstract

Risk ranking is an important component for optimization of risk management efforts. Prioritization of environmental risks is a specific area which ranks risks geographically with fair and accurate standards. Analytic tools are not always helpful for the prioritization. In some cases, especially when we are interested in environmental risk prioritization, we may consider the geographic information of our data. Therefore, we can use geographic information system (GIS) as a tool for prioritizing risks. In this study, we aim to investigate the aggregate claims of different risk classes in agricultural insurance in terms of their comparability and orderability under the dependency assumption. For this aim, we firstly classify actuarial risks of an agricultural insurance portfolio according to spatial and temporal characteristics of hazard regions. After that, we use a stochastic ordering relation called stochastic majorization that is proposed within the framework of partial order theory. We take into account the dependency of the individual claims exposed to similar environmental risks.

Key words: Aggregate claims, Partial order theory, Schur-convexity, Stochastic majorization, Crop-hail insurance.

1. Introduction

It is assumed in the classical risk theory that individual claims are independent. However, Dhaene et al. (2002a) give some examples attesting to the fact that the independency assumption is unrealistic [2]. The individual risks may be dependent since they are exposed to similar hazards or affected by similar adverse effects such as physical or financial environment. For instance, the claims of an agricultural insurance policy are contingent on the probability of the occurrence and the consequences of the same meteorological event. The crops are subjected to the same physical environment by being produced in the same geographic area. Hereby, we discuss the prioritization of the dependent aggregate claim random variables (r.v.s) of the disjoint risk classes. We assume that the claims within the risk classes are correlated. Ambagaspitiya (1998) proposes a general method in order to determine the distribution of the aggregate claims under the assumption that there are a number of dependent classes of business [1]. We use the general vectorial definition proposed in this study.

In our study, we firstly classify actuarial risks of an agricultural insurance portfolio according to spatial and temporal characteristics of hazard regions. We take into account the dependency of the individual claims exposed to similar environmental risks. After that, we use the stochastic majorization relation which is one of the stochastic ordering relations proposed within the framework of partial order theory.
The common tendency of the existing studies for comparing portfolios of correlated risks shows that the distributional properties and the moments of the aggregate claim r.v. are useful tools for the aim of prioritizing actuarial risks. Defining a sufficient measure that includes enough information about the portfolio and reflects it accurately is one of the most important tasks for evaluating the risk. The dispersion and the correlation should be considered in addition to the mean when determining the risk measures, since the expected value of the total claim r.v. does not differ under either dependency or independency cases. Because of their representing both variation and dependency in addition to the mean, we suggest using “(standardized) generalized variance” as risk measures in this study. According to our definition of risk as an aggregate claim r.v., we use these measures in the context of multivariate analysis.

This paper is organized as follows: We introduce the multivariate representation of the actuarial risks within the framework of the collective risk models in Section 2. We provide the theory of majorization and Schur-convexity in Section 3. In Section 4, we present our data set of crop-hail insurance and explain how we organize this data for the prioritization aims and give some results for a case study. Concluding remarks are made in the last section.

2. Multivariate Representation of the Actuarial Risks

In order to introduce our model setting, let us consider a crop-hail insurance portfolio. We suppose that there are \( m \) risk classes and \( p_i \) crop classes for \( i \)-th risk class with \( i = 1, 2, \ldots, m \). The aggregate claims for the \( i \)-th risk class can be considered as a \( p_i \)-variate random vector (r.vector) as follows:

\[
S^{(i)} = (S_1^{(i)}, S_2^{(i)}, \ldots, S_{p_i}^{(i)})'.
\]

Here, the aggregate loss of the \( i \)-th risk class and \( j \)-th crop class can be represented by the r.v. \( S_j^{(i)} \) and it is obtained as

\[
S_j^{(i)} = \sum_{k=1}^{N_j} X_{jk}^{(i)},
\]

where \( N_j \) is the claim number of the \( j \)-th crop class and \( X_{jk}^{(i)} \) is the claim amount of the \( k \)-th individual claim in the \( j \)-th crop class with \( i = 1, 2, \ldots, m \), \( j = 1, 2, \ldots, p_i \), and \( k = 1, 2, \ldots, N_j \).

Here, while the risk classes are disjoint, the claims within the classes are correlated. According to our data set, the risk classes are determined with respect to their environmental similarities. Therefore, we assume that the claims exposed to similar...
environmental risks, which means the claims arising in the same risk class, are dependent.

We can compare the risk classes using the risk measures related to either the aggregate claim r-vector \( \mathbf{S}^{(i)} \) or the overall aggregate claim r.v. \( S^{(i)} = \sum_{j=1}^{p_i} S_{j}^{(i)} = \mathbf{1}' \mathbf{S}^{(i)} \). In order to compare the aggregate claim r.vectors, both the generalized variance (GV) defined as

\[
GV(\mathbf{S}^{(i)}) = \det(\mathbf{\Sigma}^{(i)}) = |\mathbf{\Sigma}^{(i)}|,
\]

and standardized generalized variance (SGV) defined as

\[
SGV(\mathbf{S}^{(i)}) = |\mathbf{\Sigma}^{(i)}|^{1/p_i}
\]

can be used as risk measures representing the overall variabilities [7]. In addition to these measures, the overall aggregate claim size r.v.s can be ordered by the coefficient of variation (CV) defined as

\[
CV(\mathbf{S}^{(i)}) = \frac{\sigma^{(i)}}{\mu^{(i)}}.
\]

Here, \( \mu^{(i)} = \mathbf{1}' \mu^{(i)} \) is the mean and \( \sigma^{(i)} = \mathbf{1}' \Sigma^{(i)} \mathbf{1} \) is the standard deviation of the r.v. \( \mathbf{S}^{(i)} \) where the mean vector of the r.vector \( \mathbf{S}^{(i)} \) is represented as

\[
\mu^{(i)} = (\mu_{1}^{(i)}, \mu_{2}^{(i)}, ..., \mu_{p_i}^{(i)})',
\]

and the covariance matrix of the r.vector \( \mathbf{S}^{(i)} \) is obtained as

\[
\Sigma^{(i)} = \begin{pmatrix}
\sigma_{1}^{(i)} & \sigma_{12}^{(i)} & \cdots & \sigma_{1p_i}^{(i)} \\
\vdots & \ddots & \ddots & \vdots \\
\sigma_{p_i1}^{(i)} & \sigma_{p_i2}^{(i)} & \cdots & \sigma_{p_ip_i}^{(i)}
\end{pmatrix}.
\]

3. Ordering Risks: Inequalities

Majorization, which is an ordering relation of real-valued vectors, turns out to be a useful tool, since we are interested in the prioritization of the aggregate claim r.vectors in this study. Because it appears within the framework of partial ordering, the vectors do not need to be totally ordered, which is very advantageous for our study.

Organizers:
3.1. Theory of Majorization

After introducing our model setting in Section 2, the following definition clearly demonstrates the convenience of the majorization relation for prioritizing the aggregate claim vectors.

**Definition 3.1.1.** For $n$-dimensional vectors $x, y \in \mathbb{R}^n$, the ordering $x \preceq_{maj} y$ denotes that $y$ majorizes $x$ (or $x$ is majorized by $y$), and it is defined by [3] as follows:

$$x \preceq_{maj} y \iff \begin{cases} \sum_{i=1}^k x_{(i)} \leq \sum_{i=1}^k y_{(i)}; & k = 1, 2, \ldots, n-1 \end{cases}$$ (1)

Here, $x_{(i)}$ denotes $i$-th component of $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ in the decreasing order, i.e. the $i$-th element of the vector $x \downarrow = (x_{[1]}, x_{[2]}, \ldots, x_{[n]})$ where $x_{[1]} \geq x_{[2]} \geq \cdots \geq x_{[n]}$.

The Condition (1) is equivalent to the condition below:

$$x \preceq_{maj} y \iff \begin{cases} \sum_{i=1}^k x_{(i)} \geq \sum_{i=1}^k y_{(i)}; & k = 1, 2, \ldots, n-1 \end{cases}$$ (2)

Here, $x_{(i)}$ denotes $i$-th component of $x = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$ in the increasing order, i.e. the $i$-th element of the vector $x \uparrow = (x_{(1)}, x_{(2)}, \ldots, x_{(n)})$ where $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$ [5].

In addition, if we use the strict inequality “$<$” instead of “$\leq$” in Condition (1), or “$>$” instead of “$\geq$” in Condition (2) for $k = 1, 2, \ldots, n-1$, the ordering is called strict majorization.

3.2. Schur-convexity

"Order-preserving" functions are very beneficial in this context, since we use risk measures defined as functions to evaluate risks. A real-valued function which preserves the ordering of the majorization is said to be a "Schur-convex" function [5]. In 1923, Schur introduced the Schur-convex function that is also known by the name "Schur-increasing function". For the risk assessment, it is important to use a measure reflecting the risk of a portfolio sufficiently and accurately. Therefore, we choose a risk measure that fulfills the properties of Schur-convexity and we use it to order the aggregate claims with the stochastic majorization relation.

**Definition 3.2.1.** A real function $\varphi: \mathcal{A} \to \mathbb{R}$ for some set $\mathcal{A} \subset \mathbb{R}^n$ is said to be Schur-convex on $\mathcal{A}$ if

$$x \preceq_{maj} y \quad \forall x, y \in \mathcal{A} \quad \Rightarrow \quad \varphi(x) \leq \varphi(y).$$ (3)
\(\varphi\) is strictly Schur-convex on \(\mathcal{A}\) if \(x \prec_{\text{maj}} y\) on \(\mathcal{A} \iff \varphi(x) < \varphi(y)\) when \(x\) is not a permutation of \(y\).

Likewise, \(\varphi\) is said to be Schur-concave on \(\mathcal{A}\) if

\[
x \preceq_{\text{maj}} y\text{ on }\mathcal{A} \iff \varphi(x) \geq \varphi(y),
\]

\(\varphi\) is strictly Schur-concave on \(\mathcal{A}\) if \(x \prec_{\text{maj}} y\) on \(\mathcal{A} \iff \varphi(x) > \varphi(y)\) when \(x\) is not a permutation of \(y\).

**Remark 3.2.2.** \(\varphi(x)\) is Schur-convex on \(\mathcal{A}\) if and only if \(-\varphi(x)\) is Schur-concave on \(\mathcal{A}\).

In order to show that a function \(\varphi: \mathcal{A} \to \mathbb{R}\) with \(\mathcal{A} \subset \mathbb{R}^n\) is Schur-convex (Schur-concave), the following theorem is needed.

**Theorem 3.2.3. (Schur’s Condition)** Suppose that \(\varphi: \mathcal{I}^n \to \mathbb{R}\) is continuously differentiable where \(\mathcal{I} \subset \mathbb{R}\) is an open interval. \(\varphi\) is Schur-convex on \(\mathcal{I}^n\) if

1. \(\varphi\) is symmetric on \(\mathcal{I}^n\), and
2. \((x_i - x_j) \left( \frac{\partial \varphi}{\partial x_i} - \frac{\partial \varphi}{\partial x_j} \right) \geq 0\) for all \(1 \leq i, j \leq n\).

Therefore, if the ordering is majorization, then “\(\varphi\) is increasing” means “\(\varphi\) is Schur-convex”.

4. Application

4.1. Data

We use the claim data provided by Agricultural Insurance Pool (TARSIM) that is a governmental institution taking the responsibility for the development of agricultural insurance in Turkey. Since 95.3% of the policies of agricultural products are crop insurance policies in 2015, we focus on crop insurance. Crop insurance covers the agricultural products exposed to various sources of risk such as hail, frost, storm and flood. Within the crop insurance products, 55.3% and 39.3% of the causes of the paid loss are hail and frost, respectively. In addition, the frost hazard is covered together with the hail hazard, not by itself. Thus, we only need to take crop insurance claims caused by the hail hazard.

After we omit the data for the policies that are cancelled, we have 975,971 crop claims (including zero claims) caused by hail in 2014. Since the information about the policy holder is not provided by TARSIM, we need to obtain the individual claims.
through village codes and lot-block codes, together. After merging the same village
codes and lot-block codes, the size of the data reduced to 831,325.

4.2. Classification of the Agricultural Claims

We classify actuarial risks of an agricultural insurance portfolio according to spatial
and temporal characteristics of hazard regions. In order to do that, we use
meteorological data as inputs such as maximum temperature, minimum surface
temperature, precipitation, snow cover, snow water equivalent etc. The data is
recorded at weather stations at various times. For the unknown values in the claim
data, "spatio-temporal interpolation" techniques are used, especially in GIS.

With the help of the spatio-temporal interpolation, the meteorological quantities of a
claim can be estimated related to its location and time. Li and Revesz (2004) show
that the extension approach for the spatio-temporal interpolation gives the most
accurate results among various interpolation algorithms [4]. This method is a 3-
dimensional approach, two for space and one for time and the missing value is
interpolated using 3-D shape functions.

4.3. Ordering of the Agricultural Claim Data

After we classify the claims according to the hazard regions and crop types, we
arrange the aggregate claim vectors for the risk class \(i = 1, 2, \ldots, m\) and crop class \(j = 1, 2, \ldots, p_i\) with regard to our setting \(\mathbf{S}(i) = (S_1^i(i), S_2^i(i), \ldots, S_p^i(i))\).

In order to compare the riskiness of the aggregate claim classes using the majorization
relation, we need both a risk measure and observations represented as vectors having
the majorization relation. Once we have a Schur-convex function being taken as a
risk measure, we need to check if the majorization relation exists between vectors.
After that, we can order the functional values and infer that the riskiness of the classes
can also be ordered similarly.

In order to prioritize the aggregate claim vectors, we firstly check the properties for
the majorization given in Definition 3.1.1. Then, we use a risk measure which fulfils the
conditions of Schur-convexity provided in Theorem 3.2.3.

Risk measures are classified into two types, as safety risk measures evaluating wealth
under risk and as dispersion measures assessing the uncertainty level [6]. Since our aim
for the future is to associate this study with “decision under uncertainty”, we choose
the second class of risk measures.

In the first phase, we use the sample variance as a risk measure, which is one of the
main dispersion measures. The sample variance is a dispersion measure used for
ordering the aggregate claims of $m$ risk classes through majorization relation. In order to do that, we set the prioritization of the aggregate claims as follows:

$$S^{(k)} \preceq_{maj} S^{(l)} \iff \varphi \left( \psi(S^{(k)}) \right) \leq \varphi \left( \psi(S^{(l)}) \right),$$

(5)

where $\varphi$ is a Schur-convex function. We order the aggregate claim vectors of the $k$-th and $l$-th risk classes considering an ordering between the function values of their variances.

### 4.3.1. Schur-convexity of a risk measure

According to Marshall et al. (2009), the sample variance defined as

$$\varphi_1(x) = \varphi_1(x_1, x_2, ..., x_n) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

is strictly Schur-convex with respect to $x = (x_1, x_2, ..., x_n)$.

In order to show that $\varphi_1$ fulfills the Schur's conditions, we firstly check whether it is symmetric or not. Since this function gives the same values for all permutations of $x$, it is symmetric. For all permutations $(x_1, x_2, ..., x_n)$, the function $\varphi_1(x_1, x_2, ..., x_n)$ is the same because $(x_i - \bar{x})^2$ is the same.

Secondly, we need to show that $(x_i - x_j) \left( \frac{\partial \varphi_1}{\partial x_i} - \frac{\partial \varphi_1}{\partial x_j} \right) > 0$ for all $1 \leq i, j \leq n$ in order to prove that $\varphi_1(x)$ is strictly Schur-convex. Let $x_1 > x_2$. Then,

$$\frac{\partial \varphi_1}{\partial x_1} = \frac{1}{n} \left[ 2 \left( x_1 - \frac{x_1 + \cdots + x_n}{n} \right) \left( 1 - \frac{1}{n} \right) + 2 \left( x_2 - \frac{x_1 + \cdots + x_n}{n} \right) \left( -\frac{1}{n} \right) + \cdots + 2 \left( x_n - \frac{x_1 + \cdots + x_n}{n} \right) \left( -\frac{1}{n} \right) \right]$$

and

$$\frac{\partial \varphi_1}{\partial x_2} = \frac{1}{n} \left[ 2 \left( x_1 - \frac{x_1 + \cdots + x_n}{n} \right) \left( -\frac{1}{n} \right) + 2 \left( x_2 - \frac{x_1 + \cdots + x_n}{n} \right) \left( 1 - \frac{1}{n} \right) + \cdots + 2 \left( x_n - \frac{x_1 + \cdots + x_n}{n} \right) \left( -\frac{1}{n} \right) \right]$$

$$\Rightarrow \frac{\partial \varphi_1}{\partial x_1} - \frac{\partial \varphi_1}{\partial x_2} = \frac{2}{n} \left[ \left( x_1 - \frac{x_1 + \cdots + x_n}{n} \right) - \left( x_2 - \frac{x_1 + \cdots + x_n}{n} \right) \right] = \frac{2}{n} (x_1 - x_2)$$

Organizers:
Since $x_1 > x_2$, $x_1 - x_2 > 0$ and $\frac{\partial \varphi_1}{\partial x_1} - \frac{\partial \varphi_2}{\partial x_2} = \frac{2}{n} (x_1 - x_2) > 0$. This is true for all $1 \leq i, j \leq n$. Therefore, $\varphi_1(x)$ is strictly Schur-convex with respect to $x = (x_1, x_2, \ldots, x_n)$.

4.3.2. Rearrangement of the aggregate claim vectors

According to our main setting given in Equation (5), we rewrite the majorization definition as follows:

$$S^{(k)} \preceq_{\text{maj}} S^{(l)} \text{ iff } \begin{cases} \sum_{i=1}^{n} i S^{(k)}_i \leq \sum_{i=1}^{n} i S^{(l)}_i ; p_i = 1,2, \ldots, m - 1 \\ \sum_{i=1}^{m} S^{(k)}_i = \sum_{i=1}^{m} S^{(l)}_i \end{cases}$$

(6)

The aggregate claim vector $S^{(k)}$ for the $k$-th risk class is majorized by the aggregate claim vector $S^{(l)}$ for the $l$-th risk class if and only if the right-hand side of the equation is true. Here, we rearrange the aggregate claim vectors by sorting them into descending order.

It is obvious that the condition $\sum_{i=1}^{m} S^{(k)}_i = \sum_{i=1}^{m} S^{(l)}_i$ is very unlikely to be fulfilled because aggregate claims are continuous random values. Thus, in order to overcome this problem, we redefine the majorization relation as follows:

$$\begin{align*} S^{(k)} & \preceq_{\text{maj}} S^{(l)} \text{ iff } \begin{cases} \frac{\sum_{i=1}^{n} i S^{(k)}_i}{\sum_{i=1}^{m} S^{(k)}_i} \leq \frac{\sum_{i=1}^{n} i S^{(l)}_i}{\sum_{i=1}^{m} S^{(l)}_i} ; p_i = 1,2, \ldots, m - 1 \\ \sum_{i=1}^{m} S^{(k)}_i = \sum_{i=1}^{m} S^{(l)}_i = 1 \end{cases} \\ \sum_{i=1}^{m} S^{(k)}_i & = \sum_{i=1}^{m} S^{(l)}_i \end{align*}$$

(7)

We divide all the elements in Equation (6) by the summations to obtain Equation (7). Thus, we have two conditions to be checked in order to show that $S^{(k)}$ is majorized by $S^{(l)}$:

i. $\frac{\sum_{i=1}^{n} i S^{(k)}_i}{\sum_{i=1}^{m} S^{(k)}_i} \leq \frac{\sum_{i=1}^{n} i S^{(l)}_i}{\sum_{i=1}^{m} S^{(l)}_i}$

ii. $\sum_{i=1}^{m} S^{(k)}_i \geq \sum_{i=1}^{m} S^{(l)}_i$

4.4. Case Study

We determine 23 risk classes and 18 crop classes in our data set. We select 4 risk classes (8-th, 9-th, 19-th and 20-th) as a case study here. Considering the first and second
conditions of Equation (7), it is obtained that there is a strict majorization relation among these 4 classes. We find the following results.

\[
\sum_{j=1}^{18} S_{[j]}^{(8)} > \sum_{j=1}^{18} S_{[j]}^{(9)} > \sum_{j=1}^{18} S_{[j]}^{(19)} > \sum_{j=1}^{18} S_{[j]}^{(20)}, \quad \text{and}
\]

\[
\sum_{j=1}^{18} p_j S_{[j]}^{(8)} < \sum_{j=1}^{18} p_j S_{[j]}^{(9)} < \sum_{j=1}^{18} p_j S_{[j]}^{(19)} < \sum_{j=1}^{18} p_j S_{[j]}^{(20)}.
\]

Thus, \( S^{(8)} \prec_{\text{maj}} S^{(9)} \prec_{\text{maj}} S^{(19)} \prec_{\text{maj}} S^{(20)} \) is true for this sample.

Lastly, we check the Schur-convex functional values for these classes given in Table 4.1.

<table>
<thead>
<tr>
<th>Risk Class</th>
<th>( \phi_1 \left( \mathcal{V}(S^{(j)}) \right) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>j=8</td>
<td>0.008659253880799</td>
</tr>
<tr>
<td>j=9</td>
<td>0.015409307456010</td>
</tr>
<tr>
<td>j=19</td>
<td>0.027543754596482</td>
</tr>
<tr>
<td>j=20</td>
<td>0.047684139428587</td>
</tr>
</tbody>
</table>

Table 4.1. The sample variance values of the variance of aggregate claim vectors

From Table 4.1, the same ordering direction exists among risk classes. Therefore, we can conclude that the 8-th class is the least risky class and the 19-th class is the most risky class:

\[
S^{(8)} \prec_{\text{maj}} S^{(9)} \prec_{\text{maj}} S^{(19)} \prec_{\text{maj}} S^{(20)}
\]

\[
\Leftrightarrow \phi \left( \mathcal{V}(S^{(8)}) \right) < \phi \left( \mathcal{V}(S^{(9)}) \right) < \phi \left( \mathcal{V}(S^{(19)}) \right) < \phi \left( \mathcal{V}(S^{(20)}) \right).
\]

5. Conclusion and Future Study

We choose a sample for a case study and we use sample variance as a risk measure assessing the dispersion in the chosen sample for our portfolio. For the future studies, we will firstly look at the overall portfolio considering the multiple majorization conditions. Another goal of this study is to consider different risk measures especially the ones related to the population. We will try to prove the Schur’s conditions for different risk measures such as population variance, coefficient of variation etc. Speaking of overall portfolio, stop-loss dominance is another stochastic relation reflecting the riskiness of the aggregate claims. Therefore, we will lastly evaluate whether stop-loss premiums can be used in the frame of majorization.
References


