Insurance: Mathematics and Economics 68 (2016) 84–91

Contents lists available at ScienceDirect

Insurance: Mathematics and Economics

journal homepage: www.elsevier.com/locate/ime

Ordering Gini indexes of multivariate elliptical risks

Ranadeera Gamage Madhuka Samanthi, Wei Wei *, Vytaras Brazauskas

Department of Mathematical Sciences, University of Wisconsin-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA

ARTICLE INFO

Article history:
Received August 2015
Received in revised form
January 2016
Accepted 7 March 2016
Available online 15 March 2016

Keywords:
Elliptical distribution
Dependence structure
Comonotonicity
Usual stochastic order
Increasing convex order
Supermodular order
Pd-1 order

ABSTRACT

Gini index is a well-known tool in economics that is often used for measuring income inequality. In insurance, the index and its modifications have been used to compare the riskiness of portfolios, to order reinsurance contracts, and to summarize insurance scores (relativities). In this paper, we establish several stochastic orders between the Gini indexes of multivariate elliptical risks with the same marginals but different dependence structures. This work is motivated by the applied studies of Brazauskas et al. (2007) and Samanthi et al. (2015), who employed the Gini index to compare the riskiness of insurance portfolios. Based on extensive Monte Carlo simulations, these authors have found that the power function of the associated hypothesis test increases as portfolios become more positively correlated. The comparison of the Gini indexes (of empirically estimated risk measures) presented in this paper provides a theoretical explanation to this statistical phenomenon. Moreover, it enriches the studies of the problem of central concentration of elliptical distributions and generalizes the pd-1 order proposed by Shaked and Tong (1985).

1. Introduction

Over a hundred years ago, Corrado Gini introduced an index to measure concentration or inequality of incomes (see Gini, 1936, for English translation of the original article). It later became known as the Gini index and has been extensively studied in many fields such as economics, insurance, finance, and statistics. At the intersection of insurance and statistics, for example, the index has been used for comparing distributions of risks and prices (see Frees et al., 2011). The comparisons are usually based on insurance scores relative to price, also known as "relativities", that point to areas of potential discrepancies between risk and price distributions. After ordering both risks and prices based on relativities, one arrives at an ordered Lorenz curve that can be summarized using a Gini index. Interestingly, the Lorenz curve and Gini index defined via relativities can cope with adverse selection, help measure potential profit, and serve as useful tools in predictive modeling (for more information, see Frees et al., 2014). Moreover, Lorenz curve and Lorenz order, the concepts closely related to Gini index, have been employed by Denuit and Vermandele (1999) to order reinsurance contracts.

Other statistical applications in insurance have emphasized the fact that the Gini index is an *L*-statistic, theoretical properties of which are well-established and thus can be employed to construct statistical inferential tools. For instance, Jones and Zitikis (2005), Jones et al. (2006), and Brazauskas et al. (2007) have designed several hypothesis tests to compare the riskiness of insurance portfolios by using the Gini index. Samanthi et al. (submitted for publication) have conducted an extensive simulation study by incorporating various types of dependence between portfolios and found that the power function of the associated hypothesis test increases as portfolios become more positively correlated. The comparison of the Gini indexes (of empirically estimated risk measures) presented in this paper provides a theoretical explanation to this statistical phenomenon.

As described by Samanthi et al. (submitted for publication), the power function of the hypothesis test under consideration is a probability event involving the Gini index $\frac{1}{n^2}\sum_{1\le i,j\le n}|X_i - X_j|$, where the random variables $X_1, \ldots, X_n$ represent empirical risk measures estimated from observations on n insurance portfolios. It is also known that the random vector $\mathbf{X} = (X_1, \ldots, X_n)$ follows an asymptotically multivariate normal distribution. For more details about the design of the hypothesis test, the reader may be referred to Samanthi et al. (submitted for publication) and Brazauskas et al. (2007). In order to explain the monotonicity of the test power function, with respect to the strength of dependence, we propose the following conjecture.

Conjecture 1.1. *Let $(X_1, \ldots, X_n)$ follow a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, i.e., $(X_1, \ldots, X_n) \sim$*

* Corresponding author.
E-mail addresses: samanthi@uwm.edu (R.G.M. Samanthi), weiw@uwm.edu (W. Wei), vytaras@uwm.edu (V. Brazauskas).

http://dx.doi.org/10.1016/j.insmatheco.2016.03.005
0167-6687/

$MVN(\mathbf{0}, \mathbf{\Sigma})$. Then its Gini index $\frac{1}{n^2}\sum_{1\le i,j\le n}|X_i - X_j|$ decreases in the sense of usual stochastic order (see Section 2 for definition) as the covariance matrix $\mathbf{\Sigma}$ increases componentwise with diagonal elements remaining unchanged.

Basically, Conjecture 1.1 aims to order Gini indexes of multivariate normal risks with same marginals but different strength of dependence. Proving Conjecture 1.1 is a challenging task. This paper partially completes this task and generalizes the conclusion to elliptical distributions, yet still leaves some open problems.

Besides its usefulness in actuarial applications, the comparison of Gini indexes of multivariate elliptical risks shows its own independent interest. Intuitively, Conjecture 1.1 suggests that the probability $\mathbb{P}\left\{\frac{1}{n^2}\sum_{1\le i,j\le n}|X_i - X_j| \le t\right\}$ increases as $\mathbf{\Sigma}$ increases for any $t \ge 0$. In this sense, the study of Conjecture 1.1 falls into the scope of the problem of central concentration of elliptical distributions, which is formulated as follows: how the probability

$$\mathbb{P}_\Sigma(C) = \mathbb{P}\{(X_1, \ldots, X_n) \in C\} \tag{1.1}$$

changes according to the change of $\mathbf{\Sigma}$, where $(X_1, \ldots, X_n)$ follows an elliptical distribution with mean $\mathbf{0}$ and dispersion matrix $\mathbf{\Sigma}$?

This problem was first studied by Slepain (1962), which states that if $(X_1, \ldots, X_n)$ follows a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{\Sigma}$, then $\mathbb{P}_\Sigma(C)$ increases as $\mathbf{\Sigma}$ increases componentwise with diagonal elements remaining unchanged for any lower orthant set C. Later literature has generalized the study to elliptical distributions while regions of different shapes have been considered, such as upper orthant sets, rectangles, and convex and centrally symmetric regions. Interested readers are referred to Das Gupta et al. (1972), Joe (1990), Eaton and Perlman (1991), and Anderson (1996). All these studies imposed certain assumptions on the structure of the covariance matrix. The results derived in this paper enriches the studies on this problem in the sense that it broadens the choice of the set C.

In addition, comparison of Gini index has another fold of meaning. The methodologies can be used to generalize the pd-1 order. The pd-1 order was proposed by Shaked and Tong (1985) and used to compare the strength of dependence of exchangeable random vectors. Chang (1992) extended this concept from the perspective of stochastic majorization and explored applications in operations research. Readers are referred to Chapter 9 of Shaked and Shanthikumar (2007) for a conclusive summary of the studies on the pd-1 order. In the existing literature, the application of pd-1 order is very restrictive since the comparison applies only to exchangeable random vectors. In this paper, we manage to generalize the pd-1 order to non-exchangeable random vectors.

The rest of the paper is organized as follows. Section 2 introduces some basics about stochastic orders, elliptical distributions, and comonotonicity. Section 3 compares Gini indexes of multivariate elliptical risks in the sense of a relatively weaker order: the increasing convex order. Section 4 imposes certain assumptions on the structure of covariance matrices and establishes the usual stochastic orders between Gini indexes. In Section 5, we discuss the pd-1 order and its generalization by using similar techniques before. Section 6 provides concluding remarks of the paper.

2. Preliminaries

Throughout the paper, we use bold letters to denote vectors or matrices. For example, $\mathbf{x} = (x_1, \ldots, x_n)$ is a row vector and $\mathbf{\Sigma} = (\sigma_{ij})_{n\times n}$ is an $n \times n$ matrix. In particular, the symbol $\mathbf{0}$ denotes the row vector with all entries equal to 0, and $\mathbf{1}_{n\times n}$ denotes the $n\times n$ matrix with all entries equal to 1. The inequality between vectors or matrices denotes componentwise inequalities. For example, $(x_1, \ldots, x_n) \le (y_1, \ldots, y_n)$ implies that $x_i \le y_i$ for all $i = 1, \ldots, n$.

Consider a random vector $\mathbf{X} = (X_1, \ldots, X_n)$. Its Gini index is defined to be $\frac{1}{n^2}\sum_{1\le i,j\le n}^{n}|X_i - X_j|$. Gini index measures how dispersive the components of the random vector are. For example, if all the components are identical, then the Gini index is 0, which indicates a perfect concentration. For notational convenience, denote

$$G(\mathbf{X}) = \sum_{1\le i,j\le n}^{n} |X_i - X_j|. \tag{2.1}$$

$G(\mathbf{X})$ is the scaled Gini index and is the random variable we shall study throughout the paper. It is easy to see that $G(\mathbf{X})$ can be rewritten in terms of order statistics as follows.

$$G(\mathbf{X}) = \sum_{i=1}^{n}(4i - 2n - 2)X_{(i)}, \tag{2.2}$$

where $X_{(i)}$ denotes the ith largest component of $\{X_1, \ldots, X_n\}$.

In order to compare Gini indexes, we recall definitions of some stochastic orders.

Definition 2.1. Let X and Y be two random variables.

X is said to be smaller than Y in *usual stochastic order*, denoted as $X \le_{st} Y$, if $\mathbb{P}\{X > t\} \le \mathbb{P}\{Y > t\}$ for all $t \in \mathbb{R}$.

X is said to be smaller than Y in *increasing convex order*, denoted as $X \le_{icx} Y$, if $\mathbb{E}[u(X)] \le \mathbb{E}[u(Y)]$ for any increasing convex function u such that the expectations exist.

The above definitions are taken from Shaked and Shanthikumar (2007), which also provide the following characterization for the usual stochastic order.

Proposition 2.2. *Let X, Y be two random variables. $X \le_{st} Y$ if and only if $\mathbb{E}[u(X)] \le \mathbb{E}[u(Y)]$ for any increasing function u such that the expectations exist.*

Furthermore, in order to compare random vectors, the concept of supermodular order is needed. There is rich literature on the subject of supermodular order, see, for example, Marshall et al. (2010), Müller and Stoyan (2002), and Shaked and Shanthikumar (2007). We cite the definition of supermodular function and supermodular order from Shaked and Shanthikumar (2007).

Definition 2.3. A function $f : \mathbb{R}^n \to \mathbb{R}$ is said to be *supermodular* if for any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ it holds that

$$f(\mathbf{x}) + f(\mathbf{y}) \le f(\mathbf{x} \wedge \mathbf{y}) + f(\mathbf{x} \vee \mathbf{y}),$$

where the operators $\wedge$ and $\vee$ denote coordinatewise minimum and maximum respectively, i.e.,

$$(x_1, \ldots, x_n) \wedge (y_1, \ldots, y_n) = (\min(x_1, y_1), \ldots, \min(x_n, y_n)),$$
$$(x_1, \ldots, x_n) \vee (y_1, \ldots, y_n) = (\max(x_1, y_1), \ldots, \max(x_n, y_n)).$$

Random vector $\mathbf{X}$ is said to be smaller than random vector $\mathbf{Y}$ in the *supermodular order*, denoted as $\mathbf{X} \le_{sm} \mathbf{Y}$, if $\mathbb{E}[f(\mathbf{X})] \le \mathbb{E}[f(\mathbf{Y})]$ for any supermodular function f such that the expectations exist.

It is easy to verify that, if $\mathbf{X} \le_{sm} \mathbf{Y}$ and $\mathbf{X} \ge_{sm} \mathbf{Y}$, then $\mathbf{X} \stackrel{d}{=} \mathbf{Y}$, where $\stackrel{d}{=}$ denotes "equal in distribution". According to Kemperman (1977, Assertion (i)), we have the following result.

Proposition 2.4. *A function $f : \mathbb{R}^n \to \mathbb{R}$ is supermodular if and only if $f(x_1, \ldots, x_n)$ is supermodular as a function of (x_i, x_j) for any other fixed x_k, $k \ne i, j$ for any $1 \le i < j \le n$.*