Journal of the American Statistical Association

ISSN: 0162-1459 (Print) 1537-274X (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Extreme and Inference for Tail Gini Functionals with Applications in Tail Risk Measurement

Yanxi Hou & Xing Wang

To cite this article: Yanxi Hou & Xing Wang (2020): Extreme and Inference for Tail Gini Functionals with Applications in Tail Risk Measurement, Journal of the American Statistical Association, DOI: 10.1080/01621459.2020.1730855

To link to this article: https://doi.org/10.1080/01621459.2020.1730855

Extreme and Inference for Tail Gini Functionals with Applications in Tail Risk Measurement

Yanxi Hou*

School of Data Science, Fudan University

Xing Wang

Department of Mathematics, Illinois State University

*Corresponding author: yxhou@fudan.edu.cn.

Abstract

Tail risk analysis focuses on the problem of risk measurement on the tail regions of financial variables. As one crucial task in tail risk analysis for risk management, the measurement of tail risk variability is less addressed in the literature. Neither the theoretical results nor inference methods are fully developed, which results in the difficulty of modeling implementation. Practitioners are then short of measurement methods to understand and evaluate tail risks, even when they have large amounts of valuable data in hand. In this article, we consider the measurement of tail variability under the tail scenarios of a systemic variable by extending the Gini's methodology. As we are very interested in the limit of the proposed measures as the risk level approaches to the extreme status, we showed, by using extreme value techniques, how the tail dependence structure and marginal risk severity have influences on the limit of the proposed tail variability measures. We construct a nonparametric estimator, and its asymptotic behavior is explored. Furthermore, in order to provide practitioners with more measures for tail risk, we construct three coefficients/measures for tail risks from different views toward tail risks and illustrate them in a real data analysis.

Keywords: extreme value theory; tail copula; tail risk; Gini mean difference; measure of variability

1 Introduction

Tail-based risk measurement has been proposed in the literature to explore the *tail risk* of financial variables. The most well-known risk measures, both in the academy and industry, are the *Value-at-Risk* (VaR) and *Expected Shortfall* (ES), which are the prominent tools for tail risk analysis in the regulatory framework of insurance and banking sectors. See [3, 4, 32, 37] for discussions on tail risk and its measurement in financial econometrics. Recently, the financial crisis has led regulators and practitioners to pay attention to the impact of *systemic risk* within the financial industry. To analyze the inter-influence of risks from individual financial institutes, researchers have also studied abundant risk measures on systemic risks during the last decade. [1, 2, 5, 8, 9] discuss the impact of systemic risk and propose various new approaches to explore the risk analysis of extreme financial co-movements.

Fruitful studies focus on modeling and measuring methods of systemic risk. [2] introduces the *Conditional Value at Risk* (CoVaR) for measuring systemic risk in financial systems and applies it to some econometric models for empirical analysis. [1] proposes a systemic risk measure based on the expected loss of a financial institution's equity return conditional on the occurrence of an extreme loss in the aggregated return of the financial sector. [5] designs an aggregate risk index of systemic risk, *CATFIN*, which associates systemic risk to the VaR of the financial system. [8] provides several econometric measures of systemic risk based on network structure for financial systems. [9] proposes *SRISK* as a market-based measure of the capital shortfall of a firm conditional on a severe market decline. On the aspect of mathematical and statistical methodology, [10, 11] further develop the asymptotic properties of the risk measures in [1] under different conditions of tail heaviness by using Extreme Value Theory and their methods are very useful in statistical applications. [27] constructs a risk measure by taking into account both ES and tail dependence of two financial variables, to measure the relative risk of individual financial entity to some benchmark as a sensitive monitoring index of the market co-movement. [7] studies the extremal and asymptotic properties of coherent risk measures conditional on a systemic variable by assuming a tail comparison condition.

Mathematically speaking, one promising way is to incorporate the dependence or tail dependence structures into the modeling of extreme co-movement. The key issue is to combine the tail dependence structure with tail-based risk measurement, and properties of these newly proposed risk measures are of great interest in both theory and practice. Let (X, Y) be a pair of random loss variables, respectively, on an individual variable with interest and a benchmark variable. (X, Y) has a joint continuous distribution F with continuous marginal distributions $F_i, i = 1, 2$ respectively. Define the quantile function of F_i as $Q_i(1-\alpha) := F_i^{\leftarrow}(1-\alpha) := \inf\{x | F_i(x) \geq 1-\alpha\}$ where $\alpha \in (0,1)$ is the *risk level*. Consider some tail scenarios of the benchmark variable Y as the event $(Y > Q_2(1-\alpha))$ at level α. To quantify the uncertainty of the interesting variable X under the impact of systemic risk Y, a direct way is to model risk measures for X given the tail scenarios of Y with certain level α, which takes into account both the co-movement impact (dependence between X and Y) and risk severity (marginal quantity of X).

One issue in the above modeling is the assumption on level α. The level α is typically close to zero to indicate extreme scenarios or high-risk events of Y. Different from the traditional treatment to assume α a fixed constant, it is useful to model α as a sequence of index n, which classifies α into two categories: the intermediate level where $\alpha := \alpha_n$ satisfies $\alpha \to 0$ and $n\alpha \to \infty$ as $n \to \infty$, and the extreme level where $\alpha \to 0$ and $n\alpha \to c \in [0, \infty)$ as $n \to \infty$. Intermediate levels assume enough sample size for inference on the tail regions, while extreme levels usually require extrapolation techniques based on results of intermediate levels due to lack of sample on tail regions. In this paper, we consider both the intermediate level α and a much smaller level $p = p_n$ such that $p/\alpha \to [0,1]$ as $n \to \infty$. In real applications, an extremely small p usually implies extreme events of the systemic variable where the sample size is not big enough for researchers to make inference for risk measures.

When new ideas are concentrated on the tail risk analysis of systemic risk, less is explored for measuring the variability of the risk on tail regions. Indeed, measuring the variability of risks is another important task in risk measurement. As early as [29], Harry Markowitz has considered the variability of risks in his celebrated publication. When it comes to the tail analysis of a univariate financial variable, [19] defines a *tail-standard-deviation* (TSD) risk measure to incorporate variability on tail regions. However, TSD lacks some fundamental properties in the theory of risk measures, which is not a good risk measure in practice; see Section 1 of [20] for details. In order to solve the drawbacks of TSD, [20] proposes a Gini-type measure for both risk and variability, whose method originates from the *Gini mean difference* and the *Gini index* in a series of studies [21, 22, 23]. See [26] for a recent review of applications and inference methods of the Gini index. In welfare economics, the Gini index is a well-known concentration index for measuring the inequality in the distribution of income and wealth, which has its natural relationship with the *Lorenz curve*. However, the contributions and extensions of the Gini index and its related methods lie not only in the field of economics but also in other fields such as statistics, medicines, biology, and ecology. When it comes to risk analysis, [17, 18] generalize an ordered Lorenz curve and the associated Gini index for comparing the distribution of loss and premium, and the asymptotic property of the Gini index is studied. Another generalization of the Gini mean difference in the risk management application is the Gini-type measure in [20]. As a measure of risk variability, the Gini-type measure has many advantages over the standard deviation or tail standard deviation that we will state in the following sections. For more details about Gini's methodology, we refer to [38].

However, neither the TSD nor the Gini-type measure serves for measuring the tail variability given the impact from systemic variables, as it is challenging to incorporate the tail dependence in the risk modeling. In this paper, we extend the Gini's methodology to risk measurement of tail variability based on the tail scenarios of a benchmark variable, and further explore the extremes of the so-called *Tail Gini Functionals* (TGF) based on some fundamental results in Extreme Value Theory. Besides, we establish a statistical approach for TGF and prove the asymptotic properties based on the tail empirical processes and the tail empirical copula processes. Finally, we propose three tail coefficients/measures, the *Coefficient of Tail Gini Co-variation* (TGV), *Tail Gini Correlation* (TGR) and *Marginal Gini Shortfall* (MGS). These measures are constructed from different views toward tail risks given tail scenarios of a systemic variable. They can serve as some useful measures for tail risk analysis of systemic risks in the empirical analysis. Table 1 above summarizes the main measures/coefficients developed and applied in this paper.

We organize the rest of the paper as follows. Section 2 presents the methodology for measuring the conditional tail variability of risks and derives the extreme limits of TGF. To apply the new tail variability measure, we propose a nonparametric inference method and prove its asymptotic normality as well. In Section 3, simulated cases are conducted to understand and verify the properties of TGF as well as its estimator. In Section 4, we further define the three measures constructed from MES and TGF. The measurement issue is suggested by using the nonparametric estimators of MES and TGF. We implement a real data analysis for illustration. Finally, the conclusion is made in Section 5. Some additional simulation studies, real data analyses, and all the proofs are given in the supplementary file.

2 Main Results for Tail Gini Functionals

Measures of risk and variability are functionals mapping a random variable to some value in a sub-interval of $(-\infty,\infty]$. Measures of variability quantify the magnitude of variability of random variables. Denote $\mathcal{X}$ as a collection of loss variables and ρ as a measure of variability, that is, a functional from $\mathcal{X}$ to $[0,\infty]$. There are many desirable axiomatic properties described for a measure of variability ρ, which are systematically presented in [33] (under the name " deviation measure"):

P1 *Law invariance*: if Z and Z' are some random variables in $\mathcal{X}$ sharing the same distribution F_Z, then $\rho(Z)=\rho(Z')$.

P2 *Standardization*: $\rho(c)=0$ for all $c\in\mathbb{R}$.

P3 *Location invariance*: $\rho(Z-c)=\rho(Z)$ for all $c\in\mathbb{R}, Z\in\mathcal{X}$.

P4 *Positive homogeneity*: $\rho(\lambda Z) = \lambda\rho(Z)$ for all $\lambda > 0, Z \in \mathcal{X}$.

P5 *Sub-additivity*: $\rho(Z+W) \le \rho(Z) + \rho(W)$ for any two random variables $Z, W \in \mathcal{X}$.

A measure is called a *measure of variability* if P1-P3 are satisfied. A measure is further called a *coherent* measure of variability if P4 and P5 are also satisfied. The coherence principle for measures of risk is discussed in [6], which is different from that of a variability measure as the purpose of the two types of measures are different. It is easy to verify that the standard deviation is a coherent measure of variability under the condition that the random variable has a finite second moment. For more properties and theories about measures of risk and variability, one can find them in [30] and reference therein.

Gini-type measures originate from the work of Corrado Gini, which is discussed in detail in [24, 25, 12]. The *Gini mean difference* of a random variable *Z* with cumulative distribution F_Z such that $\mathbb{E}|Z| < \infty$ is defined as

$$\mathrm{Gini}(Z) = \mathbb{E}|Z - Z'|,$$

where Z' is an independent copy of *Z*. Note that Gini mean difference is a coherent measure of variability (see Corollary 3.1 of [20]), and it does not require the random variable to have a finite second moment, which, as a variability measure, encompasses some random variables which cannot be measured by the standard deviation. In addition, the Gini mean difference has many other equivalent representations. One important representation is the covariance-based formula,

$$\mathrm{Gini}(Z) = 4\mathrm{Cov}(Z, F_Z(Z)). \quad (2.1)$$

It is interesting to note that in the literature (2.1) is widely recognized and developed as a variability measure. See [23, 36, 16, 28] for the evolution of the covariance-based formula and its applications in econometrics. Representation

(2.1) is a mixture of the variable and its distribution, which is similar to the variance $\mathrm{Var}(Z)$ as a variability measure, but different in the sense that it reduces the marginal influence in the measure by replacing one variate Z with $F_Z(Z)$. Besides, when it comes to a pair of random variables, the covariance-based formula is very useful in measuring association. Recall that (X, Y) is a random vector with a continuous joint distribution F and continuous marginal distributions $F_i, i=1,2$. By Sklar's theorem, it is equivalent to knowing the marginal distributions F_i and survival copula C, a decomposition of the joint distribution F that C contains all the information about the dependence structure of (X, Y), and F_i contains only the information about the marginal risk severity of X and Y respectively. Thus, to measure the association between X and Y, there are three covariance measures available immediately: $\mathrm{Cov}(X,Y), \mathrm{Cov}(F_1(X), F_2(Y))$ and $\mathrm{Cov}(X, F_2(Y))$ (or $\mathrm{Cov}(F_1(X), Y)$). These three covariance measures correspond to the well-known Pearson's correlation, Spearman's rho and Gini's gamma in turn. By applying Sklar's Theorem, $\mathrm{Cov}(X,Y)$ incorporates C, F_1 and F_2; $\mathrm{Cov}(F_1(X), F_2(Y))$ only incorporates C; and $\mathrm{Cov}(X, F_2(Y))$ incorporates C and F_1, but no F_2. See [31] for more details about association measures. Thus, these three covariance measures intuitively tell us how the information about the dependence structure and marginal severity is modeled in their representations.

A good tail risk variability measure is supposed to incorporate both the marginal risk severity of X and the tail dependence structure of (X, Y) given tail scenarios of Y, and the marginal risk severity of Y should not disturb the measure either. Hence, the traditional Gini's gamma is most suitable for the extension and the covariance-based formula is very simple to interpret the tail risk variability. In practice, X may be some interesting variable, say, loss of an individual asset or portfolio; while Y could be a benchmark variable, say, loss of a financial system or benchmark portfolio. Here we talk of tail/extreme scenarios of the systemic variable Y by calling the extreme events $(Y > Q_2(1-\alpha))$, which implies Y is under distress with a certain risk level $\alpha \in (0,1)$. Based on (2.1), we introduce the measure of tail risk variability for systemic risks as follows.

Definition 1. The Tail Gini Functional of a random vector (*X*, *Y*) is defined as

$$\mathrm{TG}_{\alpha}(X;Y) = \frac{4}{\alpha}\mathrm{Cov}(X, F_2(Y) \mid F_2(Y) > 1-\alpha), \qquad (2.2)$$

where F_2 is the distribution of *Y* and $\alpha \in (0,1)$.

It is easy to see that $\mathrm{TG}_{\alpha}(X;Y)$ possesses an additive property in the first variate that

$$\mathrm{TG}_{\alpha}(X_1 + X_2;Y) = \mathrm{TG}_{\alpha}(X_1;Y) + \mathrm{TG}_{\alpha}(X_2;Y).$$

In addition, it is invariant to the marginal distribution of *Y* for any given *α*, which satisfies our expectation to exclude the marginal impact of *Y*. In risk management, the tail Gini functional of a univariate random variable has been defined in [20], which intends to measure the tail variability of a univariate random variable, as part of the so-called Gini shortfall. Also, an allocation approach based on the Gini shortfall is mentioned, and some parametric models are employed. However, for the extremal measurement of systemic risk, it is necessary to consider the rate of convergence of the risk measures when $n\alpha$ is not sufficiently large enough for the inference on tail regions. In this paper, we call (2.2) in Definition 1 as TGF for short, and develop a representation and properties for an intermediate level *α* and a much smaller level *p* (possibly an extreme level). The way to model the risk level is typically useful for including an extreme high-risk event of benchmark variable *Y* in financial applications, especially when one wants to consider some tail scenarios of systemic variables. Our method requires some basic assumptions in the Extreme Value Theory. In order to investigate the asymptotic property of TGF, it is important to model the tail dependence structure of (*X*, *Y*),

$$\mathbb{P}\left(X > Q_1(1-\alpha u), Y > Q_2(1-\alpha v)\right)/\alpha, \quad \forall u,v \in [0,1], \qquad (2.3)$$

as α approaches 0. To derive the limit of TGF, we need the following assumptions.

Assumption 1.

(1.a) There exist some $\gamma_1 > 0, \beta_1 < 0$ and an eventually positive or negative function A with $\lim_{t\to\infty} A(t) = 0$ such that

$$\lim_{t\to\infty} \frac{1}{A(1/\bar{F}_1(t))}\left(\frac{\bar{F}_1(tx)}{\bar{F}_1(t)} - x^{-1/\gamma_1}\right) = x^{-1/\gamma_1}\frac{x^{\beta_1/\gamma_1}-1}{\gamma_1\beta_1}, \quad x > 0.$$

(1.b) There exist a function $R:(0,\infty)^2 \to [0,\infty)$ and $\beta > \gamma_1, \tau < 0, T > 1$ such that, as $t \to \infty$

$$\sup_{x\in(0,\infty),y\in(0,T]}\left(\frac{tC(t^{-1}x,t^{-1}y)-R(x,y)}{x^{\beta}\wedge 1}\right) = O(t^{\tau}).$$

(1.c) As $n\to\infty, \alpha = O(n^{-1+\kappa})$ for some $0 < \kappa < \min\left\{\frac{-2\tau}{-2\tau+1}, \frac{2\gamma_1\beta_1}{2\gamma_1\beta_1+\beta_1-1}\right\}$ and $\sqrt{n\alpha}A(\alpha^{-1}) = o(1)$.

(1.c') As $n\to\infty, k = O(n^{\kappa})$ for some $0 < \kappa < \min\left\{\frac{-2\tau}{-2\tau+1}, \frac{2\gamma_1\beta_1}{2\gamma_1\beta_1+\beta_1-1}\right\}$ and $\sqrt{k}A(\frac{n}{k}) = o(1)$.

Assumption (1.a) is a second-order regular variation condition in Extreme Value Theory for the marginal distribution of X. We refer to Section 2.3 of [14] for a general introduction to the second-order regular variation condition. The extreme value index γ_1 evaluates the heaviness of the (right) tail of the distribution F_1. The larger γ_1, the heavier tail of F_1. In view of Q_1, the limit is equivalent to

$$\lim_{t\to\infty}\frac{1}{A(t)}\left(\frac{Q_1(1-1/(tx))}{Q_1(1-1/t)}-x^{\gamma_1}\right)=x^{\gamma_1}\frac{x^{\beta_1}-1}{\beta_1},\quad x>0,\ (2.4)$$

where it is based on Theorem 2.3.9 of [14]. Assumption (1.b) provides a second-order condition for (2.3). Here we extend the domain of the (survival) copula C from $[0,1]^2$ to $\mathbb{R}^2$ by assuming $C(u,v)=v$ if $u>1, 0\le v\le 1$; $C(u,v)=u$ if $v>1, 0\le u\le 1$; $C(u,v)=1$ if $u,v>1$; and $C(u,v)=0$ if $u<0$ or $v<0$. Then, based on this extension, we can write that

$$C(t^{-1}x,t^{-1}y)=\mathbb{P}(1-F_1(X)<x/t,1-F_2(Y)<y/t),\quad t,x,y>0.$$

Note that R is transformed from the so-called tail dependence function defined in (6.1.33) of [14] so it satisfies the condition of being a distribution function of a measure on $[0,\infty]^2/\{(\infty,\infty)\}$. Assumption (1.c) is a condition for an intermediate level α which is not necessary for Theorem 1 but useful in Theorems 2 - 4 below. Assumption (1.c) is equivalent to Assumption (1.c') in views of $k:=k_n$ as an intermediate order sequence such that $k\to\infty, k/n\to 0$. Therefore, in the rest of the paper, we do not explicitly state which of Assumption (1.c) or (1.c') is used. It depends on which of the intermediate level or sequence is involved. To derive the limit of (2.2), we first assume X is a non-negative loss and then a general loss variable in the following two subsections.

2.1 Non-negative Loss

We first consider the case that the random loss X is non-negative and derive the limit of $\frac{\mathrm{TG}_\alpha(X;Y)}{Q_1(1-\alpha)}$.

Theorem 1. *Suppose Assumptions (1.a) and (1.b) hold with* $0<\gamma_1<1$*. Then*

$$\lim_{\alpha\downarrow 0}\frac{\mathrm{TG}_\alpha(X;Y)}{Q_1(1-\alpha)}=\frac{2\gamma_1}{2-\gamma_1}\int_0^\infty R(x^{-1/\gamma_1},1)dx=:\theta_0.\qquad (2.5)$$

The above limit in (2.5) shows the decomposition of the extreme value index of marginal distributions and the tail dependence structure. It states that as the level α goes to zero, the ratio of TGF to the quantile function Q_1 converges to an integral θ_0 with the extreme value index γ_1 derived from the marginal distribution of X and the R function from the survival copula of (X, Y). As the R function possesses the homogeneity property such that for any $a > 0$,

$$R(ax, ax) = aR(x, y), \quad x, y \geq 0,$$

it is not surprising to have the same integral $\int_0^\infty R(x^{-1/\gamma_1}, 1)dx$ in the limit (2.5) as in the limit of *Marginal Expected Shortfall* (MES) in [10]. It implies that the limits of tail risk and tail variability coincide given the integrated tail dependence structure.

It is interesting to explore the range of θ_0. As $0 \leq R(x,1) \leq x \wedge 1$ for all $x > 0$, the lower bound of θ_0 is 0 if and only if $R(x,1) = 0$ for all $x > 0$ and the upper bound is $\frac{2\gamma_1^2}{(2-\gamma_1)(1-\gamma_1)}$ if and only if $R(x,1) = x \wedge 1$ for all $x > 0$. If (X, Y) is asymptotic independent, then θ_0 reaches its lower bound. On the other hand, if (X, Y) is comonotonic, then θ_0 reaches its upper bound. However, the inverse statements are not true since θ_0 is only modeled by $R(x, y)$ on region $(0,\infty) \times \{1\}$. The range of $\mathrm{TG}_\alpha(X;Y)$ is not standard. For example, $\mathrm{TG}_\alpha(X;Y)$ possibly takes negative values because of the conditional covariance in (2.2), and the upper bound of $\mathrm{TG}_\alpha(X;Y)$ is $\mathrm{TG}_\alpha(X;X)$ whose values change with α. When considering statistical applications of TGF, it is usually more convenient to assume that $\alpha \in [0, \alpha_0]$ provided $\theta_0 > 0$ where $\alpha_0 := \sup\{\alpha | \mathrm{TG}_\tau(X;Y) > 0, \tau \in [0, \alpha]\}$. It is reasonable in practice because $\theta_0 > 0$ is a necessary condition of normal approximation in the following Theorems 4 and 6, and $\mathrm{TG}_\alpha(X;Y)$ is eventually positive given $\theta_0 > 0$ as α approaches zero. Another recommendation is to test $\theta_0 = 0$ before the statistical application. To solve the issue of upper bound, one practical way is dividing TGF by its upper bound to construct a standardized measure bounded above by 1. The discussion is postponed to Section 4.1 when we propose three measures for tail risk measurement. In addition, when it comes to an intermediate level α, we have the following stronger version for θ_0.

Theorem 2. *Suppose Assumptions (1.a) and (1.b) hold with* $0<\gamma_1<1$. *Let* α *be an intermediate level sequence such that* $\alpha\to 0, n\alpha\to\infty$ *as* $n\to\infty$, *satisfying Assumption (1.c). Then,*

$$\lim_{n\to\infty}\sqrt{n\alpha}\,\Big|\frac{\mathrm{TG}_\alpha(X;Y)}{Q_1(1-\alpha)}-\theta_0\Big|=0, \qquad (2.6)$$

where θ_0 *is defined in (2.5).*

In order to implement the above TGF for measuring tail variability, we propose one nonparametric estimator and prove its asymptotic normality in the rest of this section. Let $\{(X_i,Y_i)\}_{i=1}^n$ be independent copies of (X, Y) and k be an intermediate sequence. Denote $F_{n2}(y)=\frac{1}{n+1}\sum_i I(Y_i\le y)$. The nonparametric estimator is

$$\hat\theta_n(k):=\frac{4n}{k^2(k-1)}\sum_{i<j}(X_i-X_j)(F_{n2}(Y_i)-F_{n2}(Y_j))I(Y_i>Y_{n-k,n},Y_j>Y_{n-k,n}), \qquad (2.7)$$

where $Y_{n,n}\ge Y_{n-1,n}\ge\ldots\ge Y_{1,n}$ are the order statistics of $Y_1,\ldots,Y_n$. The following theorem shows the consistency of $\hat\theta_n(k)$ compared to TGF.

Theorem 3. *Suppose Assumption 1 holds with* $0<\gamma_1<1/2$. *Then, as* $n\to\infty$,

$$\hat\theta_n(k)/\mathrm{TG}_{k/n}(X;Y)\xrightarrow{\mathbb{P}}1.$$

To derive the limit distribution of $\hat\theta_n(k)$, define W_R as mean zero Gaussian process on $[0,\infty]^2\setminus\{\infty,\infty\}$ with covariance structure

$$\mathbb{E}[W_R(x_1,y_1)W_R(x_2,y_2)]=R(x_1\wedge x_2,y_1\wedge y_2). \; (2.8)$$

Let

$$\Theta_1 = \int_0^1 W_R(\infty, v)dv + (\gamma_1 - \frac{3}{2})W_R(\infty, 1)$$

and

$$\Theta_2 = \int_0^\infty \int_0^1 W_R(x, v) + W_R(x, 1)dvdx^{-\gamma_1}.$$

Next, we show the asymptotic normality of the proposed estimator with intermediate levels as the following theorem.

Theorem 4. *Suppose Assumption 1 holds with* $0 < \gamma_1 < 1/2, \theta_0 > 0$ *and*

$$\Phi := \frac{4}{\theta_0}\left\{\Theta_1 \int_0^\infty R(x^{-1/\gamma_1}, 1)dx + \Theta_2\right\}.$$

Then, as $n \to \infty$,

$$\sqrt{k}\left(\frac{\hat{\theta}_n(k)}{\mathrm{TG}_{k/n}(X;Y)} - 1\right) \xrightarrow{d} \Phi,$$

where

$$\begin{cases} \mathrm{Var}(\Theta_1) & = & (\gamma_1 - 1)^2 + 1/12, \\ \mathrm{Var}(\Theta_2) & = & \int_0^\infty \int_0^1 R(x, v)(v-3)dx^{-2\gamma_1}dv - \int_0^\infty R(x,1)dx^{-2\gamma_1}, \\ \mathrm{Cov}(\Theta_1, \Theta_2) & = & \int_0^\infty \int_0^1 R(x,v)(3/2 + \gamma_1 - 2v)dx^{-\gamma_1}dv + (\gamma_1 - 3/2)\int_0^\infty R(x,1)dx^{-\gamma_1}. \end{cases} \tag{2.9}$$

Theorem 4 states equivalently that as $n \to \infty$,

$$\sqrt{k}\log\left(\frac{\hat{\theta}_n(k)}{\mathrm{TG}_{k/n}(X;Y)}\right) \xrightarrow{d} \Phi. \tag{2.10}$$

The log ratio between the estimator and the true risk measure has a centered normal limit with a complicated variance. In practice, one can apply some resampling methods to estimate the variance for construction of confidence intervals. When it comes to a much smaller risk level *p*, we can modify the estimator as

$$\tilde{\theta}_p = \left(\frac{k}{np}\right)^{\hat{\gamma}_1} \hat{\theta}_n(k), \qquad (2.11)$$

where $\hat{\gamma}_1$ is a good estimator of the extreme value index γ_1 and *k* is an intermediate order satisfying the conditions of Theorem 4. For example, we can apply the Hill estimator for γ_1 defined by

$$\hat{\gamma}_1 = \frac{1}{k}\sum_{i=0}^{k-1} \log X_{n-i,n} - \log X_{n-k,n}, \qquad (2.12)$$

where $X_{n,n} \geq X_{n-1,n} \geq \ldots \geq X_{1,n}$ are the order statistics of X_i. Asymptotic normality of $\hat{\gamma}_1$ is available in the literature (see Theorem 3.2.5 of [14]) and will be used in the proof of Corollary 1. Typically, for the Hill estimator, we have that

$$\hat{\gamma}_1 = \gamma_1 + \frac{\Gamma}{\sqrt{k}} + o_{\mathbb{P}}\left(\frac{1}{\sqrt{k}}\right),$$

where

$$\Gamma = \gamma_1\left(\int_0^1 W_R(x,\infty)\frac{dx}{x} - W_R(1,\infty)\right). \qquad (2.13)$$

Without loss of generality, we assume a consistent intermediate sequence *k* in the Hill estimator and Assumption (1.c).

Corollary 1. *Under the conditions of Theorem 4, let* $p := p_n \leq \frac{k}{n}$ *be a level sequence such that* $p \to 0$ *and* $d_n := \frac{k}{np} \to d \in [1,\infty]$ *and* $\frac{\log d_n}{k} \to 0$ *as* $n \to \infty$*. Then, as* $n \to \infty$*,*

$$\sqrt{k}\left(1\wedge\frac{1}{\log d_n}\right)\left(\frac{\tilde{\theta}_p}{\mathrm{TG}_p(X;Y)}-1\right)\xrightarrow{d}\left(1\wedge\frac{1}{\log d}\right)\Phi+\left(1\wedge\log d\right)\Gamma,$$

where Φ is defined in Theorem 4 and Γ is defined in (2.13), and in addition to (2.9),

$$\begin{cases}\mathrm{Var}(\Gamma) & = & \gamma_1^2,\\ \mathrm{Cov}(\Theta_1,\Gamma) & = & \int_0^1\int_0^1 x^{-1}R(x,v)dvdx+(\gamma_1-3/2)\int_0^1 x^{-1}R(x,1)dx\\ & & -\int_0^1 R(1,v)dv-(\gamma_1-3/2)R(1,1),\\ \mathrm{Cov}(\Theta_2,\Gamma) & = & -\frac{1}{\gamma_1}\int_0^1\int_0^1\int_0^1\left(R(x_1x_2,v)+R(x_1x_2,1)\right)dvdx_1^{-\gamma_1}dx_2^{-\gamma_1}\\ & & +\int_0^1\int_0^1 x^{-1}\left(R(x,v)+R(x,1)\right)dvdx+\int_0^1\int_0^1(\gamma_1^{-1}-x^{-1})R(x,v)dvdx^{-\gamma_1}\\ & & +\int_0^1(\gamma_1^{-1}-2x^{-1})R(x,1)dx^{-\gamma_1}+\frac{\gamma_1}{1+\gamma_1}\int_0^1\left(R(1,v)+R(1,1)\right)dv.\end{cases}\tag{2.14}$$

In the above limit, when $d = 1$, the limit becomes Φ, the same result in Theorem 4, since in this special case p is also an intermediate level. In addition, one can also find the equivalent convergence for a log ratio similar to (2.10).

2.2 A General Loss

In this subsection, we consider the case when X is real. Denote that $X^+=\max(X,0)$ and $X^-=\min(X,0)$. So we have $X=X^++X^-$. Put

$$\mathrm{TG}_\alpha^+(X;Y):=\mathrm{TG}_\alpha(X^+;Y)=\frac{4}{\alpha}\mathrm{Cov}(X^+,F_2(Y)\,|\,F_2(Y)>1-\alpha),$$

and

$$\mathrm{TG}_\alpha^-(X;Y):=\mathrm{TG}_\alpha(X^-;Y)=\frac{4}{\alpha}\mathrm{Cov}(X^-,F_2(Y)\,|\,F_2(Y)>1-\alpha).$$

To derive similar asymptotic results as in the non-negative case, we need two more conditions. These conditions are similar to conditions in Section 2.2 in [10], which are applied to deal with the negative part of the general loss.

Assumption 2.

(2.a) $\mathbb{E}|X^-|^{1/\gamma_1} < \infty$.

(2.b) As $n \to \infty, \alpha = O(n^{-1+\kappa})$ with some $0 < \kappa < -2\tau(1-\gamma_1)$.

(2.b') As $n \to \infty, k = O(n^{\kappa})$ with some $0 < \kappa < -2\tau(1-\gamma_1)$.

Note that Assumption (2.b) and (2.b') are only different in using α or k. Besides, one may easily observe that there is no conflict between the ranges of κ in Assumption (1.c) and (2.b).

Theorem 5*. Suppose Assumptions (1.a), (1.b) and (2.a) hold with* $0 < \gamma_1 < 1$*. Then, as* $n \to \infty$*,*

$$\lim_{\alpha \downarrow 0} \frac{\mathrm{TG}_\alpha(X;Y)}{\mathrm{TG}_\alpha^+(X;Y)} = 1. \quad (2.15)$$

Moreover, if the intermediate level α satisfies Assumption (1.c) and (2.b), then as $n \to \infty$,

$$\lim_{n \to \infty} \sqrt{n\alpha} \left| \frac{\mathrm{TG}_\alpha(X;Y)}{\mathrm{TG}_\alpha^+(X;Y)} - 1 \right| = 0. \ (2.16)$$

Note that θ_0 is still defined as $\lim_{\alpha \downarrow 0} \frac{\mathrm{TG}_\alpha(X;Y)}{Q_1(1-\alpha)}$. Based on the additive property of TGF, Theorem 5 states the fact that

$$\theta_0 = \lim_{\alpha \downarrow 0} \frac{\mathrm{TG}_\alpha(X;Y)}{Q_1(1-\alpha)} = \lim_{\alpha \downarrow 0} \frac{\mathrm{TG}_\alpha^+(X;Y)}{Q_1(1-\alpha)}.$$

It means that the tail variability of a general loss X is dominated by its positive component X^+ because $\mathrm{TG}_\alpha(X;Y)$ is very close to $\mathrm{TG}_\alpha^+(X;Y)$ when α is sufficiently close to zero. It is intuitively reasonable since we are measuring the right tail variability of *X*, so that the contribution to the tail variability from the negative component X^- will eventually disappear as the level α goes to zero. Therefore, the range of θ_0 is still the interval $[0, 2\gamma_1^2 / ((2-\gamma_1)(1-\gamma_1))]$. For the range of $\mathrm{TG}_\alpha(X;Y)$, it is still not standard, but we can similarly conclude that for a general loss variable *X*, if θ_0 is positive, the tail Gini functional is eventually positive considering $\alpha \in [0, \alpha_0]$ where $\alpha_0 = \sup\{\alpha | \mathrm{TG}_\tau(X;Y) > 0, \tau \in [0, \alpha]\}$. The conclusion is very useful in practice when one wants to estimate the risk measures in Definition 2.

For a real random loss *X*, we need to modify the estimator in (2.7) as

$$\hat{\theta}_n(k) := \frac{4n}{k^2(k-1)} \sum_{i<j} (X_i - X_j)(F_{n2}(Y_i) - F_{n2}(Y_j)) I(X_i > 0, X_j > 0, Y_i > Y_{n-k,n}, Y_j > Y_{n-k,n}). \tag{2.17}$$

Notice that (2.7) and (2.17) are indeed the same estimator in the case of non-negative random loss *X*. Therefore, we do not use another symbol to represent the estimator for the sake of simplicity.

Theorem 6. *Suppose Assumptions 1 and 2 hold with* $0 < \gamma_1 < 1/2$*. Then, as* $n \to \infty$*,*

$$\hat{\theta}_n(k) / \mathrm{TG}_{k/n}(X;Y) \xrightarrow{\mathbb{P}} 1.$$

Furthermore, if $\theta_0 > 0$*, then as* $n \to \infty$*,*

$$\sqrt{k}\left(\frac{\hat{\theta}_n(k)}{\mathrm{TG}_{k/n}(X;Y)} - 1\right) \xrightarrow{d} \Phi,$$

where Φ *is defined in Theorem 4.*

Similar to Corollary 1, we can apply extrapolation techniques to define $\tilde{\theta}_p$ based on the same representation (2.11) with using $\hat{\theta}_n(k)$ in (2.17) instead. The asymptotic normality of $\tilde{\theta}_p$ is guaranteed.

Corollary 2. *Under the conditions of Theorem 6, let* $p := p_n \le \frac{k}{n}$ *be a level sequence such that* $p \to 0$ *and* $d_n := \frac{k}{np} \to d \in [1, \infty]$ *and* $\frac{\log d_n}{k} \to 0$ *as* $n \to \infty$. *Then,*

$$\sqrt{k}\left(1 \wedge \frac{1}{\log d_n}\right)\left(\frac{\tilde{\theta}_p}{\mathrm{TG}_p(X;Y)} - 1\right) \xrightarrow{d} \left(1 \wedge \frac{1}{\log d}\right)\Phi + (1 \wedge \log d)\Gamma,$$

where Φ *is defined in Theorem 4 and Γ is defined in (2.13).*

3 Simulation Study

In this section, we examine the finite performance of the proposed nonparametric estimator of Tail Gini Functionals (TGF) with both intermediate level *α*, which makes $n\alpha$ large enough for normal approximation, and a much smaller risk level *p*, which is treated as extreme level in the study. Here *n* is the sample size of the simulated data. We generate the data from the following three cases in [11] of the following joint distributions, which satisfy the conditions in the theorems.

- Case I: A transformed Cauchy distribution on $(0, \infty)^2$ defined as

  $$(X, Y) = (|Z_1|^{2/5}, |Z_2|),$$

  where (Z_1, Z_2) is a standard Cauchy distribution on $\mathbb{R}^2$ with density $f(x, y) = \frac{1}{2\pi}(1 + x^2 + y^2)^{-3/2}$. It follows that $\gamma_1 = 2/5$ and $R(x, y) = x + y - \sqrt{x^2 + y^2}, x, y \ge 0$. This distribution satisfies that $\tau = -1, \beta = 2$ and $\rho_1 = -2$.

• Case II: Student-t_3 distribution on $(0,\infty)^2$ with density $f(x,y)=\frac{2}{\pi}\left(1+\frac{x^2+y^2}{3}\right)^{-5/2}, x,y>0$. It follows that $\gamma_1=1/3, R(x,y)=x+y-\frac{x^{4/3}+x^{2/3}y^{2/3}/2+y^{4/3}}{\sqrt{x^{2/3}+y^{2/3}}}, \tau=-1/3, \beta=4/3$ and $\rho_1=-2/3$.

• Case III: A transformed Cauchy distribution on the whole $\mathbb{R}^2$ defined as

$$(X,Y)=\left(Z_1^{2/5}I(Z_1\geq 0)+Z_1^{1/5}I(Z_1<0), Z_2I(Z_1\geq 0)+Z_2^{1/3}I(Z_1<0)\right).$$

This distribution satisfies that $\gamma_1=2/5, R(x,y)=x/2+y-\sqrt{x^2/4+y}, \tau=-1, \beta=2$ and $\rho_1=-2$.

In addition, for each case we consider

- Two intermediate risk levels $\alpha=0.1, 0.05$ and two extreme risk levels $p=0.01, 0.001$.
- Sample size n = 800 or 2000 or 5000.

For comparison between the estimators and true values, we evaluate these true values $\mathrm{TG}_\alpha(X;Y)$ and $\mathrm{TG}_p(X;Y)$ by using the true density functions and drawing 10000 replications with sample size 10000 from the above three cases. The true values are then approximated by the corresponding median of overall 10000 replications, which are reported in Table 2. For the extreme levels, two different $k=n\alpha$ are used to get the true value. The results do not differ much by exploiting different intermediate orders in the extrapolation. For all of the three cases, we can see that when the intermediate level $\alpha=0.1$ changes to $\alpha=0.05$, the TGF increases about 5%-30%. The TGF of extreme level p = 0.01 is nearly twice that of the intermediate level $\alpha=0.05$. The TGF of the extreme level p = 0.001 is nearly triple that of the extreme level p = 0.01. It indicates that when the benchmark Y is undergoing some tail scenarios, the impact of higher risk level results in more variability on tail regions of X.

Next, we draw 2000 replications from 3 different cases with sample size n = 800, 2000 and 5000 separately. For each replication, we compute the proposed non-parametric estimator $\hat{\theta}_n(k)$ and the corresponding extreme estimators $\tilde{\theta}_{0.01}$ and $\tilde{\theta}_{0.001}$ with different n and k. In order to compare our estimators with the true values, the log ratios $\log(\frac{\hat{\theta}_n(k)}{TG_{k/n}(X;Y)})$ and $\log(\frac{\tilde{\theta}_p}{TG_p(X;Y)})$ are calculated and the corresponding averages and standard errors are reported in Table 3. As mentioned in (2.10), the log ratio is supposed to approximate a centered normal limit, which works for extreme levels as well. We can make the following conclusions from the ratios. First, the means of the ratios are close to one for both two intermediate levels and two extreme levels, which indicates the accuracy of the two proposed estimators. The ratios in the case of intermediate levels are closer to one than those in the case of extreme levels. It is mainly due to the extrapolation techniques used for the extreme level. Second, in each case, if we compare the standard errors for different sample sizes n, it is obvious that those results with a larger sample size have smaller standard errors. On the other hand, the standard errors for level $\alpha = 0.05$ are greater than that for $\alpha = 0.1$ given the same sample size n. It is because if the sample size increases, more sample on tail regions is used in the estimation processes, which will reduce the sample variance. However, increasing the sample size on tail regions will probably increase bias, which is a common bias-variance trade-off issue in extreme value analysis. In practice, one can use the Hill estimator to choose a suitable intermediate order $k := \lfloor n\alpha \rfloor + 1$, as illustrated in the next section. In the simulation study, we select the same intermediate order for comparison in all cases. Third, for both the extreme levels p = 0.01 and p = 0.001, the estimators are slightly different when we using different k for the same sample size n. It indicates the selection of intermediate order k is very important in the empirical analysis.

In addition, we present boxplots of the log ratios between the estimators and the true values for each case. From the boxplots in Figure 1, we can see that most of the estimators are distributed symmetrically around 1 with a little heavy tail on the upper regions due to the finite sample size. Moreover, when the sample size n increases, the body of the box becomes narrower, which shows the convergence of risk measures in probability as shown in Theorems 3 and 6. In order to show the asymptotic property of the proposed estimators, we also compare the sample quantiles of log ratios for all levels with the normal quantiles by using QQ plots in Figure 2 for all cases. Most of the scatters line up on the red line, which indicates no big difference from a normal distribution. Moreover, when the sample size increase from n = 800 to 5000, the plots seem better with fewer outliers. Consequently, we can conclude that Theorem 6 provides an adequate approximation for finite sample sizes and that the simulation study shows the expected finite performance and asymptotic normality of the proposed estimators.

4 Applications and Real Data Analysis

In this section, we propose three coefficients/measures (or simply measures) in Definition 2 by using the Tail Gini Functionals and Marginal Expected Shortfall, and then apply them to a real data analysis. These measures are intended for the tail analysis of systemic risks in practice, as stated in Section 1. In order to illustrate them, we rephrase MES in our framework, which is proposed in [1] and its extremal statistical properties are further explored in [10]. Recall that we have a random loss pair (X, Y) with a continuous joint distribution F and marginal distribution $F_i, i=1,2$, and the MES in our framework is

$$\mathrm{ES}_\alpha(X;Y) = \mathbb{E}(X \mid F_2(Y) > 1-\alpha), \quad (4.1)$$

and a nonparametric estimator (for a general loss *X*) under an intermediate sequence *k* is proposed as

$$\hat{\eta}_n(k) = \frac{1}{k}\sum_{i=1}^{k} X_i I(X_i > 0, Y_i > Y_{n-k,n}). \quad (4.2)$$

If one has a much smaller risk level *p* than the intermediate level *k* / *n*, it is supposed to apply the extrapolation technique and use the estimator

$$\tilde{\eta}_p = \left(\frac{k}{np}\right)^{\hat{\gamma}_1} \frac{1}{k}\sum_{i=1}^{k} X_i I(X_i > 0, Y_i > Y_{n-k,n}), \qquad (4.3)$$

where $\hat{\gamma}_1$ is defined in (2.12). Based on the fact that given tail scenarios of a systemic variable, MES measures the expected individual loss while TGF measures its tail variability, we can extend a few traditional coefficients to the tail regions for the tail risk analysis.

One interesting application of MES and TGF is the decomposition of tail risk and tail variation based on the additive properties of the expectation- and covariance-based formulas. Consider a portfolio of *l* assets with losses X_j and weights ω_j for $j = 1, 2, \ldots, l$. Denote the total portfolio loss as the benchmark $Y = \sum_{j=1}^{l} \omega_j X_j$, our concern is to decompose the tail risk and tail variation of *Y* with respect to each asset loss. It is not difficult to see that for the decomposition of tail risk, we have

$$\mathrm{ES}_\alpha(Y;Y) = \sum_{j=1}^{l} \omega_j \mathrm{ES}_\alpha(X_j;Y). \qquad (4.4)$$

For the decomposition of the tail variation of *Y*, based on the covariance representation, we have that for any $\alpha \in (0,1)$,

$$\mathrm{Cov}(Y, F_2(Y) \mid F_2(Y) > 1-\alpha) = \sum_{j=1}^{l} \omega_j \mathrm{Cov}(X_j, F_2(Y) \mid F_2(Y) > 1-\alpha),$$

which also implies that

$$\mathrm{TG}_\alpha(Y;Y)=\sum_{j=1}^{l}\omega_j\mathrm{TG}_\alpha(X_j;Y). \qquad (4.5)$$

Therefore, the decompositions (4.4) and (4.5) have the same weights ω_j as in the portfolio benchmark *Y*, and it is natural to consider the pair $(\mathrm{ES}_\alpha(X_j;Y),\mathrm{TG}_\alpha(X_j;Y))$ as the contribution of each asset to the total tail risk and tail variation of the portfolio.

4.1 Measures for Tail Risk Analysis

It is well known that the *coefficient of variation* of a random variable *Z* with mean μ and standard deviation σ is defined as $\frac{\sigma}{\mu}$. It is a standardized measure for quantifying the dispersion of the distribution of *Z*. A small coefficient of variation (in the sense of absolute value) means a large mean μ or a small standard deviation σ, so observations are more likely to concentrate around the population mean compared with the absolute quantity of the mean. Thus, the dispersion of the distribution is summarized as small, and it gives an intuitive guarantee in prediction analysis that a few predicted observations will concentrate around the mean with distances relatively small compared to the distance between the mean and zero. Motivated by the concept of dispersion, we propose the first measure as the coefficient of Tail Gini Co-variation (TGV) in Definition 2(a). The TGV measures the dispersion of the tail distribution of *X* conditional on the tail scenarios of *Y* by summarizing the extent of concentration around the expected loss (MES) compared with the co-variation caused by the systemic variable (TGF).

The second measure is the Tail Gini Correlation (TGR) defined in Definition 2(b). In [34], Gini Correlation (or Gini's gamma) is defined as a measure for pair correlation, similar to the well-known association measures, Pearson's correlation, Kendall's tau and Spearman's rho. Properties of Gini Correlation are explored in [35, 38]. When it comes to tail risk analysis, it is natural to consider the association measure under the tail scenarios of Y. TGR is then defined as an asymmetric mixture of conditional Pearson's and Spearman's correlation coefficients. It behaves like Pearson's coefficient in X, like Spearman's coefficient in Y. By considering the risk variability on tail regions, TGR serves as a measure for the relative tail dependence structure of (X, Y) compared to the perfect positive dependence structure on tail regions. There is another interesting decomposition of tail variation similar to (4.5): recall that a portfolio consists of l assets with losses X_j and weight ω_j for $j=1,2,\ldots,l$. Let F_{1j} denote the marginal distribution of X_j and Q_{1j} denote its quantile function. Consider a new benchmark, the accumulated portfolio excess loss $Y^*=\sum_{j=1}^{l}\omega_j(X_j-Q_{1j}(1-\alpha))_+$. It is trivial to verify that

$$(X_j-Q_{1j}(1-\alpha))_+=X_j\vee Q_{1j}(1-\alpha)-Q_{1j}(1-\alpha),$$

and

$$\mathrm{TG}_\alpha(X_j\vee Q_{1j}(1-\alpha);X_j)=\mathrm{TG}_\alpha(X_j;X_j).$$

As Q_{1j}'s are all constants which have no influence on the tail variation of Y^*, the decomposition of the tail variation of Y^* is

$$\mathrm{TG}_\alpha(Y^*;Y^*)=\sum_{j=1}^{l}\omega_j\cdot\mathrm{TGR}_\alpha(X_j;Y^*)\cdot\mathrm{TG}_\alpha(X_j;X_j). \quad (4.6)$$

One may note that (4.6) not only describes the relationship between TGR and TGF in the tail variation of the accumulated portfolio excess loss, but also provides the relationship between the individual tail variation $\mathrm{TG}_\alpha(X_j;X_j)$ and the total tail variation of Y^*, which is of great interest in tail risk measurement for portfolio analysis.

The last measure is the Marginal Gini Shortfall (MGS) in Definition 2(c). For a univariate variable X, Gini Shortfall (GS) is recently proposed in Section 6 of [20] as

$$\mathrm{GS}_\alpha(X) = \mathbb{E}(X \mid X > Q_1(1-\alpha)) + \lambda \mathrm{Cov}(X, F_1(X) \mid X > Q_1(1-\alpha)).$$

It is a mixture for measuring both the tail risk and tail risk variability of a univariate loss. The loading parameter λ in GS is used for the adjustment of weights between the tail risk and variability. GS originates from the premium functionals under the Gini principle in [15]. The Gini principle indicates a measure $\mathrm{GS}_1(X) = \mathbb{E}(X) + \lambda \mathrm{Cov}(X, F_1(X))$ by considering the overall distribution of risk. So it is a special case of GS, and thus GS is generalized by replacing the expectation $\mathbb{E}(X)$ in $\mathrm{GS}_1(X)$ with the conditional tail expectation and the risk variability $\mathrm{Cov}(X, F_1(X))$ with the tail risk variability (with certain risk level α) for a univariate loss variable. In this paper, we further extend GS to MGS in (4.9), which includes the impact from the tail scenarios of the systemic variable.

Definition 2. Suppose $\mathrm{TG}_\alpha(X;Y)$ is the Tail Gini Functional in Definition 1 and $\mathrm{ES}_\alpha(X;Y)$ is the Marginal Expected Shortfall in (4.1) where α is a risk level in (0, 1). Define

(a) Coefficient of Tail Gini Co-variation

$$\mathrm{TGV}_\alpha(X;Y) = \frac{\mathrm{TG}_\alpha(X;Y)}{\mathrm{ES}_\alpha(X;Y)}. \quad (4.7)$$

(b) Tail Gini Correlation

$$\mathrm{TGR}_\alpha(X;Y) = \frac{\mathrm{TG}_\alpha(X \vee Q_1(1-\alpha);Y)}{\mathrm{TG}_\alpha(X;X)}, \qquad (4.8)$$

where $\frac{0}{0}$ is defined as 0.

(c) Marginal Gini Shortfall

$$\mathrm{GS}_{\alpha}(X;Y,\lambda)=\mathrm{ES}_{\alpha}(X;Y)+\lambda\mathrm{TG}_{\alpha}(X;Y), \quad (4.9)$$

where λ is a non-negative loading parameter.

Note that in (4.8) we take $X \vee Q_1(1-\alpha)$ in the numerator $\mathrm{TG}_{\alpha}(X \vee Q_1(1-\alpha);Y)$ to ensure the upper bound of TGR is 1 for any given α. In practice, one can use both (2.17) and (4.2) to construct plug-in estimators for the three measures. In the next subsection, we follow this idea to estimate the three measures for real data analysis.

4.2 Real Data Analysis

In the remaining section, we apply our estimators to a real data set consisting of the daily stock price of S&P 500 index and 22 companies from January 1st, 2001 to December 30th, 2016. The weekly return is transformed from 4026 observations of daily stock prices for all the stocks. Since we care more about the loss in the market, we then take the negative weekly returns and consider the (right) tail risk analysis of them. Throughout the following of this section, we use " loss" to represent the percentage of the negative weekly returns of the index and the 22 stocks. The unit of the loss is 1%. Our objective is to study the effect of the whole market on the individual stock under the extremely high-risk event of the systemic variable by estimating MES, TGF, and the other three measures proposed in Definition 2.

Table 4 shows the full names and tickers of the 22 companies plus S&P 500 index (SNP500), as well as the summary statistics of all the losses. The companies are sorted and displayed by their full names alphabetically. Figure 3 shows the boxplots of all losses whose features are also characterized in Table 4. The green boxes are plotted by the data in the whole period, while the red boxes by the data only in the year 2008. It is easy to spot that the losses in the crisis period are higher than usual, so it is useful to consider the tail scenarios of a systemic variable when analyzing systemic risks. Furthermore, the medians of green boxes are all close to 0 with values smaller than 0, which means slightly small positive returns for all companies. However, those of red boxes show positive losses in the year 2008. The boxplot of SNP500 is comparatively shorter than the others, which indicates the smaller variation of the loss data. It coincides with the fact that SNP500 is a portfolio consisting of a basket of significant stocks and from the view of the modern portfolio theory, the portfolio variance is reduced by choosing asset classes with a low or negative correlation. Thus, in the data analysis, the loss of SNP500 is treated as a systemic variable Y in the modeling method, whose tail scenarios have some great impacts from systemic risk to the individual losses X.

Before studying the effect of the whole market on the performance of the individual stock under the extreme conditions, we estimate first the extreme value index and the measures at an intermediate level for each loss. In Figure 3, all losses have more extremely high observations than extremely low observations, which shows the importance of studying the right tail risk of them. This fact enables us to apply the Hill estimator in (2.12) with a reasonable intermediate order k. Instead of using Hill plots as in the simulation study to select a uniform k for all cases, we apply the method in [13] to find the optimal k for each loss and consistently use this k as the intermediate order in other measures. The optimal k and the corresponding Hill estimators are given in the fifth and sixth columns of Table 4 separately. The corresponding TGF and MES for each company at the selected intermediate level are shown in the last two columns. From Table 4, we can observe that for all the losses, the estimated extreme value index $\hat{\gamma}_1$ is between 0.2 and 0.6. Mostly, when $\hat{\gamma}_1$ increases, the TGF and MES increase as well. However, there is no obvious strong relationship between the value of $\hat{\gamma}_1$ and the two risk measures. It is because the extreme value index is only estimated by the loss data of X while the two risk measures MES and TGF are estimated from the loss data of (X, Y), where extra information from the systemic variable is included in the procedure of tail risk measurement. In addition, Figure 4 also shows the relationship between the two risk measures and the Hill estimators at the two extremes levels $p = 0.01$ and $p = 0.001$. It shows a stronger positive tail dependence between the two risk measures and the Hill estimator when the tail scenario becomes more extreme.

To quantify the effect of the whole market on the performance of the individual stock under the extreme scenario, we calculate the above measures, including MES, TGF, TGV TGR, MGS with two levels $p = 0.01, p = 0.001$. Although some of the companies may have an estimated extreme value index slightly greater than 0.5, we still estimate those measures at extreme levels in Table 5. Recall that the losses of all the company stocks are treated as the loss variables X individually while the loss of SNP500 is treated as the benchmark variable Y consistently in the 22 cases. Since the sample size is 833, it is natural to treat these two levels as extreme levels in our method. The simulation result shows the TGF can be estimated accurately with a sample size $n = 800$ and larger. The risk measures are estimated based on the optimal k and the estimator $\hat{\gamma}_1$. They measure the effect of the market on the individuals from different aspects at extreme levels, and they can be used to set a threshold in the decision making process and premium calculation in asset pricing. Note that the TGR is calculated by using results of intermediate level k / n due to extrapolation techniques in (2.11), which cancels the first factor related to the level p. This calls for the further study of an alternative estimation of the TGR for extreme levels.

Figure 5 is a collection of tail loss-variability ratio plots. The upper left graph shows the mean and standard error for all the companies which describe the overall loss-variability pattern. One can observe that most of the companies concentrate into a small group while a few companies, AIG, C, LNC, and HIG, stay far away from the group, either suffering a high loss like LNC or large variability like AIG. The other three graphs in Figure 5 show the square root of TGF and MES of all the companies at the intermediate level and two extreme levels p = 0.01 and p = 0.001 separately. There is an positive relationship between the square root of TGF and MES, especially when the level becomes smaller. Besides, there is an interesting fact across the three plots that most of the companies stay in a relatively stable position compared to the group while HIG is moving away from them when the level decreases, showing both higher tail loss and variability. It indicates HIG may suffer more from the systemic risk than the others given the system is under tail scenarios. The systemic risk has influences on other companies as well. Company C and LNC have relatively greater MES. AIG and MET have greater increments than C and LNC when the level decreases.

Table 5 summarizes all the estimated measures in Definition 2 for these companies. One may see that most of the companies have TGR greater than 0.5, which indicates a relatively high dependence with the systemic variable on tail regions so that the impact from it is not ignorable. One can also observe that the TGV generally decreases when the level decreases, and except ALL and CB, the others have lower TGV than SNP500. Therefore, the estimated tail dispersions are very similar for most stocks. Finally, MGS also indicates that HIG encounters the greatest influence from the market, which is not surprising as it has both greater MES and TGF. MGS displayed in Table 5 is calculated using λ = 1; it equally weighs the effect of expected loss and variability. However, dependent on different objectives, λ can be adjusted to make a trade-off between the expected loss and variability. To visualize the results in Table 5, Figure 6 plots the histograms of the nine risk measures and shows the ranges and counts of each measure at the two levels.

5 Conclusion

In this paper, we consider the tail-based risk measurement for tail variability given the impact from tail scenarios of some benchmark/systemic variable. The proposed method is useful for the tail analysis of systemic risks. We define the Tail Gini Functional as a measure for tail variability extended from Gini's methodology. By assuming both intermediate levels and extreme levels, we apply conditions/results in Extreme Value Theory to our models and explore the limit representation of TGF. The limit shows a decomposition between the tail dependence structure and marginal risk quantity, i.e., the limit of tail survival copula and extreme value index of the marginal distribution. Besides, we propose a nonparametric estimator for the TGF and its asymptotic normality is proved for statistical inference. Furthermore, we construct three measures for tail risk analysis by using both TGF and MES. These measures provide some insightful metrics of risks on tail regions in a real data analysis.

Acknowledgments

We thank the Associate Editor and two anonymous referees for various constructive comments and suggestions. Yanxi Hou's research was partially supported by National Natural Science Foundation of China Grant 71803026 and 71991471, and by Science and Technology Commission of Shanghai Municipality Project 19511120700.

References

[1] Acharya, V.V., Pedersen, L. H., Philippon, T. and Richardson, M., 2017. Measuring Systemic Risk. *The Review of Financial Studies, 30(1), pp.2-47.*

[2] Adrian, T. and Brunnermeier, M.K., 2016. CoVaR. *American Economic Review, 106(7), pp.1705-41.*

[3] Agarwal, V., Ruenzi, S. and Weigert, F., 2017. Tail risk in hedge funds: A unique view from portfolio holdings. *Journal of Financial Economics, 125(3), pp.610-636.*

[4] Almeida, C., Ardison, K., Garcia, R. and Vicente, J., 2017. Nonparametric tail risk, stock returns, and the macroeconomy. *Journal of Financial Econometrics, 15(3), pp.333-376.*

[5] Allen, L., Bali, T.G. and Tang, Y., 2012. Does systemic risk in the financial sector predict future economic downturns?. *The Review of Financial Studies, 25(10), pp.3000-3036.*

[6] Artzner, P., Delbaen, F., Eber, J.M. and Heath, D., 1999. Coherent measures of risk. *Mathematical Finance, 9(3), pp.203-228.*

[7] Asimit, A.V. and Li, J., 2016. Extremes for coherent risk measures. *Insurance: Mathematics and Economics, 71, pp.332-341.*

[8] Billio, M., Getmansky, M., Lo., A.W. and Pelizzon, L., 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics 104(3), pp.535-559.*

[9] Brownlees, C. and Engle, R.F., 2016. SRISK: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies, 30(1), pp.48-79.*

[10] Cai, J.J., Einmahl, J.H., Haan, L. and Zhou, C., 2015. Estimation of the marginal expected shortfall: the mean when a related variable is extreme. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 77(2), pp.417-442.*

[11] Cai, J.J., Chavez-Demoulin, V. and Guillou, A., 2017. Modified marginal expected shortfall under asymptotic dependence. *Biometrika, 104(1), pp.243-249.*

[12] Ceriani, L. and Verme, P., 2012. The origins of the Gini index: extracts from Variabilitàe Mutabilità (1912) by Corrado Gini. *The Journal of Economic Inequality, 10(3), pp.421-443.*

[13] Cheng, S. and Peng, L., 2001. Confidence intervals for the tail index. *Bernoulli, 7(5), pp.751-760.*

[14] de Haan, L. and Ferreira, A., 2006. Extreme Value Theory: an Introduction. *New York: Springer.*

[15] Denneberg, D., 1990. Premium calculation: why standard deviation should be replaced by absolute deviation. *ASTIN Bulletin: The Journal of the IAA, 20(2), pp.181-190.*

[16] Fei, J.C., Ranis, G. and Kuo, S.W., 1978. Growth and the family distribution of income by factor components. *The Quarterly Journal of Economics, 92(1), pp.17-53.*

[17] Frees, E.W., Meyers, G. and Cummings, A.D., 2011. Summarizing insurance scores using a Gini index. *Journal of the American Statistical Association, 106(495), pp.1085-1098.*

[18] Frees, E.W., Meyers, G. and Cummings, A.D., 2014. Insurance ratemaking and a Gini index. *Journal of Risk and Insurance, 81(2), pp.335-366.*

[19] Furman, E. and Landsman, Z., 2006. Tail variance premium with applications for elliptical portfolio of risks. *ASTIN Bulletin: The Journal of the IAA, 36(2), pp.433-462.*

[20] Furman, E., Wang, R. and Zitikis, R., 2017. Gini-type measures of risk and variability: Gini shortfall, capital allocations, and heavy-tailed risks. *Journal of Banking* & *Finance, 83, pp.70-84.*

[21] Gini, C., 1909. Il diverso accrescimento delle classi sociali e la concentrazione della ricchezza. Giornale degli economisti, 38, pp.27-83.

[22] Gini, C., 1912. Variabilità e mutabilità: Contributo allo studio delle distribuzioni e delle relazioni statistiche, Vol. III(part II). *Bologna: Cuppini.*

[23] Gini, C., 1914. Sulla misura della concentrazione e della variabilità dei caratteri. *Atti del Reale Istituto veneto di scienze, lettere ed arti, 73, pp.1203-1248.*

[24] Giorgi, G.M., 1990. Bibliographic portrait of the Gini concentration ratio. *Metron, 48(1-4), pp.183-221.*

[25] Giorgi, G.M., 1993. A fresh look at the topical interest. *Metron, 51(1-2), pp.83-98.*

[26] Giorgi, G.M. and Gigliarano, C., 2017. The Gini concentration index: a review of the inference literature. *Journal of Economic Surveys, 31(4), pp.1130-1148.*

[27] He, Y., Hou, Y. and Peng, L, 2017. Statistical Inference for a Relative Risk Measure. *Journal of Business & Economic Statistics, preprint.*

[28] Lerman, R.I. and Yitzhaki, S., 1984. A note on the calculation and interpretation of the Gini index. *Economics Letters, 15(3-4), pp.363-368.*

[29] Markowitz, H., 1952. Portfolio selection. *The Journal of Finance, 7(1), pp.77-91.*

[30] McNeil, A.J., Frey, R. and Embrechts, P., 2015. Quantitative Risk Management: Concepts, Techniques and Tools-revised edition. *Princeton university press.*

[31] Nelsen, R.B., 2007. An introduction to copulas. *Springer Science* & *Business Media.*

[32] Kelly, B. and Jiang, H., 2014. Tail risk and asset prices. *The Review of Financial Studies, 27(10), pp.2841-2871.*

[33] Rockafellar, R.T., Uryasev, S. and Zabarankin, M., 2006. Generalized deviations in risk analysis. *Finance and Stochastics, 10(1), pp.51-74.*

[34] Schezhtman, E. and Yitzhaki, S., 1987. A Measure Of Association Based On Gin's Mean Difference. *Communications in statistics-Theory and Methods, 16(1), pp.207-231.*

[35] Schechtman, E. and Yitzhaki, S., 1999. On the proper bounds of the Gini correlation. *Economics letters, 63(2), pp.133-138.*

[36] Stuart, A., 1954. The correlation between variate-values and ranks in samples from a continuous distribution. *British Journal of Statistical Psychology, 7(1), pp.37-44.*

[37] Van Oordt, M.R. and Zhou, C., 2016. Systematic tail risk. *Journal of Financial and Quantitative Analysis, 51(2), pp.685-705.*

[38] Yitzhaki, S. and Schechtman, E., 2012. The Gini methodology: A primer on a statistical methodology (Vol. 272). *Springer Science* & *Business Media.*

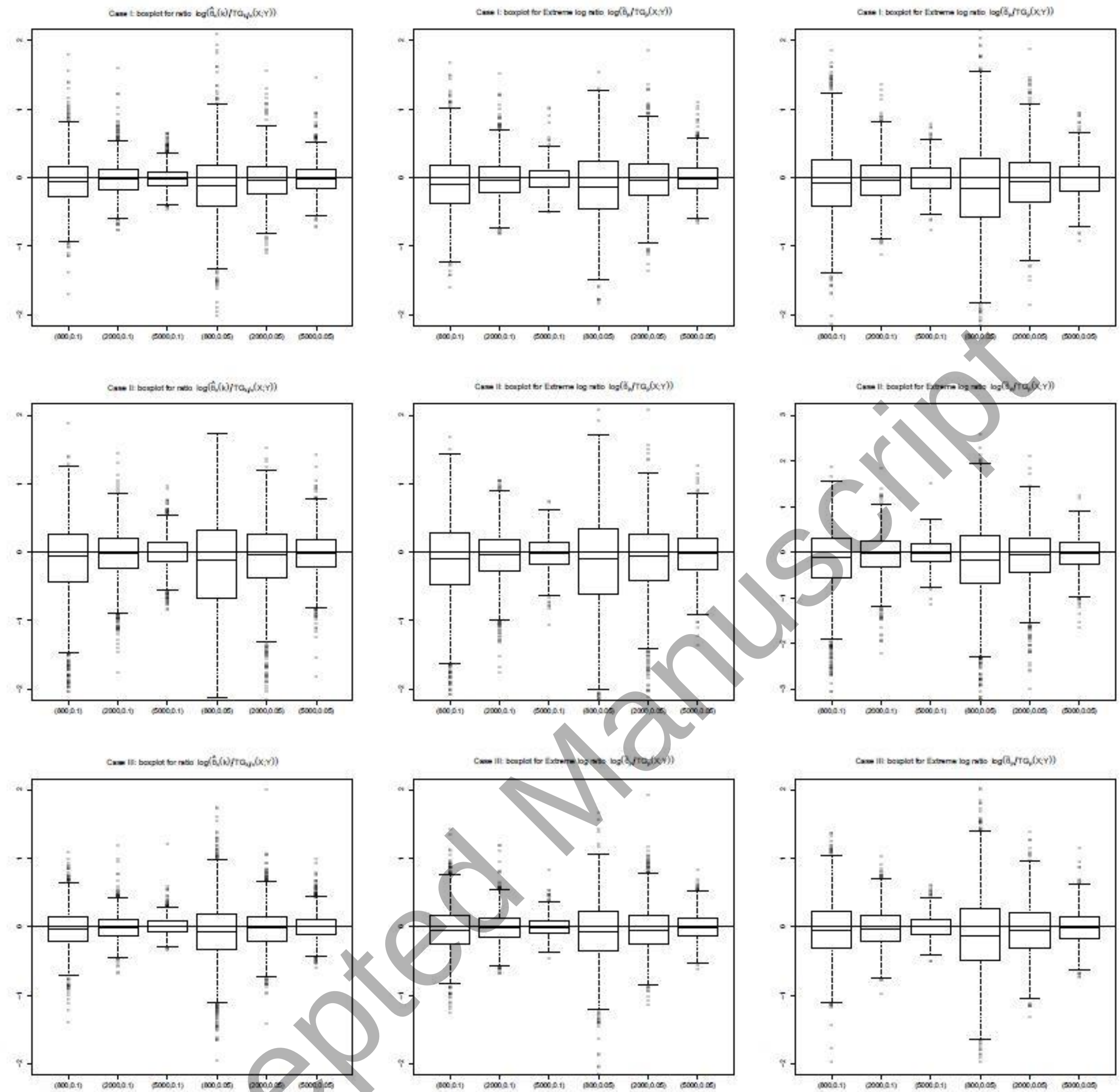

Fig. 1 Simulation Study: Boxplots of log ratios for case I-III. The left column shows cases with intermediate levels for $(n, k/n)$, while the middle and right columns show cases with extreme levels for $(n, k/n, p = 0.01)$ and $(n, k/n, p = 0.001)$ respectively. The first row shows the result of case I while the second and last row show the results for case II and case III separately.

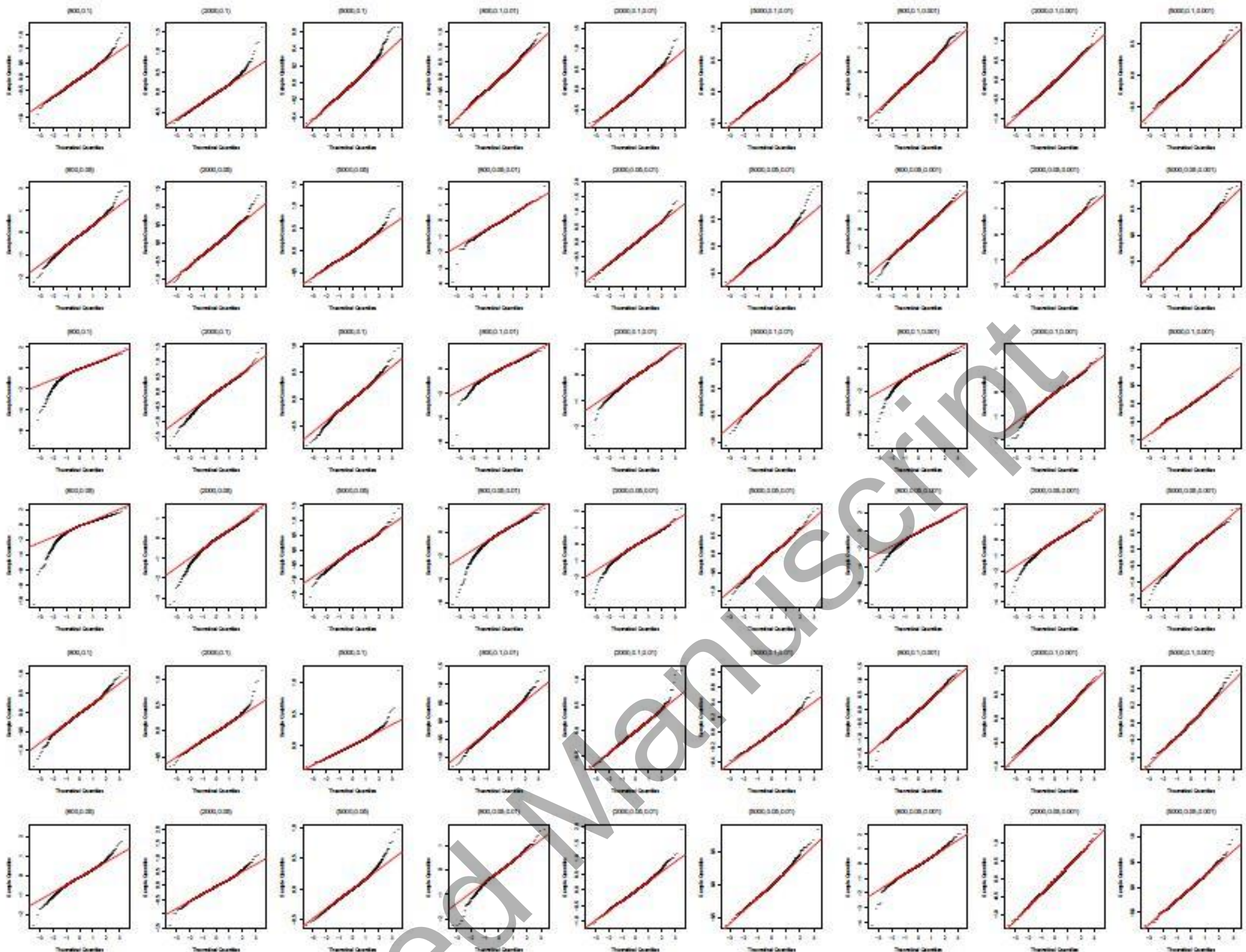

Fig. 2 Simulation Study: QQ-plots of log ratios for case I-III. The left three columns show cases with intermediate levels for $(n, k/n)$, while the middle and right three columns show cases with extreme levels for $(n, k/n, p = 0.01)$ and $(n, k/n, p = 0.001)$ separately. The first two rows show the results for case I, while the middle and last two rows are the results for case II and case III separately.

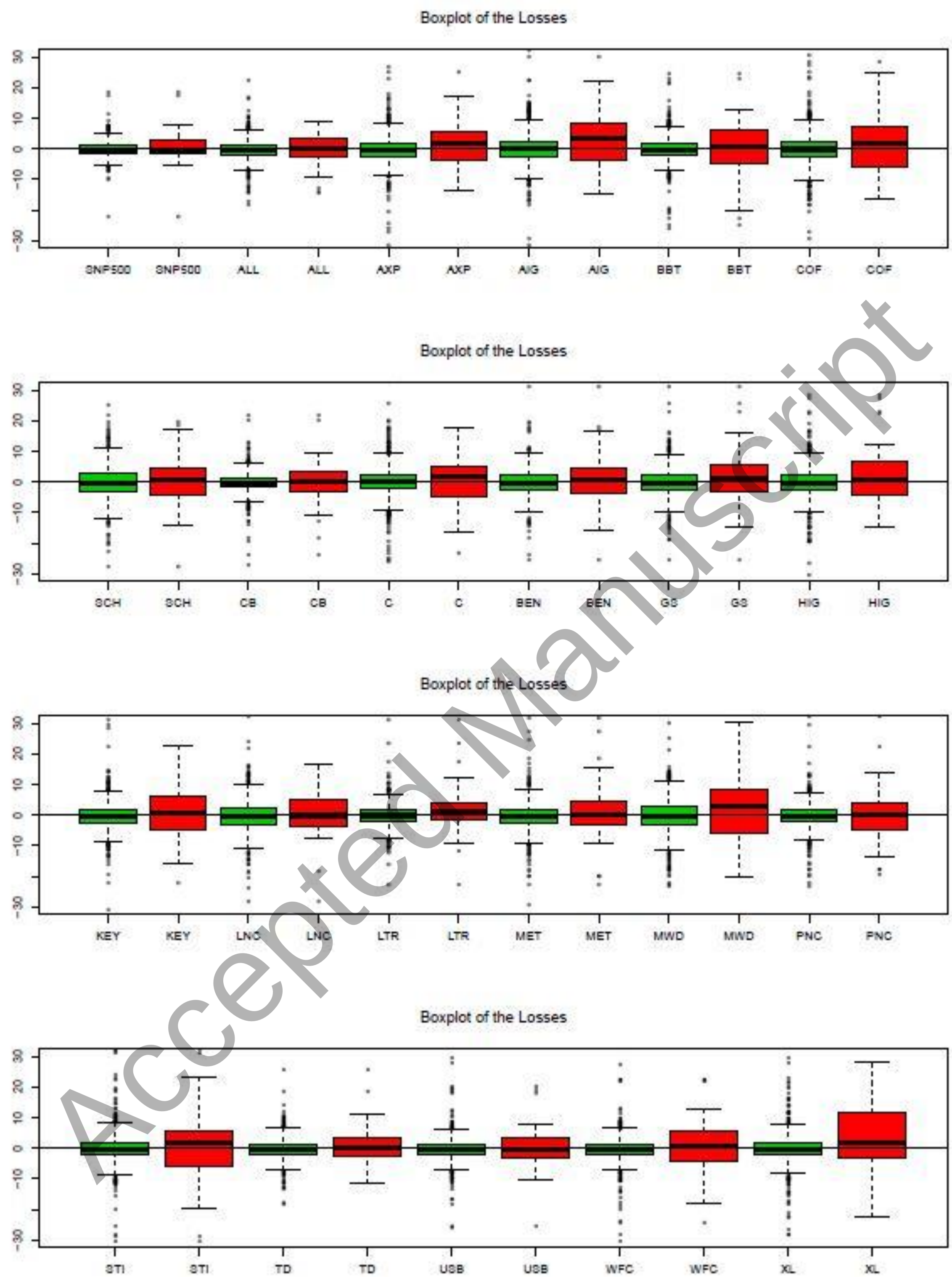

Fig. 3 Real Data Analysis: Boxplots of the losses.

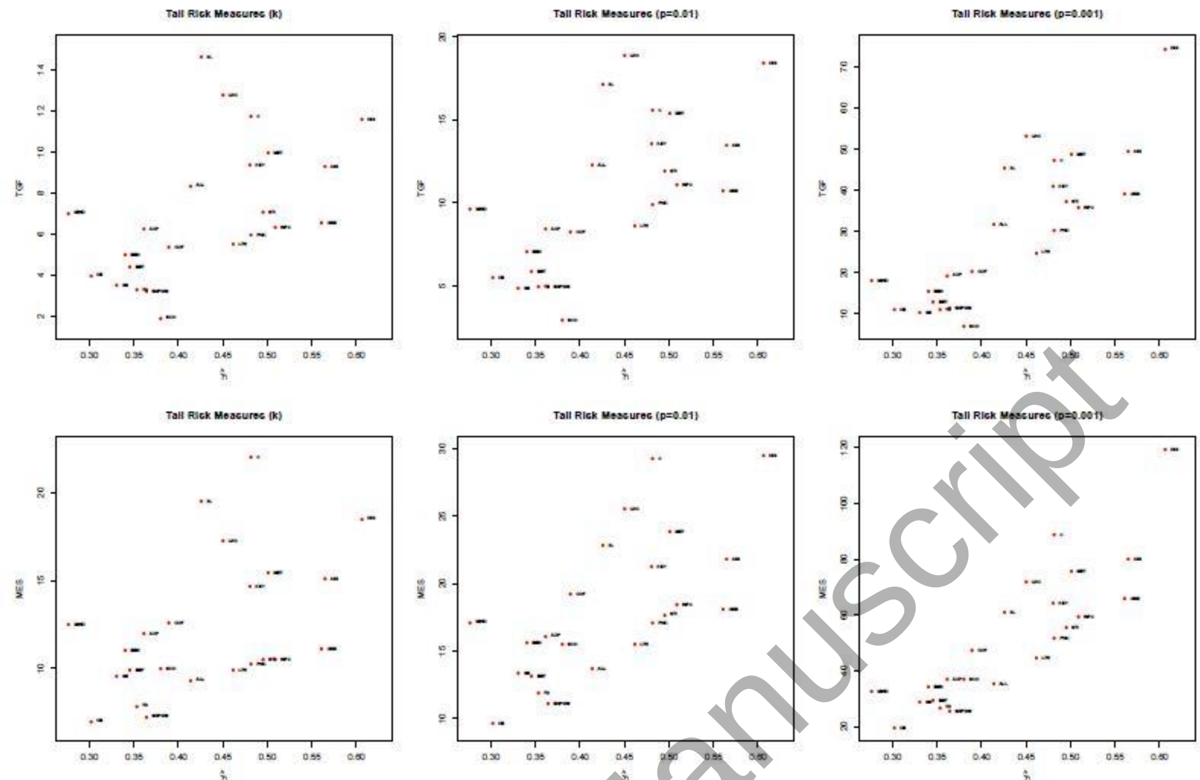

Fig. 4 Real Data Analysis: Compare tail risk measures and the extreme value index for the companies.

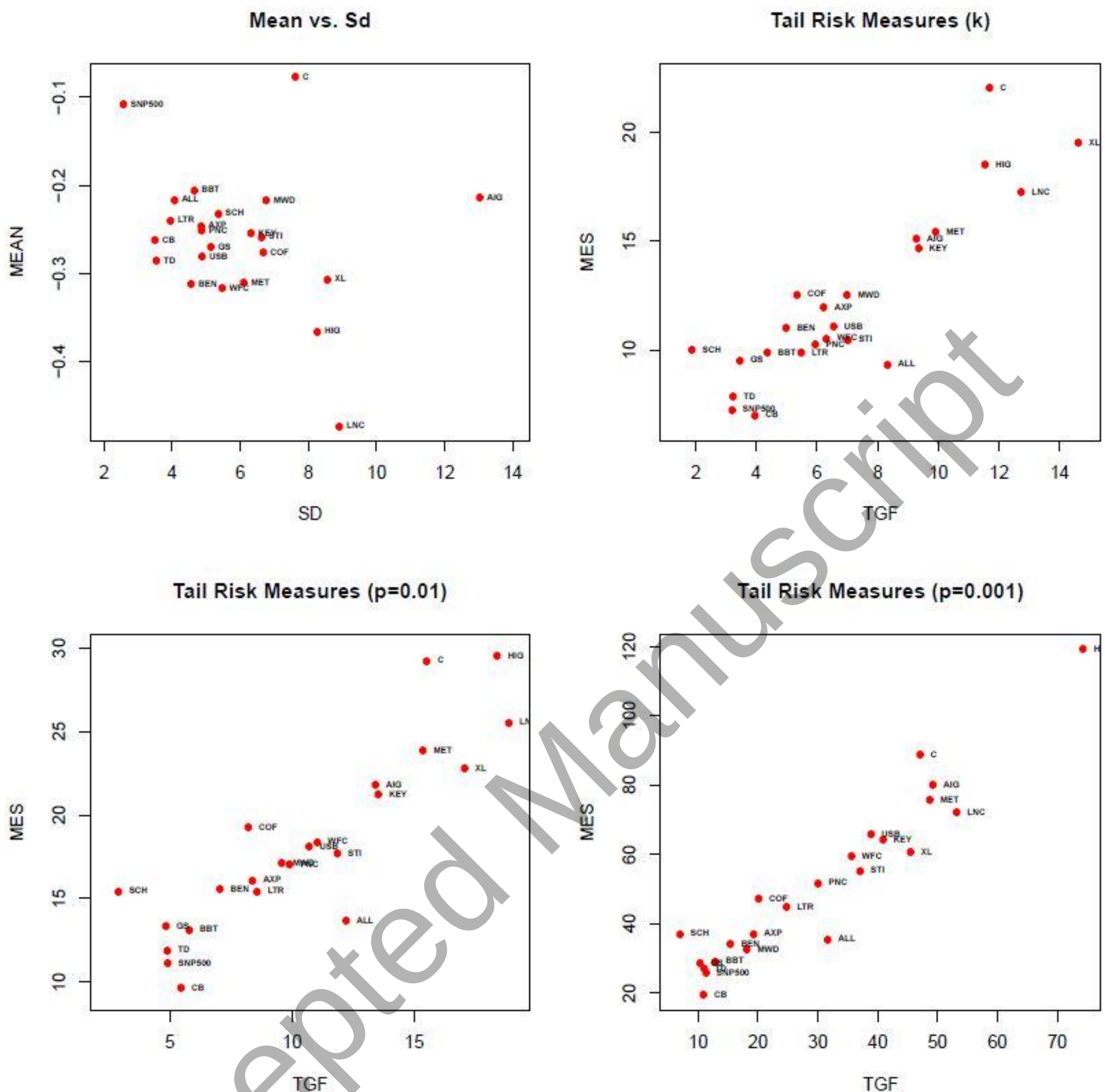

Fig. 5 Real Data Analysis: Compare tail risk measures of stock loss for the companies.

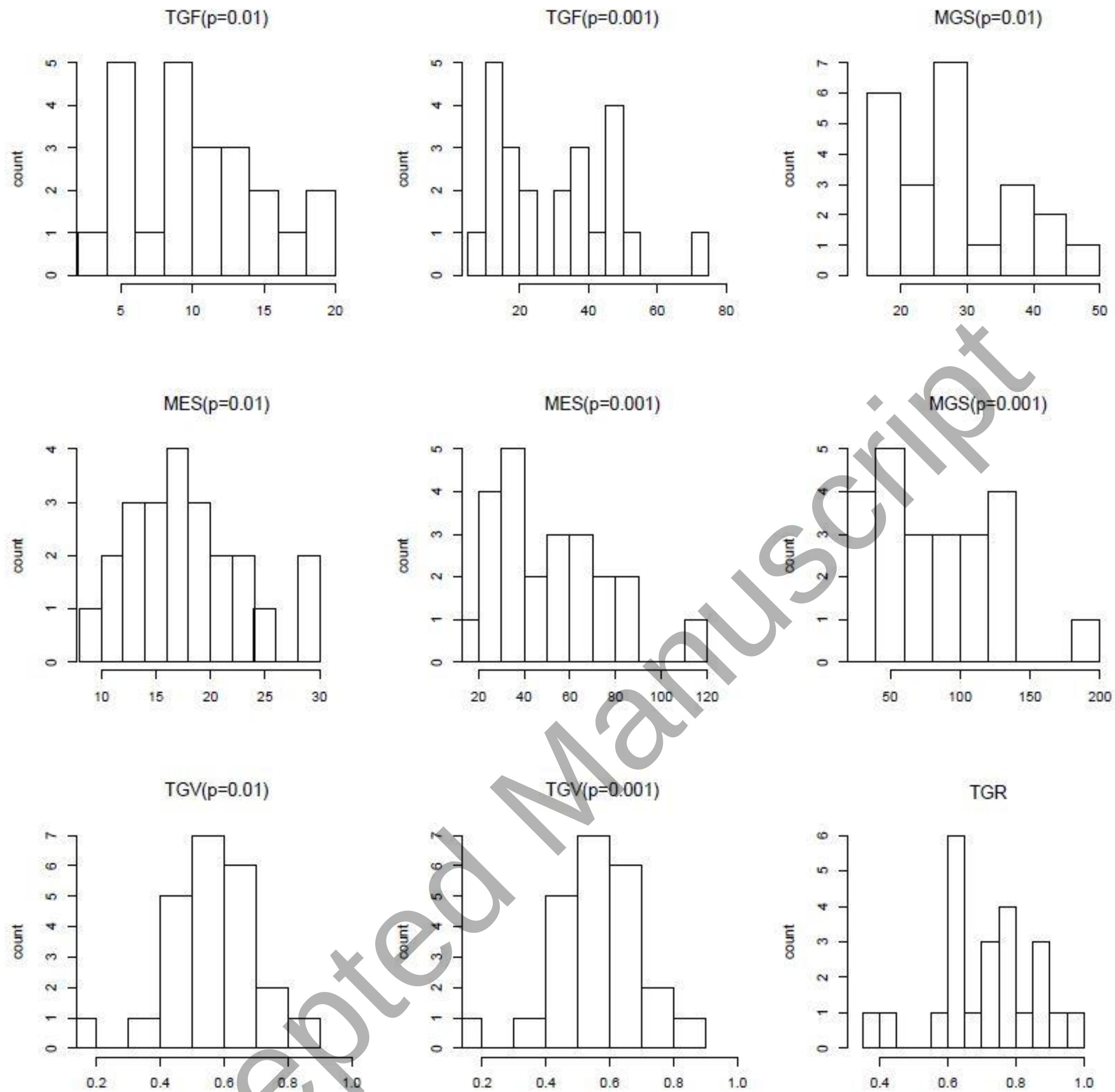

Fig. 6 Real Data Analysis: Histogram graphs of the tail risk measures for the stock loss of all the companies.

Table 1 Summary of Names and Mathematical Symbols of Measures/Coefficients for Tail Risk Analysis of Systemic Risk.

Name	Short Name	Mathematical Symbols	Definition
Tail Gini Functional	TGF	$\mathrm{TG}_\alpha(X;Y)$	(2.2)
Marginal Expected Shortfall	MES	$\mathrm{ES}_\alpha(X;Y)$	(4.1)
Tail Gini Co-variation	TGV	$\mathrm{TGV}_\alpha(X;Y)$	(4.7)
Tail Gini Correlation	TGR	$\mathrm{TGR}_\alpha(X;Y)$	(4.8)
Marginal Gini Shortfall	MGS	$\mathrm{GS}_\alpha(X;Y,\lambda)$	(4.9)

Table 2 Approximation to the true values of Tail Gini Functional

TGF	$\alpha=0.1$	$\alpha=0.05$	p = 0.01		p = 0.001	
			$\alpha=0.1$	$\alpha=0.05$	$\alpha=0.1$	$\alpha=0.05$
Case I	1.4709	1.9334	3.7061	3.6732	9.3574	9.2135
Case II	0.8162	1.0303	2.0905	1.8837	5.3427	4.4667
Case III	1.2613	1.3475	3.2343	2.5746	8.2936	6.4949

Table 3 Simulation Study: Means of the ratios of the proposed estimators of Tail Gini Functional and the true values for $\alpha=0.1,0.05, p=0.01,0.001$ and sample size $n=800,2000,5000$ are reported with corresponding standard deviation given in the brackets.

$\hat{\theta}_n(k)$	$\alpha = 0.1$			$\alpha = 0.05$		
	n = 800	n = 2000	n = 5000	n = 800	n = 2000	n = 5000
Case I	1.0038(0.4016)	1.0092(0.2736)	1.0063(0.1599)	1.0040(0.5722)	1.0177(0.3507)	1.0141(0.2379)
Case II	1.0114(0.5616)	1.0278(0.3591)	1.0255(0.2331)	1.0119(0.7663)	1.0305(0.5020)	1.0160(0.3258)
Case III	1.0005(0.2843)	1.0040(0.1910)	1.0115(0.1300)	1.0278(0.5510)	1.0155(0.3275)	1.0200(0.1983)
$\tilde{\theta}_p(p = 0.01)$	$\alpha = 0.1$			$\alpha = 0.05$		
	n = 800	n = 2000	n = 5000	n = 800	n = 2000	n = 5000
Case I	1.0081(0.4753)	1.0219(0.3218)	1.0186(0.2040)	1.0187(0.6323)	1.0407(0.4150)	1.0219(0.2649)
Case II	1.0327(0.6135)	1.0126(0.3770)	1.0050(0.2355)	1.0193(0.8346)	1.0367(0.5839)	1.0264(0.3569)
Case III	1.0088(0.3531)	1.0151(0.2380)	1.0028(0.1481)	1.0319(0.5609)	1.0241(0.3782)	1.0116(0.2142)
$\tilde{\theta}_p(p = 0.001)$	$\alpha = 0.1$			$\alpha = 0.05$		
	n = 800	n = 2000	n = 5000	n = 800	n = 2000	n = 5000
Case I	1.0649(0.6259)	1.0231(0.4150)	1.0217(0.2649)	1.0724(0.8329)	1.0376(0.5046)	1.0317(0.3111)
Case II	1.0388(0.6971)	1.0361(0.5839)	1.0144(0.3569)	1.0696(1.0460)	1.0777(0.6536)	1.0361(0.3841)
Case III	1.0423(0.4667)	1.0166(0.3782)	1.0148(0.2142)	1.0657(0.7621)	1.0251(0.4084)	1.0309(0.2980)

Table 4 Real Data Analysis: Summary statistics and MES, TGF for the companies where k is the optimal intermediate order.

Firms	Ticker	mean(%)	sd(%)	k	$\hat{\gamma}_1$	$\hat{\eta}_n(k)$	$\hat{\theta}_n(k)$
S&P500	SNP500	-0.1085	2.5648	27	0.3638	3.2013	7.2367
Allstate Corporation	ALL	-0.2166	4.0677	21	0.4135	8.3298	9.3148
American Express	AXP	-0.2463	4.8473	19	0.3612	6.2242	11.9435
AIG	AIG	-0.2145	13.0303	16	0.5646	9.2806	15.0828
BB&T Corporation	BBT	-0.2056	4.6462	19	0.3455	4.3588	9.8665
Capital One	COF	-0.2764	6.6715	25	0.3893	5.3462	12.5370
Charles Schwab	SCH	-0.2320	5.3550	26	0.3801	1.8779	10.0017
Chubb Corporation	CB	-0.2625	3.4969	24	0.3017	3.9589	6.9842
Citigroup	C	-0.0777	7.6221	15	0.4816	11.6944	22.0130
Franklin Resources	BEN	-0.3119	4.5585	23	0.3399	4.9858	10.9959
Goldman Sachs	GS	-0.2702	5.1328	23	0.3305	3.4521	9.5197
Hartford Financial Services	HIG	-0.3658	8.2655	18	0.6060	11.5395	18.5009
Keycorp	KEY	-0.2551	6.3176	18	0.4803	9.3506	14.6699
Lincoln National	LNC	-0.4731	8.9078	20	0.4499	12.7341	17.2130
Loews Corporation	LTR	-0.2400	3.9540	22	0.4613	5.4761	9.8504
MetLife	MET	-0.3110	6.1031	20	0.5007	9.9073	15.3929
Morgan Stanley	MWD	-0.2176	6.7548	26	0.2765	6.9940	12.4850
PNC Financial	PNC	-0.2519	4.8686	26	0.3801	1.8779	10.0017
SunTrust Banks	STI	-0.2587	6.6256	24	0.4949	7.0211	10.4634
Toronto Dominion Bank	TD	-0.2853	3.5340	27	0.3530	3.2257	7.8509

Firms	Ticker	mean(%)	sd(%)	k	$\hat{\gamma}_1$	$\hat{\eta}_n(k)$	$\hat{\theta}_n(k)$
UBS	USB	-0.2810	4.8749	20	0.5608	6.5429	11.0576
Wells Fargo	WFC	-0.3172	5.4643	25	0.5085	6.3129	10.5131
XL Capital	XL	-0.3069	8.5637	12	0.4254	14.6109	19.4920

Table 5 Real Data Analysis: Comparison of tail risk measures with extreme levels for the companies where $\lambda = 1$ in MGS.

Stock X		$p = 0.01$				$p = 0.001$			
	TGR	MES	TGF	TGV	MGS	MES	TGF	TGV	MGS
SNP500	1.0000	11.1002	4.9104	0.4424	16.0106	25.6514	11.3475	0.4424	36.9989
ALL	0.8442	13.6532	12.2094	0.8943	25.8626	35.3801	31.6387	0.8943	67.0188
AXP	0.8755	16.0874	8.3837	0.5211	24.4712	36.9574	19.2598	0.5211	56.2172
AIG	0.3737	21.8044	13.4164	0.6153	35.2208	80.0164	49.2348	0.6153	129.2512
BBT	0.7822	13.1186	5.7955	0.4418	18.9140	29.0651	12.8403	0.4418	41.9054
COF	0.5655	19.2313	8.2010	0.4264	27.4323	47.1328	20.0992	0.4264	67.2319
SCH	0.4097	15.4159	2.8945	0.1878	18.3104	36.9890	6.9450	0.1878	43.9340
CB	0.7691	9.6111	5.4479	0.5668	15.0590	19.2522	10.9128	0.5668	30.1650

Stock X		p = 0.01				p = 0.001			
C	0.6499	29.2218	15.5241	0.5312	44.7459	88.5777	47.0568	0.5312	135.6345
BEN	0.7033	15.5298	7.0415	0.4534	22.5713	33.9692	15.4023	0.4534	49.3715
GS	0.6034	13.3174	4.8293	0.3626	18.1467	28.5075	10.3377	0.3626	38.8452
HIG	0.6513	29.5103	18.4063	0.6237	47.9166	119.1129	74.2937	0.6237	193.4066
KEY	0.7128	21.2391	13.5378	0.6374	34.7769	64.1802	40.9086	0.6374	105.0887
LNC	0.8552	25.5260	18.8840	0.7398	44.4101	71.9215	53.2072	0.7398	125.1287
LTR	0.9118	15.4171	8.5708	0.5559	23.9879	44.5930	24.7904	0.5559	69.3834
MET	0.7873	23.8655	15.3604	0.6436	39.2259	75.5868	48.6495	0.6436	124.2363
MWD	0.7021	17.1035	9.5812	0.5602	26.6848	32.3309	18.1114	0.5602	50.4423
PNC	0.6108	17.0302	9.8995	0.5813	26.9296	51.6327	30.0136	0.5813	81.6463
STI	0.6301	17.6659	11.8541	0.6710	29.5199	55.2184	37.0524	0.6710	92.2707
TD	0.8571	11.8901	4.8853	0.4109	16.7754	26.8013	11.0118	0.4109	37.8131
USB	0.6438	18.0718	10.6932	0.5917	28.7650	65.7432	38.9006	0.5917	104.6438

Stock X		$p = 0.01$				$p = 0.001$			
WFC	0.6487	18.3840	11.0392	0.6005	29.4232	59.2853	35.5998	0.6005	94.8851
XL	0.7913	22.7667	17.0656	0.7496	39.8323	60.6346	45.4508	0.7496	106.0854