Research Article

Risk Measurement by G-Expected Shortfall

Ziting Pei , Xuhui Wang , and Xingye Yue

Center for Financial Engineering and Math Center for Interdisciplinary Research School of Mathematical Sciences,
Soochow University, Suzhou 215006, China

Correspondence should be addressed to Ziting Pei; ziting_pei@163.com

Received 11 December 2020; Revised 29 March 2021; Accepted 9 April 2021; Published 21 April 2021

Academic Editor: Mostafa M. A. Khater

Copyright © 2021 Ziting Pei et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

G-expected shortfall (G-ES), which is a new type of worst-case expected shortfall (ES), is defined as measuring risk under infinite distributions induced by volatility uncertainty. Compared with extant notions of the worst-case ES, the G-ES can be computed using an explicit formula with low computational cost. We also conduct backtests for the G-ES. The empirical analysis demonstrates that the G-ES is a reliable risk measure.

1. Introduction

The Basel Committee on Banking Supervision publicly released the new market risk framework, “Fundamental Review of the Trading Book” (FRTB) on January 14, 2016, to address the shortcomings of the prior market risk capital framework, Basel 2.5, and to design a minimum capital standard to apply market risk more uniformly across jurisdictions. The FRTB suggests using expected shortfall (ES) at the 97.5% confidence level to replace the 10-day value-at-risk (VaR) and stressed VaR at the 99% confidence level because the ES is a coherent risk measure that satisfies all axioms proposed in Artzner et al. [1] and it prioritizes tail risk to a greater degree. However, ES is more sensitive than VaR in estimating errors on distributions. If there is no good model for the tail of the distribution, then the ES value may be quite misleading; that is, the accuracy of the ES estimation is heavily affected by the accuracy of the tail modelling. An alternative method is to consider the worst-case ES.

This paper presents a new and simple method to calculate the worst-case ES when facing infinite distributions induced by volatility uncertainty. We employ a newly developed probability theory, G-expectation (G-normal distribution) established by Peng [2] to define the worst-case ES, which we call the G-ES. The G-expectation is a sublinear expectation that is the supremum of a set of linear

expectations. The G-normal distribution is a distribution defined under the sublinear expectation. Hence, the quantiles of the G-normal distribution and the average of their tails are natural candidates for the worst-case VaR and ES. We explain the advantages of using the theory of sublinear expectation to characterize the worst-case risk exposure in Section 2 in detail.

The G-ES can be easily backtested since it has an explicit formula, whereas the ES is usually difficult to be backtested when the model is uncertain, although significant progress was made in this direction. For instance, Du and Escanciano [3] adjust the conditional backtests (Christoffersen [4] and Berkowitz et al. [5]) and the unconditional backtest (Kupiec [6]) for VaR to ES.

Our contribution is threefold:

- (1) We provide an explicit formula to compute the G-ES, making it easy to conduct a backtest.
- (2) The G-ES method can be applied to high-dimensional portfolio risk management.
- (3) Dynamic backtests are conducted for worldwide indexes. The G-ES performs robustly and reliably in the backtests.

Many papers investigate the worst-case ES. Under a moment-cone uncertainty set, Natarajan et al. [7] define the worst-case conditional VaR measure (CVaR (CVaR is an

equivalent concept to ES)) using a semidefinite programming method that generalizes the idea of the worst-case VaR introduced by Ghaoui et al. [8] to the worst-case CVaR. Zhu and Fukushima [9] study the worst-case CVaR under mixture distribution uncertainty, box uncertainty, and ellipsoidal uncertainty. An explicit formula is not given, but it can be solved efficiently by numerical methods. Compared with Zhu and Fukushima [9] and Natarajan et al. [7], the G-ES has an explicit formula with a low computational cost. We assume that the mean and volatility vary within bounded intervals and that the return follows the G-normal distribution, which is a worst-case distribution induced by mean and volatility uncertainty. This kind of worst-case distribution covers not only a set of normal distributions but also some other distributions, although we cannot specify all of them. In previous work, Pei et al. [10] show that the G-normal distribution covers some bimodal skew-normal distributions.

Chen et al. [11] and Natarajan et al. [12] study the worst-case CVaR for positions with uncertain distributions, but with given means and variances, and obtain closed-form expressions. However, their worst-case CVaR often heavily overestimates risks. The empirical tests of nine international indexes from 2007 to 2018 suggest that the G-ES is robust and reliable most of the time.

The rest of this paper is organized as follows. In Section 2, we explain the basics of the G-normal distribution and the G-ES. We provide the definition and several properties of the G-ES in Section 3. The empirical backtests are conducted in Section 4. Section 5 concludes the paper.

2. G-Normal Distribution

2.1. Background. Peng [13–15] established G-stochastic analysis, which is a path analysis that extends the classical Wiener analysis to a framework of sublinear expectation on events field $\Omega = C_0([0, +\infty), \mathbb{R})$, the space of all $\mathbb{R}$ -valued continuous paths $(\omega_t)_{t \in \mathbb{R}^+}$ with $\omega_0 = 0$, equipped with a uniform norm on compact subspaces. Notions such as the G-normal distribution, G-Brownian motion, and G-expectation were introduced (see Appendix or Peng's review paper [16] and summative book [2]). The representation for G-expectation [17],

$$\mathbb{E}[\cdot] = \sup_{P \in \mathcal{P}} E_P[\cdot], \quad (1)$$

indicates that G-expectation naturally induces a weakly compact set of probabilities $\mathcal{P}$. The G-Brownian motion (B_t) is a martingale under each $P \in \mathcal{P}$ [18], and there exists a

unique adapted process (σ_t^P) such that $\underline{\sigma} \leq \sigma_t^P \leq \bar{\sigma}$ almost surely (a.s.) and

$$B_t = \int_0^t \sigma_s^P dW_s^P, \quad \forall t \geq 0, \quad P - \text{a.s.}, \quad (2)$$

where (W_t^P) is a standard Brownian motion under the linear expectation E_P . There are infinite probability measures in the set $\mathcal{P}$, as one process (σ_t^P) induces one probability measure. In this paper, we aim to identify the worst-case ES in the setting of ambiguous volatility.

A G-normal distributed random variable X is characterized by the solution of a Hamilton–Jacobi–Bellman equation (HJB). Let $u(t, x) := \mathbb{E}[\varphi(x + \sqrt{t} X)]$, where φ is bounded and Lipschitz-continuous. In particular, $u(1, 0) = \mathbb{E}[\varphi(X)]$. Peng [2] shows that a random variable X is G-normal distributed if u is the viscosity solution of the following partial differential equation (PDE):

$$\begin{cases} \partial_t u - G(D^2 u) = 0, \\ (t, x) \in [0, \infty) \times \mathbb{R}^d, \\ u(0, x) = \varphi(x), \end{cases} \quad (3)$$

where $D^2 u$ is the Hessian matrix of u ; that is, $D^2 u = (\partial_{x^i x^j}^2 u)_{i,j=1}^d$ and $G(A) = (1/2) \sup_{\gamma \in \Gamma} \text{tr}[\gamma \gamma^\top A]$, $A \in \mathbb{S}(d)$; $\mathbb{S}(d)$ denotes the space of $d \times d$ symmetric matrices; Γ represents the set of all possible covariance matrices, which is a given bounded, closed, and convex subset of $\mathbb{R}^{d \times d}$. PDE (3) is the so-called G-heat equation. If Γ is a singleton, then it is the classical heat equation. When $d = 1$, the volatility belongs to an interval $[\underline{\sigma}, \bar{\sigma}]$, where $\bar{\sigma}^2 = \mathbb{E}[X^2]$ and $\underline{\sigma}^2 = -\mathbb{E}[-X^2]$, and the mean is $\mathbb{E}[X] = \mathbb{E}[-X] = 0$. In this case, we denote $X \sim \mathcal{N}(0, [\underline{\sigma}^2, \bar{\sigma}^2])$.

2.2. The Worst-Case Distribution. Note that to obtain the best-case distribution function $\bar{F}(v) := \sup_{P \in \mathcal{P}} P(X \leq v)$ and the worst-case distribution $\underline{F}(v) := \inf_{P \in \mathcal{P}} P(X \leq v)$ (given a confidence level $\alpha \in (0, 1)$ and for the same position X , the VaR obtained by $\underline{F}(v)$ is always greater than or equal to the one obtained by $\bar{F}(v)$), so we identify $\bar{F}(v)$ as the best-case distribution while $\underline{F}(v)$ as the worst-case distribution) is equivalent to solving the PDE (3) with the initial condition of indicator type. We find that the “similar solutions” method (Bluman and Cole [19]) works for the PDE (3) with initial condition $\varphi(x) = 1_{\{x \leq v\}}$ or $\varphi(x) = 1_{\{x > v\}}$ (see Pei et al. [10]). We assume that X follows a 1-dimensional G-normal distribution with $\underline{\sigma} > 0$. Pei et al. [10] show that

$$\begin{aligned}\bar{F}(v) &:= \sup_{P \in \mathcal{P}} P(X \leq v) \\ &= \begin{cases} \frac{\bar{\sigma} - \underline{\sigma}}{\bar{\sigma} + \underline{\sigma}} + \frac{2\underline{\sigma}}{\bar{\sigma} + \underline{\sigma}} N\left(\frac{v}{\underline{\sigma}}\right), & \text{if } v \geq 0, \\ \frac{2\bar{\sigma}}{\bar{\sigma} + \underline{\sigma}} N\left(\frac{v}{\bar{\sigma}}\right), & \text{if } v < 0, \end{cases} \quad (4)\end{aligned}$$

$$\begin{aligned}\underline{F}(v) &:= \inf_{P \in \mathcal{P}} P(X \leq v) \\ &= \begin{cases} \frac{\underline{\sigma} - \bar{\sigma}}{\bar{\sigma} + \underline{\sigma}} + \frac{2\bar{\sigma}}{\bar{\sigma} + \underline{\sigma}} N\left(\frac{v}{\bar{\sigma}}\right), & \text{if } v \geq 0, \\ \frac{2\underline{\sigma}}{\bar{\sigma} + \underline{\sigma}} N\left(\frac{v}{\underline{\sigma}}\right), & \text{if } v < 0, \end{cases} \quad (5)\end{aligned}$$

where $N(\cdot)$ is the distribution function of the standard normal distribution. When $\underline{\sigma} = 0$, the best- and worst-case distribution functions are the limits of those in (4) and (5), respectively, as $\underline{\sigma}$ converges to zero. It is easy to check that the functions $\bar{F}(\cdot)$ and $\underline{F}(\cdot)$ are distribution functions for some random variables with the corresponding best/worst-case density functions:

$$\bar{f}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\bar{\sigma}} \cdot \frac{2\bar{\sigma}}{\bar{\sigma} + \underline{\sigma}} \cdot e^{-(x^2/2\bar{\sigma}^2)}, & x < 0; \\ \frac{1}{\sqrt{2\pi}\underline{\sigma}} \cdot \frac{2\underline{\sigma}}{\bar{\sigma} + \underline{\sigma}} \cdot e^{-(x^2/2\underline{\sigma}^2)}, & x \geq 0, \end{cases} \quad (6)$$

$$\underline{f}(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\underline{\sigma}} \cdot \frac{2\underline{\sigma}}{\bar{\sigma} + \underline{\sigma}} \cdot e^{-(x^2/2\underline{\sigma}^2)}, & x < 0; \\ \frac{1}{\sqrt{2\pi}\bar{\sigma}} \cdot \frac{2\bar{\sigma}}{\bar{\sigma} + \underline{\sigma}} \cdot e^{-(x^2/2\bar{\sigma}^2)}, & x \geq 0. \end{cases} \quad (7)$$

A close look reveals that $\bar{f}(x) = \underline{f}(-x)$.

The best/worst-case distributions have more skewness than the normal distribution does. Figure 1 compares the best/worst-case density functions with a family of normal density functions with volatility varying within $[\underline{\sigma}, \bar{\sigma}] = [0.01, 0.03]$. A look at Figure 1 suggests that uncertain volatility results in a skewed density function. For example, the best-case density function $\bar{f}$ has a left fat tail, such that the best-case distribution function $\bar{F}(v) := \sup_{\sigma \in [\underline{\sigma}, \bar{\sigma}]} P(X \leq v)$ is greater than any normal distribution with volatility in $[\underline{\sigma}, \bar{\sigma}]$. We use the best-case density function to calculate the G-ES for return data and the worst-case density function to calculate the G-ES for loss data. It is worth noting that the best- (worst-) case distribution function is not the maxima (minima) of the set of normal distributions, which suggests that the set $\mathcal{P}$ not only includes normal distributions, but some other distributions that are not normally distributed. Pei et al. [10] specify that

FIGURE 1: The best-case density function $\bar{f}$, the worst-case density function $\underline{f}$, and density functions for a family of zero-mean normal distributions with standard deviations varying in $[\underline{\sigma}, \bar{\sigma}]$.

the set $\mathcal{P}$ may include some bimodal skew-normal distributions.

The following proposition calculates the skewness and kurtosis of the worst-case density function $\underline{f}$.

Proposition 1. Let X be a random variable with density function $\underline{f}$. Then,

(i) Its mean is $\mu = \sqrt{(2/\pi)}(\bar{\sigma} - \underline{\sigma})$ and the standard deviation is

$$\sigma = \left(\left(1 - \frac{2}{\pi}\right)\bar{\sigma}^2 + \left(\frac{4}{\pi} - 1\right)\bar{\sigma}\underline{\sigma} + \left(1 - \frac{2}{\pi}\right)\underline{\sigma}^2 \right)^{(1/2)}. \quad (8)$$

(ii) Its skewness is

$$\begin{aligned} & \frac{\sqrt{2\pi}(1-k)}{\left(\pi\left((1-(2/\pi))(1+k^2) + ((4/\pi)-1)k\right)\right)^{(1/2)}} \\ & - \frac{2\sqrt{2}(\pi-3)(1-k)^3}{\left(\pi\left((1-(2/\pi))(1+k^2) + ((4/\pi)-1)k\right)\right)^{(3/2)}}, \end{aligned} \quad (9)$$

and kurtosis is

$$\begin{aligned} & \frac{(3\pi-4)(1+k^5) - 8k(1+k^3)}{\pi(1+k)\left((1-(2/\pi))(1+k^2) + ((4/\pi)-1)k\right)^2} \\ & + \frac{12(\pi k^2 - (1-k)^4)}{\pi^2\left((1-(2/\pi))(1+k^2) + ((4/\pi)-1)k\right)^2}, \end{aligned} \quad (10)$$

where $k = \underline{\sigma}/\bar{\sigma}$. Moreover, both the skewness and kurtosis are decreasing with respect to k .

Proof. Let E be the linear expectation under which X has the worst-case density $\underline{f}$. From equation (7), we can get

$$\begin{aligned}
\mu &= E[X] \\
&= \frac{2\underline{\sigma}}{\bar{\sigma} + \underline{\sigma}} \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\underline{\sigma}}} e^{-(x^2/2\underline{\sigma}^2)} x dx + \frac{2\bar{\sigma}}{\bar{\sigma} + \underline{\sigma}} \int_0^{+\infty} \frac{1}{\sqrt{2\pi\bar{\sigma}}} e^{-(x^2/2\bar{\sigma}^2)} x dx \\
&= -\frac{2\underline{\sigma}^3}{\bar{\sigma} + \underline{\sigma}} \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\underline{\sigma}}} d\left(e^{-(x^2/2\underline{\sigma}^2)}\right) - \frac{2\bar{\sigma}^3}{\bar{\sigma} + \underline{\sigma}} \int_0^{+\infty} \frac{1}{\sqrt{2\pi\bar{\sigma}}} d\left(e^{-(x^2/2\bar{\sigma}^2)}\right) \\
&= \frac{-2\underline{\sigma}^2}{\sqrt{2\pi}(\bar{\sigma} + \underline{\sigma})} + \frac{2\bar{\sigma}^2}{\sqrt{2\pi}(\bar{\sigma} + \underline{\sigma})} \\
&= \sqrt{\frac{2}{\pi}}(\bar{\sigma} - \underline{\sigma}), \\
E[X^2] &= \frac{2\underline{\sigma}}{\bar{\sigma} + \underline{\sigma}} \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\underline{\sigma}}} e^{-(x^2/2\underline{\sigma}^2)} x^2 dx + \frac{2\bar{\sigma}}{\bar{\sigma} + \underline{\sigma}} \int_0^{+\infty} \frac{1}{\sqrt{2\pi\bar{\sigma}}} e^{-(x^2/2\bar{\sigma}^2)} x^2 dx \\
&= \frac{\underline{\sigma}^3}{\bar{\sigma} + \underline{\sigma}} + \frac{\bar{\sigma}^3}{\bar{\sigma} + \underline{\sigma}} \\
&= \underline{\sigma} - \bar{\sigma} \underline{\sigma} + \bar{\sigma}^2,
\end{aligned} \tag{11}$$

then, the standard deviation is $\sigma = (E[X^2] - (E[X])^2)^{(1/2)}$
 $2) = ((1 - (2/\pi))\bar{\sigma}^2 + ((4/\pi) - 1)\bar{\sigma}\underline{\sigma} + (1 - (2/\pi))\underline{\sigma}^2)^{(1/2)}$.

For (ii), by direct calculation, the skewness is

$$\begin{aligned}
E\left[\frac{(X - \mu)^3}{\sigma^3}\right] &= \frac{E[X^3] - 3\mu\sigma^2 - \mu^3}{\sigma^3} \\
&= \frac{1}{\sigma^3} \left(-\frac{2\underline{\sigma}}{\bar{\sigma} + \underline{\sigma}} \int_{-\infty}^0 \frac{1}{\sqrt{2\pi\underline{\sigma}}} x^2 d\left(e^{-(x^2/2\underline{\sigma}^2)}\right) - \frac{2\bar{\sigma}^3}{\bar{\sigma} + \underline{\sigma}} \int_0^{+\infty} \frac{1}{\sqrt{2\pi\bar{\sigma}}} x^2 d\left(e^{-(x^2/2\bar{\sigma}^2)}\right) \right) \\
&\quad - \frac{(3\mu\sigma^2 + \mu^3)}{\sigma^3} = \frac{4(\bar{\sigma} - \underline{\sigma})(\bar{\sigma}^2 + \underline{\sigma}^2)}{\sqrt{2\pi}\sigma^3} - \frac{(3\mu\sigma^2 + \mu^3)}{\sigma^3} \\
&= \frac{\sqrt{2\pi}(1 - k)}{\left(\pi((1 - (2/\pi))(1 + k^2) + ((4/\pi) - 1)k)\right)^{(1/2)}} - \frac{2\sqrt{2}(\pi - 3)(1 - k)^3}{\left(\pi((1 - (2/\pi))(1 + k^2) + ((4/\pi) - 1)k)\right)^{(3/2)}}.
\end{aligned} \tag{12}$$

The kurtosis is

$$\begin{aligned}
E\left[\frac{(X - \mu)^4}{\sigma^4}\right] &= \frac{E[X^4] - 4\mu E[X^3] + 6\mu^2 E[X^2] - 4\mu^3 E[X] + \mu^4}{\sigma^4} \\
&= \frac{3\sqrt{\pi}(\bar{\sigma}^5 + \underline{\sigma}^5) + 8\sqrt{2}\mu(\underline{\sigma}^4 - \bar{\sigma}^4) + 6\sqrt{\pi}\mu^2(\bar{\sigma}^3 + \underline{\sigma}^3) + 4\sqrt{2}\mu^3(\underline{\sigma}^2 - \bar{\sigma}^2)}{\sqrt{\pi}(\bar{\sigma} + \underline{\sigma})\sigma^4} + \frac{\mu^4}{\sigma^4} \\
&= \frac{(3\pi - 4)(1 + k^5) - 8k(1 + k^3)}{\pi(1 + k)((1 - (2/\pi))(1 + k^2) + ((4/\pi) - 1)k)^2} + \frac{12(\pi k^2 - (1 - k)^4)}{\pi^2((1 - (2/\pi))(1 + k^2) + ((4/\pi) - 1)k)^2}.
\end{aligned} \tag{13}$$

Define $f(k) := E((X - \mu)^3/\sigma^3)$ and $g(k) := E[(X - \mu)^4/\sigma^4]$. Then, taking the derivative of $f(\cdot)$, we have

$$\begin{aligned} f'(k) &= \frac{(1+k)(-\sqrt{2/\pi})((1-(2/\pi))(1+k^2) + ((4/\pi)-1)k) + 3(\pi-3)(2/\pi)^{(3/2)}(1-k)^2}{2((1-(2/\pi))(1+k^2) + ((4/\pi)-1)k)^{(5/2)}} \\ &= (\sqrt{2}/\pi\sqrt{\pi})(1+k)((-16+5\pi)k^2 + (32-11\pi)k + (-16+5\pi))/2((1-(2/\pi))(1+k^2) + ((4/\pi)-1)k)^{\frac{5}{2}} \\ &= \frac{\sqrt{2}}{\pi\sqrt{\pi}} \frac{I_1 \cdot I_2}{I_3} \\ &< 0, \quad \text{for } 0 \leq k \leq 1, \end{aligned} \quad (14)$$

where $I_1 = 1+k > 0$, $I_2 = (-16+5\pi)k^2 + (32-11\pi)k + (-16+5\pi) < 0$, and $I_3 = 2((1-(2/\pi))(1+k^2) + ((4/\pi)-1)k)^{(5/2)} > 0$ for $0 \leq k \leq 1$.

$$\begin{aligned} g'(k) &= \frac{\pi(1-k^2)((64-30\pi+3\pi^2)k^2 + (-128+68\pi-9\pi^2)k + 64-30\pi+3\pi^2)}{((\pi-2)k^2 - (\pi-4)k + \pi-2)^3} \\ &= \pi \frac{I_4 \cdot I_5}{I_6} \\ &\leq 0, \quad \text{for } 0 \leq k \leq 1, \end{aligned} \quad (15)$$

where $I_4 = 1-k^2 \geq 0$, $I_5 = (64-30\pi+3\pi^2)k^2 + (-128+68\pi-9\pi^2)k + 64-30\pi+3\pi^2 < 0$, and $I_6 = ((\pi-2)k^2 - (\pi-4)k + \pi-2)^3 > 0$ for $0 \leq k \leq 1$. Consequently, Proposition 1 holds true.

The variable k measures the uncertainty of the volatility. Figure 2 demonstrates that as the uncertainty of the volatility decreases (k increases), the skewness and kurtosis decrease. When $k = 1$, the equation corresponds to a normal distribution with skewness 0 and kurtosis 3. Hence, the worst-case distribution has the elasticity to fit data by adjusting its skewness and kurtosis. The worst-case distribution enhances a lot in skewness and kurtosis from normal distributions. However, the skewness and kurtosis of $\underline{f}$ are below those of the Gumbel distribution, which has skewness 1.14 and kurtosis 5.4. Thus, we find a distribution with skewness and kurtosis between those of the normal and Gumbel distributions. $\square$

3. G-ES

ES is the average value of the loss greater than the VaR for a given confidence level (Rockafellar and Uryasev [20, 21]). It estimates how much the tail loss exceeds the VaR given the distribution of the investment within a fixed period. Let X be

a random variable that represents the possible loss and let F be its distribution function. For a given confidence level $\alpha \in (0, 1)$, the traditional ES is defined as

$$ES_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_t(X) dt, \quad (16)$$

where VaR_t is the t -quantile for a given distribution. However, for a financial position, future distributions are often not precisely known, even if we have a past distribution in hand. We now define the G-ES, which incorporates the distribution uncertainty.

3.1. Definition of G-ES. Let $\mathcal{P}$ be the set of probability measures induced by processes (σ_t) valued in $[\underline{\sigma}, \bar{\sigma}]$. Since we consider the worst-case scenario, and it only makes sense to consider nonnegative risk exposures, we employ the first part of (5) to calculate the G-ES. The following lemma for G-VaR comes from Pei et al. [10].

Lemma 1. Assume that X follows a 1-dimensional G-normal distribution, that is, $X \sim \mathcal{N}(0, [\underline{\sigma}^2, \bar{\sigma}^2])$. We have the following explicit formula for the G-VaR:

FIGURE 2: The monotonicity of the skewness and kurtosis of $\underline{f}$ with respect to k .

$$G - \text{VaR}_\alpha(X) = -\bar{\sigma}N^{-1}\left(\frac{\bar{\sigma} + \underline{\sigma}}{2\bar{\sigma}}(1 - \alpha)\right), \quad \text{for } \alpha \in \left[\frac{\underline{\sigma}}{\bar{\sigma} + \underline{\sigma}}, 1\right]. \quad (17)$$

Similar to the traditional ES, we can define the G-ES as the average of tail G-VaRs.

Definition 1. Assume that $X \sim \mathcal{N}(0, [\underline{\sigma}^2, \bar{\sigma}^2])$. For a confidence level α , the G-ES is defined as

$$G - \text{ES}_\alpha(X) = \frac{1}{1 - \alpha} \int_\alpha^1 G - \text{VaR}_t(X) dt. \quad (18)$$

By Lemma 1, the G-ES has the following closed form.

Theorem 1. Assume $X \sim \mathcal{N}(0, [\underline{\sigma}^2, \bar{\sigma}^2])$. We obtain the G-ES for X in the following closed form:

$$G - \text{ES}_\alpha(X) = \frac{1}{1 - \alpha} \sqrt{\frac{2}{\pi}} \frac{\bar{\sigma}^2}{\bar{\sigma} + \underline{\sigma}} \exp\left\{-\frac{1}{2} \left[N^{-1}\left(\frac{\bar{\sigma} + \underline{\sigma}}{2\bar{\sigma}}(1 - \alpha)\right)\right]^2\right\}, \quad \alpha \in \left[\frac{\underline{\sigma}}{\bar{\sigma} + \underline{\sigma}}, 1\right]. \quad (19)$$

Proof. From equation (17) and definition (18), for $\alpha \in [\frac{\underline{\sigma}}{\bar{\sigma} + \underline{\sigma}}, 1]$, we have

$$\begin{aligned} G - \text{ES}_\alpha(X) &= \frac{-\bar{\sigma}}{1 - \alpha} \int_\alpha^1 N^{-1}\left(\frac{\bar{\sigma} + \underline{\sigma}}{2\bar{\sigma}}(1 - t)\right) dt \\ &= \sqrt{\frac{2}{\pi}} \frac{\bar{\sigma}^2}{(1 - \alpha)(\bar{\sigma} + \underline{\sigma})} \int_{N^{-1}\left(\frac{\bar{\sigma} + \underline{\sigma}}{2\bar{\sigma}}(1 - \alpha)\right)}^{-\infty} t e^{-(t^2/2)} dt \\ &= \frac{\bar{\sigma}^2}{\sqrt{2\pi}(1 - \alpha)(\bar{\sigma} + \underline{\sigma})} \int_{N^{-1}\left(\frac{\bar{\sigma} + \underline{\sigma}}{2\bar{\sigma}}(1 - \alpha)\right)}^{-\infty} e^{-(t^2/2)} dt \\ &= \frac{1}{1 - \alpha} \sqrt{\frac{2}{\pi}} \frac{\bar{\sigma}^2}{\bar{\sigma} + \underline{\sigma}} \exp\left\{-\frac{1}{2} \left[N^{-1}\left(\frac{\bar{\sigma} + \underline{\sigma}}{2\bar{\sigma}}(1 - \alpha)\right)\right]^2\right\}. \end{aligned} \quad (20)$$

The proof is complete.

From equation (19), we see that we can calculate the G-ES only via several parameters, avoiding the computation of a set of G-VaRs. Its simple operability could be appealing for the financial industries.

Similar to the traditional ES, it is easy to see that G-ES has the following properties. $\square$

Proposition 2. Given a confidence level α , the G-ES satisfies the following axioms:

(1) *Monotonicity:* $G - \text{ES}_\alpha(X) \geq G - \text{ES}_\alpha(Y)$ if $X \geq Y$

(2) *Translation invariance:* $G - \text{ES}_\alpha(X - c) = G - \text{ES}_\alpha(X) - c$ for $c \in \mathbb{R}$

(3) *Homogeneity:* $G - \text{ES}_\alpha(\lambda X) = \lambda G - \text{ES}_\alpha(X)$ for $\lambda \geq 0$

Proof. The three properties are direct consequences of definition (18). $\square$

3.2. G-ES for a Portfolio. In this section, we show how to calculate the G-ES for a portfolio. Our results show that doing so is equivalent to computing the weighted volatility of the portfolio, which dramatically simplifies the computation.

Proposition 3. Assume that X follows a d -dimensional G-normal distribution. Then, for each $a = (a_1, \dots, a_d) \in \mathbb{R}^d$, $X^a = (a_1 X_1, \dots, a_d X_d)$,

$$G - \text{ES}(X^a) = G - \text{ES}_\alpha\left(X_{\left(\sigma_{-a^T a}^2, \sigma_{a^T a}^2\right)}\right), \quad (21)$$

where $X_{(\sigma_{-a^T a}^2, \sigma_{a^T a}^2)} \sim \mathcal{N}(0, [\sigma_{aa^T}^2, \sigma_{-a^T a}^2])$ with $\sigma_{a^T a}^2 = 2G(a^T a)$ and $\sigma_{-a^T a}^2 = -2G(-a^T a)$.

Proof. By Peng [2], for a d -dimensional G-normal distributed random variable X , for each vector $a = (a_1, \dots, a_d) \in \mathbb{R}^d$, $X^a = \langle a, X \rangle$ follows a 1-dimensional G-normal distribution. Consequently, Proposition 3 holds true.

Note that the dimension d of the portfolio can be of several hundreds or thousands because calculating the G-ES of a portfolio is simply equivalent to finding the weighted volatility bounds σ_{-aa^T} and σ_{aa^T} . Given the set of possible covariance matrices, it is not costly to determine σ_{-aa^T} and σ_{aa^T} by a certain search algorithm. Particularly, the following corollary states a result with given boundaries of covariance matrices for a portfolio. $\square$

Corollary 1. Let $\bar{C}$ and $\underline{C}$ be two $d \times d$ dimensional nonnegative definite matrices such that $\bar{C} - \underline{C}$ is also nonnegative definite. Assume that X follows a d -dimensional G-normal distribution with covariance matrix C satisfying $\underline{C} \leq C \leq \bar{C}$ (for two matrices A and B , $A \leq B$ means $B - A$ is nonnegative definite). Then, for each vector $a = (a_1, \dots, a_d) \in \mathbb{R}^d$, the G-ES of the portfolio $X^a = (a_1 X_1, \dots, a_d X_d)$ is

$$G - ES(X^a) = G - ES_\alpha \left(X_{((\underline{\sigma}^*)^2, (\bar{\sigma}^*)^2)} \right), \quad (22)$$

where $X_{((\underline{\sigma}^*)^2, (\bar{\sigma}^*)^2)} \sim \mathcal{N}(0, [(\underline{\sigma}^*)^2, (\bar{\sigma}^*)^2])$ with

$$\begin{aligned} (\bar{\sigma}^*)^2 &= \text{tr}(\bar{C} a^T a), \\ (\underline{\sigma}^*)^2 &= \text{tr}(\underline{C} a^T a), \end{aligned} \quad (23)$$

in which $\text{tr}(\cdot)$ represents the trace of the matrix.

Proof. From Proposition 3, we obtain $X^a \sim \mathcal{N}(0, [(\underline{\sigma}^*)^2, (\bar{\sigma}^*)^2])$ with $(\bar{\sigma}^*)^2 = 2G(a^T a) = \sup_{\underline{C} \leq C \leq \bar{C}} [\text{tr}(C a^T a)]$ and $(\underline{\sigma}^*)^2 = -2G(-a^T a) = \inf_{\underline{C} \leq C \leq \bar{C}} [\text{tr}(\bar{C} a^T a)]$. Then, by Guo et al. [22], we have

$$\sup_{\underline{C} \leq C \leq \bar{C}} [\text{tr}(C a^T a)] = \text{tr}(\underline{C} a^T a) + \sum_{i=1}^d \gamma_i^+, \quad (24)$$

where $\gamma_i, i = 1, \dots, d$, are the eigenvalues of $L^T a^T a L$ with $LL^T = \bar{\sigma}^2 - \underline{\sigma}^2$ being the Cholesky decomposition. It is easy

to verify that $L^T a^T a L$ is a nonnegative definite matrix, so $\sum_{i=1}^d \gamma_i^+ = \sum_{i=1}^d \gamma_i = \text{tr}(L^T a^T a L) = \text{tr}((\bar{C} - \underline{C}) a^T a)$ and $\sup_{\underline{C} \leq C \leq \bar{C}} [\text{tr}(C a^T a)] = \text{tr}(\underline{C} a^T a) + \text{tr}((\bar{C} - \underline{C}) a^T a) = \text{tr}(\bar{C} a^T a)$. As for $\sigma_{-a^T a}^2$,

$$\inf_{\underline{C} \leq C \leq \bar{C}} [\text{tr}(C a^T a)] = - \sup_{\underline{C} \leq C \leq \bar{C}} [\text{tr}(-C a^T a)] = \text{tr}(\underline{C} a^T a) - \sum_{i=1}^d \hat{\gamma}_i^+, \quad (25)$$

where $\hat{\gamma}_i, i = 1, \dots, d$, are the eigenvalues of $-L^T a^T a L$. Since $-L^T a^T a L$ is a nonpositive definite matrix, we have $\hat{\gamma}_i^+ = 0, i = 1, \dots, d$; therefore, $\sum_{i=1}^d \hat{\gamma}_i^+ = 0$. Consequently, Corollary 1 holds true. $\square$

Corollary 2. Let $\bar{C}$ be a 2×2 nonnegative definite matrix, and assume that (X_1, X_2) follows a two-dimensional G-normal distribution with covariance matrix C satisfying $k^2 \bar{C} \leq C \leq \bar{C}, 0 \leq k \leq 1$. Then, for a given $\alpha \in [(\underline{\sigma}/\bar{\sigma} + \underline{\sigma}), 1]$, $G - ES_\alpha(X_1 + X_2) \leq G - ES_\alpha(X_1) + G - ES_\alpha(X_2)$.

Corollary 2 states the subadditivity of the G-ES. Particularly, we can take $k = 0$; that is, $0 \leq C \leq \bar{C}$.

Proof of Corollary 2. Since (X_1, X_2) are 2-dimensional G-normal distributed, X_1 and X_2 are 1-dimensional G-normal distributed random variables. Without loss of generality, we assume $X_1 \sim \mathcal{N}(0, [\underline{\sigma}_1^2, \bar{\sigma}_1^2])$, $X_2 \sim \mathcal{N}(0, [\underline{\sigma}_2^2, \bar{\sigma}_2^2])$, and the matrix $\bar{C} = \begin{pmatrix} \bar{\sigma}_1^2 & \rho_0 \bar{\sigma}_1 \bar{\sigma}_2 \\ \rho_0 \bar{\sigma}_1 \bar{\sigma}_2 & \bar{\sigma}_2^2 \end{pmatrix}$. By the condition $k^2 \bar{C} \leq C \leq \bar{C}$, we obtain $\underline{\sigma}_1^2/\bar{\sigma}_1^2 = \underline{\sigma}_2^2/\bar{\sigma}_2^2 = k^2$. From Corollary 1, we obtain $G - ES_\alpha(X_1 + X_2) = G - ES_\alpha(X_{(\sigma_{-a^T a}^2, \sigma_{a^T a}^2)})$, where $X_{(\sigma_{-a^T a}^2, \sigma_{a^T a}^2)} \sim \mathcal{N}(0, [\sigma_{-a^T a}^2, \sigma_{a^T a}^2])$, $\sigma_{-a^T a}^2/\sigma_{a^T a}^2 = k^2(\bar{\sigma}_1^2 + 2\rho_0 \bar{\sigma}_1 \bar{\sigma}_2 + \bar{\sigma}_2^2)/(\bar{\sigma}_1^2 + 2\rho_0 \bar{\sigma}_1 \bar{\sigma}_2 + \bar{\sigma}_2^2) = k^2$. Observing that $-N^{-1}((1+k/2)(1-\alpha)) \geq 0$, and by the explicit formula of the G-ES, we obtain

$$\begin{aligned} G - ES_\alpha(X_1 + X_2) &= \frac{1}{1-\alpha} \sqrt{\frac{2}{\pi}} \frac{(\bar{\sigma}_1^2 + 2\rho_0 \bar{\sigma}_1 \bar{\sigma}_2 + \bar{\sigma}_2^2)^{(1/2)}}{1+k} \exp \left\{ -\frac{1}{2} \left[N^{-1} \left(\left(\frac{1+k}{2} \right) (1-\alpha) \right) \right]^2 \right\} \\ &\leq \frac{1}{1-\alpha} \sqrt{\frac{2}{\pi}} \frac{\bar{\sigma}_1 + \bar{\sigma}_2}{1+k} \exp \left\{ -\frac{1}{2} \left[N^{-1} \left(\left(\frac{1+k}{2} \right) (1-\alpha) \right) \right]^2 \right\} \\ &= G - ES_\alpha(X_1) + G - ES_\alpha(X_2). \end{aligned} \quad (26)$$

The proof is complete.

We now present a numerical example (this numerical example is worked on ThinkPad T450 (CPU 2.4 GHz, Intel Core i7-5500U, Memory 8 GB)). Given the number of assets, we randomly generate two nonnegative definite matrices $\bar{C}$ and $\underline{C}$ as the upper and lower bounds of the covariance matrix, satisfying that $\bar{C} - \underline{C}$ is a nonnegative definite.

Table 1 reports the weighted volatility bounds, the G-ES with a 97.5% confidence level, and the time cost to calculate the G-ES, given the bounds of the volatility matrices. As we see, it takes only several seconds to calculate the G-ES for a 5,000-dimensional portfolio (the time does not include that the generation of the volatility matrixes). A close look also shows that as the number of assets increases, the G-ES

TABLE 1: The time costs of calculating the G-ES for a portfolio, assuming that the proportion of money invested in each asset is the same; that is, $a = (1/d, \dots, 1/d)$.

	Number of assets				
	10	100	500	1000	5000
$\sigma_{-a^T a}$	0.0433	0.0128	0.0056	0.0039	0.0017
$\sigma_{a^T a}$	0.2492	0.1525	0.0983	0.0819	0.0548
$G - ES_{0.975}$	0.1304	0.0525	0.0219	0.0162	0.0072
CPU time (seconds)	0.0008	0.0013	0.0125	0.0792	5.5610

decreases, which is in line with the basic principle of portfolio diversification. $\square$

Remark 1. Empirically, we can also find the two boundary matrices $\underline{C}$ and $\overline{C}$. Suppose there are N covariance matrices $C^n = [c_{ij}^n]_{1 \leq i \leq d, 1 \leq j \leq d}$, $n = 1, \dots, N$. If one of these covariance matrices C^k for some k satisfies $C^k \leq C^n$ ($C^k \geq C^n$) for all $n = 1, \dots, N$, then we set $\underline{C} = C^k$ ($\overline{C} = C^k$, respectively). Otherwise, we define $\underline{C}$ as the zero matrix, and $\overline{C} = [\bar{c}_{ij}]_{1 \leq i \leq d, 1 \leq j \leq d}$, where $\bar{c}_{ii} = \max_{n=1,2,\dots,N} \{c_{ii}^n\} + \sum_{j=1, j \neq i}^d (\max_{n=1,2,\dots,N} \{c_{ij}^n\} - \min_{n=1,2,\dots,N} \{c_{ij}^n\})$ for $i = 1, 2, \dots, d$, and $\bar{c}_{ij} = \max_{n=1,2,\dots,N} \{c_{ij}^n\}$ for $i, j = 1, 2, \dots, d$, and $i \neq j$. It is easy to check that, for each n , $\overline{C} - C^n$ is a diagonally dominant symmetric real matrix, and its main diagonal elements are $\bar{c}_{ii} - c_{ii}^n \geq 0$, $i = 1, 2, \dots, d$, $n = 1, \dots, N$. By Ye [23], $\overline{C} - C^n \geq 0$ for $n = 1, \dots, N$.

Another way to obtain the bounds $\underline{C}$ and $\overline{C}$ is by the perturbation of a prediction matrix C . For instance, with C in hand, we can define $\underline{C} = k_1 C$ and $\overline{C} = k_2 C$ as the bounds, where $0 \leq k_1 \leq 1 \leq k_2$.

4. Backtest

In this section, we test the robustness of the G-ES empirically. We use daily loss data of nine global market indexes: the CSI300, SHSCI, and SZSCI from China; the S&P500, DJIA, and NASDAQ from the US; and the CAC40, FTSE100, and DAX from Europe. The sample period is from 1 January 2007 to 31 December 2018. The loss data are normalized with zero mean. We use the exponentially weighted moving average (EWMA, see Hull [24]) model to predict daily volatility.

Peng et al. [25] obtain the upper/lower volatility bounds by dividing the data into several windows and then taking the maximum volatility as $\bar{\sigma}$ and the minimum as $\mathbb{S}((d))$. They use window sizes of 250, 500, and 1000 days to make predictions. However, even for 250 days, the G-VaR does not change sensitively over time. In particular, for 1000 days, the

G-VaR graph is nearly a straight line for the four years from 2009 to 2012 for the S&P500 index.

To avoid this insensitivity, for the G-ES at the 97.5% confidence level, we multiply the daily volatility by 1.16/0.84 to obtain the upper/lower bounds of volatility. We tested several multipliers and choose these two multipliers. The empirical calibration later shows that we can obtain a reliable G-ES in this way, not only for the China indexes but also for the US and Europe indexes. However, there is no consensus on how to determine volatility ambiguity because different sample sizes or different partitions of the data may yield different volatility bounds. We must therefore find the best method. We first adjust the multipliers for the CSI300 and then apply the best results (1.16/0.84) to the other indexes. We find that the G-ES performs robustly and reliably most of the time. Furthermore, for different confidence levels, we should adopt different levels of volatility ambiguity, that is, different multipliers. Due to limited space, we show only several ESs at the 97.5% confidence level based on the CSI300, S&P500, and CAC40 for the years 2007 to 2012. Figure 3 shows that the G-ES performs similarly to the Gumbel-ES, (the formula for the Gumbel-ES is $(1/(1-\alpha)) \int_{\alpha}^1 (\theta - \beta \ln(-\ln t)) dt$, where θ and β are the location and scale parameters, respectively, and α is the confidence level) while the G-VaR is close to the normal-ES with the same confidence level. A close look indicates that the G-ES is obviously more robust than the G-VaR. Interestingly, Figures 4 and 5 also show the same pattern.

To confirm the reliability of the G-ES, we first apply a nonparametric method of Acerbi and Szekely [26] to backtest the ES models. Then we conduct a comparative backtest following Nolde and Ziegel [27].

Let L_i represent the percentage loss in day i ($i = 1, \dots, T$). These losses are distributed according to a real but unknown distribution F_i and forecasted by a predictive distribution P_i . We assume that the random variables $\vec{L} = \{L_i\}$ are independent of each other. Let $I_i = 1_{\{L_i - \text{VaR}_{\alpha,i} > 0\}}$ be the indicator function of a VaR violation.

We now describe this test, derived from the representation of the ES as an unconditional expectation (as for the G-ES, E corresponds to the worst-case distribution $\underline{F}$)

$$ES_{\alpha,i} = E \left[\frac{L_i \cdot I_i}{1 - \alpha} \right], \quad (27)$$

which suggests defining the following test statistic:

$$Z(\vec{L}) = \frac{\sum_{i=1}^T (L_i \cdot I_i / ES_{\alpha,i})}{T(1 - \alpha)} - 1. \quad (28)$$

The null and the alternative hypothesis will be

$$H_0: P_i^{[1-\alpha]} = F_i^{[1-\alpha]}, \quad (29)$$

$$\forall i, H_1: ES_{\alpha,i}^{F_i} \geq ES_{\alpha,i}, \forall i \text{ and } ">" \text{ holds for some } i, \text{VaR}_{\alpha,i}^{F_i} \geq \text{VaR}_{\alpha,i}, \forall i,$$

FIGURE 3: Comparison of several ESs and G-VaR for the CSI300 at the $\alpha = 97.5\%$ confidence level.

FIGURE 4: Comparison of several ESs and G-VaR for the S&P500 at the $\alpha = 97.5\%$ confidence level.

FIGURE 5: Comparison of several ESs and G-VaR for CAC40 at the $\alpha = 97.5\%$ confidence level.

respectively, where $P_i^{[1-\alpha]} = \min\{1, (1 - P_i/1 - \alpha)\}$ and $F_i^{[1-\alpha]} = \min\{1, (1 - F_i/1 - \alpha)\}$ denote the tail distributions of L_i when $L_i > \text{VaR}_{\alpha,i}$, and $ES_{\alpha,i}^{F_i}$ and $\text{VaR}_{\alpha,i}^{F_i}$ denote the values of the risk measures when $L_i \sim F_i$. Let E_{H_0} and E_{H_1} be the conditional expectations under hypothesis H_0 and H_1 , respectively. Similar to Acerbi and Szekely [26], it is easy to obtain $E_{H_0}[Z] = 0$ and $E_{H_1}[Z] > 0$. To calculate the p value $p = P_Z(Z(\vec{l}))$ of a realization $Z(\vec{l})$, we need to simulate the distribution P_Z under H_0 . The first step is to simulate $L_i^m \sim P_i, \forall i, \forall m = 1, \dots, M$ independently. Second, we must calculate $Z^m := Z(\vec{L}^m)$ for each m . Therefore, we estimate the p value by $p = \sum_{m=1}^M 1_{\{Z^m > Z(\vec{l})\}}/M$, where M is a suitably large number of scenarios. Given a significance level ϕ , the test is not rejected if $p > \phi$. When $p \leq \phi$, then we conclude that the model underestimates the shortfall risk, and hence, the model does not pass the test. According to Acerbi and Szekely [26], if $p > 1 - \phi$, then the model overestimates the shortfall risk.

Table 2 shows the results of the test for each model. A large p value means a large value of the corresponding ES. “ $\times$ ” means that the p value is less than 0.05 or greater than 0.95; that is, the corresponding model does not pass the test, while “ $\checkmark$ ” implies that the p value belongs to the interval (0.05, 0.95]; that is, the corresponding model passes the test. As we can see, the G-ES passes all the tests for the nine international indexes, while the Gumbel-ES does not pass the tests for indexes SZSCI, CAC40, and DAX because it overestimates the shortfall risks. The normal-ES does not pass the test in any case because it underestimates the shortfall risks.

Remark 2. We also apply the backtesting methodology of McNeil and Frey [28] to the three ES models for a 12-year sample period of the nine worldwide indexes. Given a significance level $\phi = 5\%$, normal-ES does not pass the test for any of the nine international indexes. The G-ES passes the test only for the DAX index, while the Gumbel-ES passes most tests for the international indexes except for the S&P500 index.

However, as shown in Figures 3–5 and Table 2, the G-ES is closed to and outperforms the Gumbel-ES. Moreover, Acerbi and Szekely [26] indicate the test method in McNeil and Frey [28] alone or its variation is not valid to backtest models for ES because the accuracy of VaR forecast affects the outcomes of this test significantly. This is also supported by findings in Roccioletti (2016, Chap. 5, Sect. 5.3) [29].

Next, we conduct the comparative backtests introduced by Nolde and Ziegel [27]. We consider the pair $\Theta = (\text{VaR}_\alpha, ES_\alpha)$. Let $\mathbb{N}$ be the set of natural numbers and $\{R_i\}_{i \in \mathbb{N}}$ and $\{R_i^*\}_{i \in \mathbb{N}}$ be two sequences of predictions of Θ , which are referred to as the internal model and the standard model, respectively. Let S be a consistent scoring function for Θ . Then, we say $\{R_i\}_{i \in \mathbb{N}}$ S -dominates $\{R_i^*\}_{i \in \mathbb{N}}$ if $E[S(R_i, L_i) - S(R_i^*, L_i)] \leq 0$, for all $i \in \mathbb{N}$.

Now we conduct the comparative backtests proposed in Nolde and Ziegel [27]. For a consistent scoring function S and a set with n data, the sample average is defined as follows:

TABLE 2: The p values for a 12-year sample period of nine worldwide indexes for three ES models at the significance level $\phi = 5\%$.

	Normal-ES	G-ES	Gumbel-ES
CSI300	0.0000 (×)	0.3908 (✓)	0.9400 (✓)
SHSCI	0.0000 (×)	0.2280 (✓)	0.7396 (✓)
SZSCI	0.0000 (×)	0.5204 (✓)	0.9517 (×)
S&P500	0.0000 (×)	0.0666 (✓)	0.5298 (✓)
DJIA	0.0000 (×)	0.1081 (✓)	0.6617 (✓)
NASDAQ	0.0000 (×)	0.1684 (✓)	0.7588 (✓)
CAC40	0.0000 (×)	0.9338 (✓)	0.9997 (×)
FTSE100	0.0000 (×)	0.4543 (✓)	0.9484 (✓)
DAX	0.0000 (×)	0.8887 (✓)	0.9994 (×)

Note: “×” means that the corresponding ES model does not pass the test, while “✓” implies that the corresponding ES model passes the test.

$$\bar{S} := \frac{1}{n} \sum_{i=1}^n S(R_i, L_i). \quad (30)$$

Given a confidence level $\alpha = 97.5\%$, Table 3 reports the sample average of consistent scoring functions S_1 and S_2 , respectively, along with the corresponding model rankings, where S_1 and S_2 are defined as

$$S_1((r_1, r_2), x) = 1\{x > r_1\} \frac{x - r_1}{2\sqrt{r_2}} + (1 - \alpha) \frac{r_1 + r_2}{2\sqrt{r_2}},$$

$$S_2((r_1, r_2), x) = 1\{x > r_1\} \frac{x - r_1}{r_2} + (1 - \alpha) \left(\frac{r_1}{r_2} - 1 + \log(r_2) \right). \quad (31)$$

Combining the outcomes of Part A and Part B of Table 3, we see that, in most cases rankings obtained from both consistent scoring functions coincide with each other except for indexes SZSCI, DJIA, and NASDAQ. In general, for the Chinese and the American stock indexes, the scoring functions rank the G-ES as the best or second best performing model. For the European stock indexes, the two scoring functions result in a good agreement with G-ES being the best forecaster. For all the nine indexes, the normal-ES shows the worst performance according to both the two scoring functions.

As a complement to model rankings in Table 3, now we compare the internal model with a given standard model using the test method proposed in Fissler et al. [30] and Nolde and Ziegel [27]. Define

$$\lambda^* := \limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E(S(R_i, L_i) - S(R_i^*, L_i)),$$

$$\lambda_* := \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n E(S(R_i, L_i) - S(R_i^*, L_i)). \quad (32)$$

Then the comparative backtesting hypotheses can be formulated as

$$H_0^-: \lambda^* \leq 0, \quad H_0^+: \lambda_* \geq 0. \quad (33)$$

TABLE 3: Comparing the predictive performance of different forecasting models for a 12-year sample period of nine worldwide indexes based on the average scores $\bar{S}_1$ and $\bar{S}_2$, scaled by $1 - \alpha$ for presentation purpose, respectively, along with the corresponding model ranks (in parentheses).

Part A			
	Normal-ES	G-ES	Gumbel-ES
CSI300	0.2192 (3)	0.2148 (1)	0.2150 (2)
SHSCI	0.2148 (3)	0.2100* (2)	0.2100** (1)
SZSCI	0.2313 (3)	0.2271 (1)	0.2276 (2)
S&P500	0.1775 (3)	0.1731* (2)	0.1731** (1)
DJIA	0.1696 (3)	0.1663 (1)	0.1665 (2)
NASDAQ	0.1868 (3)	0.1826 (1)	0.1829 (2)
CAC40	0.1866 (3)	0.1849 (1)	0.1860 (2)
FTSE100	0.1701 (3)	0.1682 (1)	0.1683 (2)
DAX	0.1849 (3)	0.1829 (1)	0.1830 (2)

Part B			
CSI300	-3.0563 (3)	-3.1223 (1)	-3.1221 (2)
SHSCI	-3.0845 (3)	-3.1617 (2)	-3.1676 (1)
SZSCI	-2.9043 (3)	-2.9747 (2)	-2.9768 (1)
S&P500	-3.4161 (3)	-3.5173 (2)	-3.5290 (1)
DJIA	-3.5224 (3)	-3.6046 (2)	-3.6122 (1)
NASDAQ	-3.3176 (3)	-3.4112 (2)	-3.4166 (1)
CAC40	-3.3588 (3)	-3.4034 (1)	-3.3933 (2)
FTSE100	-3.5474 (3)	-3.5995 (1)	-3.5967 (2)
DAX	-3.3801 (3)	-3.4264 (1)	-3.4226 (2)

Note. Before rounding off, *the number is 0.21004, **the number is 0.21002, *the number is 0.17310, and **the number is 0.17307.

We define

$$\Delta_n \bar{S} := \frac{1}{n} \sum_{i=1}^n (S(R_i, L_i) - S(R_i^*, L_i)). \quad (34)$$

Given a significance level $\phi = 5\%$, we use the test statistic

$$T_1 = \frac{\Delta_n \bar{S}}{\hat{\sigma}_n / \sqrt{n}}, \quad (35)$$

where $\hat{\sigma}_n^2$ represents a heteroscedasticity and autocorrelation consistent estimator of the asymptotic variance, $\sigma_n^2 = \text{var}(\sqrt{n} \Delta_n \bar{S})$, and then we obtain an asymptotic level- ϕ test of H_0^+ if we reject the null hypothesis when $N(T_1) \leq \phi$, and of H_0^- if we reject the null hypothesis when $1 - N(T_1) \leq \phi$. Based on the outcome of the tests of H_0^+ and H_0^- , we say that the internal model fails the comparative backtest if H_0^- is rejected (i.e., it is in the red region of Figure 6). The internal model passes the backtest if H_0^+ is rejected (i.e., it is in the green region of Figure 6). The internal model needs further investigation if neither H_0^+ nor H_0^- can be rejected (i.e., it is in the yellow region of Figure 6).

Figure 6 displays the traffic light matrices for three ES models from 1 January 2007 to 31 December 2018. Due to limited space, we only show outcomes for the indexes of CSI300, S&P500, and CAC40. Along the vertical axis, we consider hypothetical “standard” models with the investigated “internal” models displayed along the horizontal axis. The red (green) cells correspond to situations in which the comparative backtest is failed (passed), while yellow cells indicate cases where no conclusive evidence is available to

FIGURE 6: Traffic light matrices for $(\text{VaR}_\alpha, \text{ES}_\alpha)$ forecasts at the significance level $\phi = 5\%$ for CSI300 (top row), S&P500 (middle row), and CAC40 (bottom row), respectively. The left and right parts are based on the scoring functions S_1 and S_2 , respectively.

pass or fail the comparative backtest. The left part of Figure 6 corresponds to the scoring function S_1 , while the right part corresponds to the scoring function S_2 . For the CSI300 index and the S&P500 index, the two scoring functions result in a good agreement with normal-ES being the worst forecaster (i.e., failing the comparative backtests against all the other models) and the two scoring functions cannot identify differences of performance between the G-ES and the Gumbel-ES at the given significance level. As for the CAC40 index, we show that the G-ES outperforms Gumbel-ES under both scoring functions. In particular, the scoring function S_2 is better at identifying models than the scoring function S_1 .

5. Conclusion

We have presented a simple method, the G-ES, to measure shortfall risk by incorporating volatility uncertainty. We have extended G-ES to compute risks for a portfolio with low cost, which either has closed-form formulas or can be implemented by simple numerical computation. The dimension of the portfolio can be of several hundreds. The empirical tests show that G-ES performs well based on the test statistic Z introduced by Acerbi and Szekely [26] and on the comparative backtests following Nolde and Ziegel [27]. Compared with extant ES models, the G-ES is robust and reliable.

Appendix

A. Basic Knowledge about G-Expectation

In this section, we recall some basic knowledge about Peng's G-stochastic calculus. Readers are referred to [2] for more information.

We denote by $\mathbb{S}(d)$ the collection of $d \times d$ symmetric matrices and $\mathbb{S}_+(d)$ the positive-semidefinite elements of $\mathbb{S}(d)$. Let $\Omega = C_0([0, +\infty), \mathbb{R})$ denote the space of all $\mathbb{R}$ -valued continuous paths $(\omega_t)_{t \in \mathbb{R}^+}$ with $\omega_0 = 0$. $\mathcal{B}(\Omega)$ denotes the Borel σ -algebra of Ω . Let $\mathcal{H}$ be a linear space of real functions defined on Ω such that if $X_1, \dots, X_d \in \mathcal{H}$, then $\varphi(X_1, \dots, X_d) \in \mathcal{H}$ for each $\varphi \in C_{l,\text{Lip}}(\mathbb{R}^d)$, where $C_{l,\text{Lip}}(\mathbb{R}^d)$ denotes the linear space of (local Lipschitz) functions φ satisfying $|\varphi(x) - \varphi(y)| \leq C(1 + |x|^m + |y|^m)|x - y|$, $\forall x, y \in \mathbb{R}^d$, for some $C > 0$, $m \in \mathbb{N}$ depending on φ . $\mathcal{H}$ is considered as a space of "random variables". In this case, $X = (X_1, \dots, X_d)$ is called an d -dimensional random vector, denoted by $X \in \mathcal{H}^d$.

Definition A.1. A sublinear expectation $\mathbb{E}$ on $\mathcal{H}$ is a functional $\mathbb{E}: \mathcal{H} \rightarrow \mathbb{R}$ satisfying the following properties: for all $X, Y \in \mathcal{H}$, we have

- (a) Monotonicity: if $X \geq Y$, then $\mathbb{E}[X] \geq \mathbb{E}[Y]$
- (b) Constant preserving: $\mathbb{E}[c] = c, \forall c \in \mathbb{R}$
- (c) Subadditivity: $\mathbb{E}[X + Y] \leq \mathbb{E}[X] + \mathbb{E}[Y]$
- (d) Positive homogeneity: $\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X], \forall \lambda \geq 0$

Definition A.2. Let X_1 and X_2 be two d -dimensional random vectors defined on the sublinear expectation spaces $[\Omega, \mathcal{H}, \mathbb{E}]$. They are called identically distributed, denoted by $X_1 \stackrel{d}{=} X_2$, if

$$\mathbb{E}[\varphi(X_1)] = \mathbb{E}[\varphi(X_2)], \quad \forall \varphi \in C_{l,\text{Lip}}(\mathbb{R}^d). \quad (\text{A.1})$$

Definition A.3. In a sublinear expectation space $(\Omega, \mathcal{H}, \mathbb{E})$, a random vector $Y \in \mathcal{H}^d$ is said to be independent of another random vector $X \in \mathcal{H}^d$ under $\mathbb{E}$ if for each test function $\varphi \in C_{l,\text{Lip}}(\mathbb{R}^{2d})$ we have

$$\mathbb{E}[\varphi(X, Y)] = \mathbb{E}[\mathbb{E}[\varphi(x, Y)]_{x=X}]. \quad (\text{A.2})$$

Definition A.4. (G-normal distribution). A d -dimensional random vector $X = (X_1, \dots, X_d)$ in a sublinear expectation space $(\Omega, \mathcal{H}, \mathbb{E})$ is called G-normal distributed if for each $a, b > 0$ we have

$$aX + b\bar{X} \stackrel{d}{=} \sqrt{a^2 + b^2} X, \quad (\text{A.3})$$

where $\bar{X}$ is an independent copy of X .

Remark A.1. It is easy to check that $\mathbb{E}[X] = \mathbb{E}[-X] = 0$. The so-called "G" is related to $G: \mathbb{S}(d) \rightarrow \mathbb{R}$ defined by

$$G(A) = \frac{1}{2} \mathbb{E}[\langle AX, X \rangle]. \quad (\text{A.4})$$

Assume X follows a normal distribution. For each $(t, x) \in [0, \infty) \times \mathbb{R}^d$, define $u(t, x) := \mathbb{E}[\varphi(x + \sqrt{t} X)]$, where φ is bounded and Lipschitz-continuous. Peng [2] shows that X is G-normal distributed if u is the viscosity solution of the following HJB equation:

$$\begin{cases} \partial_t u - G(D^2 u) = 0, \\ (t, x) \in (0, \infty) \times \mathbb{R}^d, \\ u(0, x) = \varphi(x), \end{cases} \quad (\text{A.5})$$

where $D^2 u$ is the Hessian matrix of u . Similar to the classical theory of stochastic calculus, G-Brownian motion and related stochastic integral are also defined by Peng [2] and extended by Li and Peng [31] and Song [32].

Data Availability

The data in this paper can be used publicly.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Authors' Contributions

Xingye Yue contributed to the conception of the study, and Xuhui Wang helped perform the analysis with constructive discussions. They both contributed significantly to analysis and manuscript preparation. Ziting Pei performed the data analyses and wrote the manuscript.

Acknowledgments

The authors are grateful to professors Zaichao Du and Sishui Wang for useful discussions. This work was supported by the Natural Science Foundation of China (nos. 11971342 and 12071335).

References

- [1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath, "Coherent measures of risk," *Mathematical Finance*, vol. 9, no. 3, pp. 203–228, 1999.
- [2] S. Peng, "Nonlinear expectations and stochastic calculus under uncertainty with robust CLT and G-Brownian motion," *Probability Theory and Stochastic Modelling*, Springer, Berlin, Germany, 2019.
- [3] Z. Du and J. C. Escanciano, "Backtesting expected shortfall: counting for tail risk," *Management Science*, vol. 63, no. 4, pp. 940–958, a, 2016.
- [4] P. F. Christoffersen, "Evaluating interval forecasts," *International Economic Review*, vol. 39, no. 4, pp. 841–862, 1998.
- [5] J. Berkowitz, P. Christoffersen, and D. Pelletier, "Evaluating value-at-risk models with desk-level data," *Management Science*, vol. 57, no. 12, pp. 2213–2227, 2011.
- [6] P. H. Kupiec, "Techniques for verifying the accuracy of risk measurement models," *The Journal of Derivatives*, vol. 3, no. 2, pp. 73–84, 1995.

- [7] K. Natarajan, D. Pachamanova, and M. Sim, "Constructing risk measures from uncertainty sets," *Operations Research*, vol. 57, no. 5, pp. 1129–1141, 2009.
- [8] L. E. Ghaoui, M. Oks, and F. Oustry, "Worst-case value-at-risk and robust portfolio optimization: a conic programming approach," *Operations Research*, vol. 51, no. 4, pp. 543–556, 2003.
- [9] S. Zhu and M. Fukushima, "Worst-case conditional value-at-risk with application to robust portfolio management," *Operations Research*, vol. 57, no. 5, pp. 1155–1168, 2009.
- [10] Z. Pei, X. Wang, Y. Xu, and X. Yue, "A worst-case risk measure by G-VaR," *Forthcoming in Acta Mathematicae Applicatae Sinica, English Series*, 1–25, Springer, Berlin, Germany, 2020.
- [11] L. Chen, S. He, and S. Zhang, "Tight bounds for some risk measures, with applications to robust portfolio selection," *Operations Research*, vol. 59, no. 4, pp. 847–865, 2011.
- [12] K. Natarajan, M. Sim, and J. Uichanco, "Tractable robust expected utility and risk models for portfolio optimization," *Mathematical Finance*, vol. 20, no. 4, pp. 695–731, 2010.
- [13] S. Peng, "G-expectation, G-Brownian motion and related stochastic calculus of Ito's type," in *Stochastic Analysis and Applications, the Abel Symposium 2005, Abel Symposia 2*, pp. 541–567, Springer-Verlag, Berlin, Germany, 2006.
- [14] S. Peng, "G-Brownian motion and dynamic risk measure under volatility uncertainty," 2007, <http://arxiv.org/abs/0711.2834.v1>.
- [15] S. Peng, "Multi-dimensional G-Brownian motion and related stochastic calculus under G-expectation," *Stochastic Processes and Their Applications*, vol. 118, no. 12, pp. 2223–2253, 2008.
- [16] S. Peng, "Survey on normal distributions, central limit theorem, Brownian motion and the related stochastic calculus under sublinear expectations," *Science in China Series A: Mathematics*, vol. 52, no. 7, pp. 1391–1411, 2009.
- [17] M.-s. Hu and S.-g. Peng, "On representation theorem of G-expectations and paths of G-Brownian motion," *Acta Mathematicae Applicatae Sinica, English Series*, vol. 25, no. 3, pp. 539–546, 2009.
- [18] H. M. Soner, N. Touzi, and J. Zhang, "Martingale representation theorem for the G-expectation," *Stochastic Processes and Their Applications*, vol. 121, no. 2, pp. 265–287, 2011.
- [19] G. W. Bluman and J. D. Cole, *Similarity Methods for Differential Equations*, Springer Science & Business Media, Berlin, Germany, 2012.
- [20] R. T. Rockafellar and S. Uryasev, "Optimization of conditional value-at-risk," *The Journal of Risk*, vol. 2, no. 3, pp. 21–41, 2000.
- [21] R. T. Rockafellar and S. Uryasev, "Conditional value-at-risk for general loss distributions," *Journal of Banking & Finance*, vol. 26, no. 7, pp. 1443–1471, 2002.
- [22] W. Guo, J. Zhang, and J. Zhuo, "A monotone scheme for high-dimensional fully nonlinear PDEs," *The Annals of Applied Probability*, vol. 25, no. 3, pp. 1540–1580, 2015.
- [23] Q. Ye, "Computing singular values of diagonally dominant matrices to high relative accuracy," *Mathematics of Computation*, vol. 77, no. 264, pp. 2195–2230, 2008.
- [24] J. Hull, *Risk Management and Financial Institutions*, + Web Site, John Wiley & Sons, Hoboken, NJ, USA, 2012.
- [25] S. Peng, S. Yang, and J. Yao, "Improving value-at-risk prediction under model uncertainty," *Forthcoming in Journal of Financial Econometrics*, pp. 1–44, 2020.
- [26] C. Acerbi and B. Szekely, "Back-testing expected shortfall," *Risk*, vol. 27, no. 11, pp. 76–81, 2014.
- [27] N. Nolde and J. F. Ziegel, "Elicitability and backtesting: perspectives for banking regulation," *Annals of Applied Statistics*, vol. 11, no. 4, pp. 1833–1874, 2017.
- [28] A. J. McNeil and R. Frey, "Estimation of tailed-related risk measures for heteroscedastic financial time series: an extreme value approach," *Journal of Empirical Finance*, vol. 7, no. 3–4, pp. 271–300, 2000.
- [29] S. Roccioletti, *Backtesting Value at Risk and Expected Shortfall*, Springer, Berlin, Germany, 2015.
- [30] T. Fissler, J. F. Ziegel, and T. Gneiting, *Expected Shortfall Is Jointly Elicitable with Value at Risk implications for Backtesting*, pp. 58–C61, Risk Magazine, London, UK, 2016.
- [31] X. Li and S. Peng, "Stopping times and related Ito's calculus with G-Brownian motion," *Stochastic Processes and Their Applications*, vol. 121, no. 7, pp. 1492–1508, 2011.
- [32] Y. Song, "Some properties on G-evaluation and its applications to G-martingale decomposition," *Science China Mathematics*, vol. 54, no. 2, pp. 287–300, 2011.