Research Article

Estimation of Tail Risk and Moments Using Option Prices with a Novel Pricing Model under a Distorted Lognormal Distribution

Yan Chen, Ya Cai, and Chengli Zheng

1School of Mathematics and Statistics, Central China Normal University, Wuhan, China
2School of Economics and Business Administration, Financial Engineering Research Center, Central China Normal University, Wuhan, China

Correspondence should be addressed to Chengli Zheng; zhengchengli168@163.com

Received 29 May 2020; Accepted 15 June 2020; Published 20 July 2020

Guest Editor: Wenguang Yu

Copyright © 2020 Yan Chen 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.

Risk measures based on the trading option prices in the market are forward-looking, such as VIX. We propose a new method combining distorted lognormal distribution with interpolation to price options accurately and then estimate tail risk. Our method can price the option of any strikes between the maximum and the minimum value of strikes in the real market, which reduces the instability and inaccuracy of using the limited option to measure the risk. In addition, our novel method treats the underlying asset price as a stochastic indicator rather than a fixed indicator as described in previous research studies for risk measurement. Moreover, even if the available sample size is very small, we can measure the risk stably and precisely after interpolation. Finally, the empirical test results of SP500 market show that this method has good performance, especially for the option markets with sparse strikes.

1. Introduction

Risk measure is always being regarded as one of the most important parts for risk management. To quantify the risk well, the future distribution of the risk must be characterized, which involves forecasting and is difficult. The traditional method is based on the assumption that the history has a habit of repeating itself. So, the historical time series are used to estimate the risk measures. This method is backward-looking, which will never contain information about future.

To solve this problem, people turn to options market. Derivative markets are viewed as forward-looking markets because the prices of the derivative instruments traded in those markets incorporate the market participants’ perception of future market changes, which is extra information compared to those from stock markets. As a publicly traded market, the options market including stock options and index options can bring all the opinions about the future information of investors together. So, the options market can be exploited to compute the risk measures of their underlying assets. The standard approach is to determine the volatility of a stock or a stock index that is implied by today’s market prices of traded options.

The VIX volatility index is one of the application methods of option price. It is disseminated by the Chicago Board Options Exchange (CBOE) following the results of Whaley [1] and has attracted much attention in recent years. The index is computed from current price on a wide range of out-of-the-money European style call and put options written on the SP500 equity index, and it is built to serve as a model-free option-implied return volatility measure for the SP500 index over the coming 30 days, expressed in annualized percentage terms. VIX captures attention from the real market, and it is even labeled “fear gauge” and routinely cited by the media when describing current market conditions or “investor sentiment.” Besides the popularity in practice, VIX is increasingly used in financial economics, especially it provides a model-free measurement of the expected value of the SP500 return variation under the risk-neutral (Q probability) measure. Hence, the VIX embodies a model-free market volatility forecast. The literature about VIX can be classified in three main directions. One direction is about the forecast power of variation in the VIX for future realized return volatility. For example, Jiang and Tian [2] thought that the
information content of the VIX volatility forecast is superior to alternative implied volatility measures as well as forecasts based on historical volatility. The second direction emphasizes the fact that the VIX measure is constructed directly from observed option prices so it is bound to incorporate any pricing of variance risk that may be embedded in the market prices. Thus, VIX does not represent a pure estimator of future return volatility for the underlying asset because it includes compensation for variance risk as well. Instead, the wedge between regular time series forecasts for volatility, developed under the actual or objective (P probability) measure, and the VIX, representing the risk-neutral return volatility forecast, may be interpreted as a market volatility risk premium (see, e.g., Bondarenko [3]; Bollerslev et al. [4]; and Carr and Wu [5] and their extensions Todorov [6] and Todorov and Bollerslev [7]). The third one is about seeking to draw inference regarding the stochastic properties of the underlying market return volatility process directly from the high-frequency behavior of the VIX (see Todorov and Tauchen [8] and Jiang and Tian [9]). Andersen et al. [10] considered a novel Corridor Implied Volatility index (CX) based on a range of strikes with high-frequency data including jumps and asymmetries. For more information about model-free volatility indices, see Gonzalez-Perez [11].

Besides research on VIX, there are some others using the ideas behind this kind of model-free estimation from options. Linders et al., [12] proposed a novel group of herd behavior indices (HIX), which are model-free and risk-neutral, derived from available option data from the market. Their numerical illustration showed that the HIX based on the Dow Jones is identical to the CIX from Dhaene et al., [13], which further demonstrated that model-free estimation is an appropriate approach to calculate the HIX. Related empirical research studies are limited and need to be further explored.

This idea of model-free estimation based on options data is very important and contributes to reduce model error dramatically. However, their applications are based on the options market, where the number of options traded are limited. Usually, the strikes of options are located mainly close to the strike of at-the-money option, and the options with strikes far from that are less, even sparse. However, when this model-free method is applied, the options used mainly are those out-of-the-money options with strikes far from that of at-the-money option. Moreover, there are not enough options traded in some markets, such as China; the size of samples of data is very small. Both of the above cases will lead to non-negligible computation errors of risk measurement.

The second problem of the current method to compute risk measures is attributable to range estimation of risk measures. Andersen et al., [14] thought that some extreme options value with strikes on the tails might be noises and will make the estimation of risk to be not correct and not stable. To make it robust, the corridor method is developed, discarding the data of tails and only considering a range of strikes. Dhaene et al., [13] have the similar idea of range when building their downside heard behavior indices (DHIX). In fact, because the trading options cannot cover all ranges of strikes from zero to infinity, current risk measures based on market data are all within the strike range from the minimum to the maximum in the real market. The ranges of those measures are consistent with the fixed range of strikes, including Andersen et al. [14] and Linders et al. [12]. But the real range of risk measures must be built on the range of underlying assets, which is a stochastic process, so the range must be random, too. Apparently, it is not suitable to describe risk features with a fixed range and there are differences between the two ranges of fixed and random. Therefore, the current method (proposed by VIX, Andersen et al. [14] and Linders et al. [12]) cannot deal with this problem well because they only used the option prices traded in the market and measure risk on a fixed range of strikes.

To make some improvements on above problems, we propose a new method based on distorted lognormal distribution and interpolation in this paper. Firstly, according to the characteristics of market price of options, we introduce a distorted variable to traditional lognormal distribution of underlying asset. This distorted variable is the function of the strikes, and it makes the options price of market to fit very well. Through market data, the distorted variables of options traded in the market can be estimated, and then using interpolation, all the distorted variables of arbitrary strikes between the minimum strike to maximum strike in the real market can be obtained, and then the prices of all the options (with or without options traded in the real market) can be computed well. This method can enlarge the size of samples of data from limited traded options, which will make the computation of risk measures more precise and robust. To test validation of our method, we decrease the size of samples gradually. And we present the whole equations for risk measures under the stochastic range using this method; it shows that our new method can solve the second problem very well. The empirical results show that the differences between the two ranges of fixed and random can be very big.

Besides the traditional risk measure such as volatility which is computed on the whole range from zero to infinity and their corridor versions, we also consider the computations of tail risk, including left tail and right tail, and their range versions, discarding the extreme points of the tails. The empirical results from SP500 market show that our novel method of distorted lognormal distribution combined with interpolation is very good, especially for the option markets with sparse strikes.

The rest of the paper is arranged as follows. In Section 2, the idea of distorted lognormal distribution will be introduced, and the empirical estimations of distorted variables and their interpolations will be presented. In Section 3, we will use this method to compute moments and tail risk estimations, compared with the current traditional method; especially, the stochastic ranges are treated for three different kinds of range moments and tail risk. Section 4 concludes the paper.
2. Distorted Lognormal Distributions

Precise option price is the basis of computing risk measurement with the model-free estimation. But many empirical results (see, e.g., MacBeth and Merville [15]; Gultekin et al. [16]; and Long and Officer [17]) show that the traditional B-S-M model mispriced the options, sometimes undervalued options and sometimes overvalued options out of the money or in the money. So, we think this is one kind of investing psychology or behavior. The investors prefer to adjust subjective probability according to the moneyness. This makes the final equilibrium market prices vary with the moneyness, which may be the source of the puzzle of volatility smiles and smirks. So, we describe this adjustment of probability by a parameter \( m = m(\lambda) \), which varies with the moneyness \( \lambda = K/S_0 \). We can view that the market distorts the probability with the moneyness. The idea of distorting the normal distribution has been accepted and developed gradually to fit the "spike and fat tail" distribution better (see, e.g., Hamada and Sherris [18]; Godin et al. [19]; Labuschagne and Offwood [20]; Gerber and Shiu [21]; and Xiao-nan et al. [22]). Specifically, we assume that the distortion function is as follows:

\[
f(S_T) = \frac{e^{m(\lambda)S_T}}{E[e^{m(\lambda)|S_T}]},
\]

where \( S_T \) is the price of underlying asset at the time \( T \) and \( E \) means expectation. Then, we can use this distortion function to options pricing.

2.1. Theoretical Approach. In this section, we will explain how to define distorted variables of \( m(\lambda) \) for a call option and \( q(\lambda) \) for a put option. Given the price of a call and of a put as a function of \( K, \sigma, r, T, S_0 \), and \( m \). Now we consider the option pricing model based on the distortion function as equation (1). For simplicity, we only consider to distort the lognormal distribution here.

Under the traditional assumptions, an underlying stock price \( S \) follows the stochastic process as follows:

\[
\frac{dS}{S} = rdT + \sigma dB,
\]

where \( B \) is a Brownian motion under risk-neutral probability measure \( Q \), \( r \) is the risk-free rate, and \( \sigma \) is the constant volatility of \( S \). So, we have

\[
S_T = S_0e^{(r-\sigma^2/2)T+\sigma\sqrt{T}x},
\]

where \( x \sim N(0, 1) \) with a probability distribution as

\[
p(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.
\]

Here, we just distort the probability of \( x \) directly with weight function as

\[
f(x) = \frac{e^{m}}{E[e^{mx}]},
\]

\[
m(\lambda) \text{ is omitted as } m, \text{ briefly. So, the probability } Q \text{ is transformed to } Q^*, \text{ and } p(x) \text{ is distorted to } p^*(x) \text{ as}
\]

\[
p^*(x) = f(x)p(x) = e^{mx-(m^2/2)} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} = \frac{1}{\sqrt{2\pi}} e^{-(x-m)^2/2}.
\]

Obviously, it still follows a normal distribution. Based on this normal distribution, we can get the closed solution for European options easily.

For European call:

\[
c^*_0 = E^Q \left[ (S_T - K)1_{S_T > K} \right] e^{-rT}
\]

\[
= e^{-rT} \int_{S_T > K} (S_T - K)1_{S_T > K} p^*(x) dx
\]

\[
= S_0 e^{m\sqrt{T}} N(d_1^*) - Ke^{-rT} N(d_2^*),
\]

where

\[
d_1^* = \frac{\log(K/S_0) - (r - \sigma^2/2)T}{\sigma\sqrt{T}} + \sigma\sqrt{T} + m,
\]

\[
d_2^* = d_1^* - \sigma\sqrt{T}.
\]

Accordingly, when \( m = 0 \), equation (7) is transformed into the traditional B-S-M model. We can estimate \( m(\lambda) \) from the real market data of call options, and then we can price for call options by (7).

Here, it is further illustrated that this distorted method can solve the problem of volatility smiles and smirks, which is described as follows:

\[
c_0(\tilde{\sigma}, K) = \overline{c}_0,
\]

\[
\tilde{\sigma} = c_0^{-1}(\overline{c}_0, K),
\]

where \( c_0(\tilde{\sigma}, K) \) is the price based on the traditional B-S-M model, \( \overline{c}_0 \) is the price in the real market, and \( \tilde{\sigma} \) is the implied volatility. For the same underlying assets with the same \( S_0, T, r \), but different strike price \( K \), the implied volatility \( \tilde{\sigma} \) varies with strike price \( K \), which means the traditional B-S-M model does not work in the real market. Here, we show that our new model (equation (7)) can value the options well and solve the problem of volatility smiles and smirks. We suppose that our price is the same as the real market price:

\[
\overline{c}_0 = c^*_0,
\]

and guess the shape of \( m(\lambda) \) through the shape of volatility smiles and smirks, and then we compute the implied volatility as follows:

\[
\tilde{\sigma} = c_0^{-1}(\overline{c}_0, K).
\]

According to the shapes of volatility smiles and smirks reported, there are some kinds of distortion function that can be chosen, such as logistic, exponential, quadric, and their composite shapes. Here, we just show an example with quadric function as

\[
m(\lambda) = a(\lambda - 1)^21_{\lambda \leq 1} + b(\lambda - 1)^21_{\lambda \geq 1}.
\]

Equation (13) means \( m \) is zero at \( \lambda = 1 \) and is asymmetrical around \( \lambda = 1 \). To show the volatility smiles and smirks, we use an example. We suppose that
$S_0 = 100, T = 1, r = 0.05, \text{ and } \lambda = 0.25, \text{ and we select proper value for parameters } a \text{ and } b \text{ in equation (13). Figure 1 shows the example of } m(\lambda); \text{ Figures 2 and 3 show the corresponding volatility smile and smirk from equation (12).}

For European put, we denote distortion as $q(\lambda)$ because we think it is different from that of call options. At this situation, we can get the closed solution for European put options easily.

$$
p_0^* = E^0\left[ (K - S_T)1_{S_T < K} \right] e^{-rT} = e^{-rT} \int_{S_T < K} (K - S_T)1_{S_T < K} p^*(x) \, dx \tag{14}
$$

where the definitions of $d_1^*$ and $d_2^*$ are the same as equations (8) and (9).

$$
d_1^* = \frac{-\left(\log(K/S_0) - (r - \sigma^2/2)T\right)}{\sigma \sqrt{T}} + \sigma \sqrt{T} + q, \tag{15}
d_2^* = d_1^* - \sigma \sqrt{T}.
$$

Apparently, when $q = 0$, equation (14) is transformed into the traditional B-S-M model.

Similarly, we can estimate $q(\lambda)$ from the real market data of put options, and then we can price the put options by equation (14). And we show volatility smiles and smirks with put option similarly. We still use quadric function as

$$
q(\lambda) = a(\lambda - 1)^2 1_{\lambda > 1} + b(\lambda - 1)^2 1_{\lambda < 1}. \tag{16}
$$

Equation (16) means $q$ is zero at $\lambda = 1$ and is asymmetrical around $\lambda = 1$. To show the volatility smiles and smirks, we still use the example with $S_0 = 100, T = 1, r = 0.05, \text{ and } \lambda = 0.25, \text{ and we select proper value for parameter } a \text{ and } b \text{ in equation (16). Figure 4 shows the example of } q(\lambda); \text{ Figures 5 and 6 show the corresponding volatility smile and smirk from equation (12). To show a little difference, here we show left smirk in Figure 6, while in Figure 3, we show a right smirk.}

So, we show that our distorted method can explain the puzzle of volatility smiles and smirks well with simulation examples theoretically. To show that this method can be good in practice, we will use the real market data to fit in next section.

2.2. Practice. We now consider a date with options available on the SP500 index and all the maturity shorter than 3 years.

We collect all data about options whose underlying asset is SP500 index. The whole data are collected from the Reuters database. The data period starts from the date of Feb 7, 2017, so the Current date is Feb 7, 2017. We have 2666 types of prices of European options, half of which are calls and the others are puts. The maturity of options listed in Table 1 has 11 types and each maturity corresponds to a different risk-free rate $r$ (%). On that day, the close price of SP500 index is 2293, which means $S_0 = 2293$, and the volatility $\sigma = 0.125$.

The scatter graph of options market price with different moneyness is shown in Figure 7. The basic descriptive statistics are listed in Table 2. From Table 2, we can see that the moneyness $\lambda$ in the real market changes around in the interval of $[0.035, 1.55]$. Its mean and median values are lower than 1 and the kurtosis can be nearly approximate to that of normal distribution. Both the market prices of calls and puts are right-skewness distributions, which means the market is more likely to be optimistic, especially for the puts market due to its significantly high kurtosis. Surely, the basic
descriptive statistics show a little difference among different maturities; here we omit them. In Table 2, \( cP_0 \) means a mixture of all the put and call options. 

Now, we represent the theoretical \( m(\lambda) \) (for calls) and \( q(\lambda) \) (for puts); for each strike, we solve for the exact value of \( \lambda \) so that the market price of the option matches the model price. We assume that our option pricing model based on distorted lognormal distributions is reasonable and equals the real market price:

\[
c_0^*(m(\lambda)) = c_0,
\]

and then by using equation (17), we can estimate \( m(\lambda) \) for calls as

\[
m(\lambda) = c_0^{*\lambda}(c_0).
\]

Similarly, we can estimate \( q(\lambda) \) for puts as

\[
q(\lambda) = p_0^{*\lambda}(p_0).
\]

We call \( m(\lambda) \) estimated from real market implied \( m \) and \( q(\lambda) \) implied \( q \). Then, we get all the implied \( m \) and \( q \) at all the 11 types of maturities \( T \). Figures 8 and 9 show the scatter graphs for all the implied \( m \) and \( q \) with different moneyness at all 11 types of maturities \( T \); the basic descriptive statistics are listed in Table 3. From Figures 8 and 9, it can be seen that the implied \( m \) and \( q \) at different maturities \( T \) varying with the moneyness \( \lambda \) have the similar trend, which means that the relationship between these two variables can be described by the same kind of function. We can find that the shapes of \( q \) are almost the same as we guessed in Figure 4; we can find a function similar to equation (13) to fit them well. But the shapes of \( m \) are not the same as we guessed in Figure 1, only similar; they are more complicated, with a big convex around \( \lambda = 1 \) (see Figures 9 and 10). Figures 10 and 11 show the relationship between implied \( m \) (or \( q \)) and the moneyness \( \lambda \) at maturities \( T = 0.28 \). If we can find suitable functions to fit them very well, we can use the method of distorted lognormal distribution we proposed to all the prices of options with all the strikes from zero to \(+\infty\) (or big enough), whether they are traded in the real market or not. Then, we can use these options to compute the moments and tail risk measures more precisely. However, we cannot find a good function to fit them well, which may be our work in future. So, we use another method—interpolation—to extend the number of strikes. Though the interpolation cannot extend the range of strikes outside of the range of real market, it can enlarge the number of strikes inside that range, where there are no options traded in the real market. As we know, the number of options traded in the market is limited.

The method of interpolation will be useful for the markets where there are not enough options. To verify this, we narrow the number of samples and then make interpolation treatment. Figures 10 and 11 show the original data and their interpolations. The interpolation method here chosen is spline. The blue dot is the original data, there are 63 points for maturity \( T = 0.28 \), and the red curve denotes interpolation. We can see that even for S&P500 options market, the points are not enough, especially for the right side where \( \lambda \) is high. For some option markets, such as some single stock-based option market and Chinese options market, there are not enough options traded, which will be similar to Figure 12. In Figure 12, we decrease the number of samples by deleting part of points. From Figures 12(a)–12(d), the number of samples decreases and becomes more and more sparse. Here, we do not delete the first and the last...
original points to hold the same range of strikes. In the real market, that will not be the truth. For the traditional method, to replicate risk measures by options, enough samples are very important, especially for tail risk. In case of Figure 12(d), there are only 12 points; if we compute the tail risk, there are only one or two points that can be used, and the result may be very bad. Our interpolation method can overcome this problem.

In the next section, we will use this method to price the options and apply these prices to compute moments and tail risks.

3. Applications to Moments and Tail Risk Estimation

In this section, we will use the methods suggested by Andersen et al. [14] and LDS (2014, 2015) to construct moments and tail risk measures via options with a range of strike prices, and for comparison, we will use the pricing model we proposed to compute them.

3.1. Theory. Firstly, we will explain how to compute $E[f(S)]$ with $f$ twice differential as a function of the prices of the options and approximation for the finite number of strikes available.
Table 3: Descriptive statistics of $m$ and $q$.

| $T$ | Max | Min | Mean | Std | Skewness | Kurtosis | Max | Min | Mean | Std | Skewness | Kurtosis |
|-----|-----|-----|------|-----|----------|----------|-----|-----|------|-----|-----------|----------|
| 0.03 | 5.15 | -0.51 | -0.17 | 0.58 | 5.46 | 39.50 | -0.27 | -71.41 | -10.4 | 11.5 | -2.02 | 8.69 |
| 0.1  | 4.82 | -0.65 | -0.17 | 0.54 | 6.99 | 56.60 | -0.09 | -48.17 | -4.79 | 6.64 | -3.09 | 15.90 |
| 0.2  | 0.30 | -0.67 | -0.21 | 0.17 | -1.48 | 4.78 | -0.05 | -24.93 | -3.34 | 4.34 | -2.37 | 9.67 |
| 0.28 | 1.55 | -0.66 | -0.16 | 0.33 | 2.79 | 15.12 | -0.06 | -9.05 | -2.62 | 2.65 | -0.80 | 2.41 |
| 0.35 | 3.11 | -0.66 | -0.01 | 0.63 | 3.28 | 14.05 | -0.08 | -24.90 | -4.01 | 4.98 | -1.84 | 6.79 |
| 0.6  | 1.17 | -0.62 | -0.14 | 0.25 | 2.30 | 14.18 | -0.11 | -18.97 | -3.02 | 3.70 | -1.95 | 7.40 |
| 0.85 | 0.65 | -0.54 | -0.16 | 0.15 | 0.87 | 11.16 | -0.13 | -24.92 | -3.30 | 4.22 | -2.47 | 10.97 |
| 0.95 | 1.32 | -0.58 | -0.11 | 0.29 | 3.21 | 15.62 | -0.15 | -23.90 | -3.01 | 4.04 | -2.58 | 11.36 |
| 1.35 | 0.02 | -0.48 | -0.14 | 0.12 | -0.71 | 3.47 | -0.17 | -19.52 | -2.55 | 3.29 | -2.82 | 12.63 |
| 1.87 | 0.03 | -0.37 | -0.13 | 0.10 | 0.01 | 1.74 | -0.19 | -16.26 | -2.41 | 2.82 | -2.43 | 10.20 |
| 2.87 | 0.02 | -0.31 | -0.13 | 0.11 | -0.23 | 1.60 | -0.27 | -12.72 | -1.84 | 2.04 | -2.87 | 13.29 |

Figure 10: Implied $q$ with different $\lambda$ for $T = 0.28$.

Figure 11: Implied $m$ with different $\lambda$ for $T = 0.28$.

Figure 12: Continued.
We will use the same approach as the paper on corridor VIX [14] or on HIX [23, 24]. This approach is from the idea of Carr and Madan [25].

From Carr and Madan [25], for any twice continuously differentiable function $f(x)$, it can be expressed as follows:

$$f(x) = f(F) + f'(F)(x - F) + \int_{F}^{\infty} f''(v)(x - v)^+ \, dv + \int_{0}^{E} f''(v)(v - x)^+ \, dv.$$  \hfill (20)

Linders et al. [12] used this to express the swap rate $E[f(S/E[S])]$ with $F = 1$, as follows:

$$E\left[ f\left(\frac{S}{E[S]}\right) - f(1) \right] = E\left[ \int_{1}^{E} f''(v)\left(\frac{S}{E[S]} - v\right)^+ \, dv + \int_{0}^{E} f''(v)\left(v - \frac{S}{E[S]}\right)^+ \, dv \right].$$

$$= \frac{1}{E[S]^2} \left( \int_{E[S]}^{\infty} f''\left(\frac{K}{E[S]}\right) E[(S - K)^+] \, dK + \int_{0}^{E[S]} f''\left(\frac{K}{E[S]}\right) E[(K - S)^+] \, dK \right)$$

$$= \frac{e^{rT}}{(E[S])^2} \left( \int_{E[S]}^{\infty} f''\left(\frac{K}{E[S]}\right) C(K) \, dK + \int_{0}^{E[S]} f''\left(\frac{K}{E[S]}\right) P(K) \, dK \right).$$  \hfill (21)

where $C[K] = E[(S - K)^+]e^{-rT}$ and $P[K] = E[(K - S)^+]e^{-rT}$. Equation (21) means that for any twice continuously differentiable function $f(x)$, its expectation can be expressed as the integral of European puts and calls at different strikes. Because the options market is looking forward, they can react with the information more actively for the future than the underlying assets markets, so we can use this method to construct some kind of measures in a way of model-free estimation to forecast the market more precisely.

Apparently, equation (21) requires that there must exist options with all the strikes from 0 to infinity continuously. However, only some of them exist in the real options market with limited discrete strikes. To deal with this problem, Linders et al. [12] used the composite trapezoidal rule; they approximated formula (21) as follows:

$$E\left[ f\left(\frac{S}{E[S]}\right) - f(1) \right] \approx \frac{e^{rT}}{(E[S])^2} \left( \sum_{i=1}^{b} \frac{f''\left(\frac{K_i}{E[S]}\right)}{2} \Delta K_i - \frac{f''\left(\frac{K_0}{E[S]}\right)}{2} \left(\frac{E[S] - K_0}{E[S]}\right)^2 \right).$$  \hfill (22)
where

\[
\Delta K_i = \begin{cases} 
K_{i+1} - K_{i}, & \text{if } i = -l, \\
K_{i+1} - K_{i-1}, & \text{if } i = -l + 1, \ldots, h - 1, \\
K_h - K_{h-l}, & \text{if } i = h,
\end{cases}
\]

and

\[
Q[K_i] = \begin{cases} 
P[K_i], & \text{if } K_i < K_0, \\
\frac{C[K_i] + P[K_i]}{2}, & \text{if } K_i = K_0, \\
C[K_i], & \text{if } K_i > K_0,
\end{cases}
\]

Throughout the whole paper, we will assume that there are only a finite number of European options with maturity \( T \). In particular, for the underlying asset (SP500 index), the strikes of the traded puts and calls are denoted by \( K_i, i = -l, -l + 1, \ldots, h - 1, h \), with

\[
0 = K_{-l} < \cdots < K_{-1} < K_0 \leq E[S] < K_1 < K_2 < K_h < K_{h+1},
\]

(24)

where \( K_{h+1} = F^{-1}_S(1) \) is assumed to be finite. In reality, the underlying asset and call option based in it have unknown upward potential. For discrete approximation, here a finite upper bound can be chosen arbitrarily large. Note that as long as there is at least one strike \( K \) for which the prices \( C[K] \) and \( P[K] \) are traded, the forward rate \( E[S] \) can be computed in model-free way using the put-call parity. In practical situations, we follow the methodology proposed in Chicago Board Options Exchange [26] to determine the forward rate of the SP500 index:

\[
E[S] = e^{rT} (C[K^*] - P[K^*]) + K^*, \quad (25)
\]

with

\[
K^* = \arg \min_{K \in [K_{-h}, K_{h+1}]} |C[K] - P[K]|. \quad (26)
\]

Now we use this method to estimate some kinds of measures.

**Example 1.** Estimation of \( E[S^0] \).

The first is to estimate moments of \( E[S^0] \). When \( \theta = 1 \), \( E[S] \) can be estimated.

When \( \theta = 2 \), \( f'^\theta (K_i/E[S]) = 2 \), we have

\[
E\left[ \left( \frac{S}{E[S]} \right)^2 - 1 \right] = \frac{\text{Var}[S]}{(E[S])^2} \approx 2e^{rT} \left( \sum_{i=-l}^{h} Q[K_i] \Delta K_i \right)
\]

(27)

When \( \theta = 3 \), \( f'^\theta (K_i/E[S]) = 6K_i/E[S] \), we have

\[
E\left[ \left( \frac{S}{E[S]} \right)^3 - 1 \right] \approx \frac{e^{rT}}{(E[S])^3} \left( \sum_{i=-l}^{h} 6K_i Q[K_i] \Delta K_i \right) - \frac{6K_0}{E[S]} \left( \frac{E[S] - K_h}{E[S]} \right)^2.
\]

(28)

**Example 2.** Estimation of range of variance \( RVar(S) \).

In the real market, the extreme situations are not stable; this means that the options price with extreme strikes may not be good to estimate the variance, so these extreme strike situations must be discarded. Andersen et al. [14] used this method to replicate implied volatility; they called it Corridor Implied Volatility Index, or \( CX \), and through the empirical test, they proved that this \( CX \) is more robust than the ordinary Implied Volatility Index (VIX). In fact, the basic theory they considered is as follows:

\[
E\left[ f\left( \frac{S}{E[S]} \right) - f(1) \right] 1_{S \in [K_{-l}, K_{l}]} = E\left[ \int_0^\infty f''(v) \left( \frac{S}{E[S]} - v \right) \frac{1_{S \in [K_{-l}, K_{l}]} dv}{E[S] v} \right] + E\left[ \int_0^1 f''(v) \left( v - \frac{S}{E[S]} \right) \frac{1_{S \in [K_{-l}, K_{l}]} dv}{E[S] v} \right]
\]

\[
= \frac{1}{(E[S])^2} \left( \int_0^\infty f''(K/E[S]) E\left[ (S - K)^1_{S \in [K_{-l}, K_{l}]} \right] dK + \int_0^{E[S]} f''(K/E[S]) E\left[ (K - S)^1_{S \in [K_{-l}, K_{l}]} \right] dK \right)
\]

\[
= \frac{e^{rT}}{(E[S])^2} \left( \int_0^\infty f''(K/E[S]) C_R[K] dK + \int_0^{E[S]} f''(K/E[S]) P_R[K] dK \right).
\]

(29)

where \( C_R[K] = E\left[ (S - K)^1_{S \in [K_{-l}, K_{l}]} \right] e^{-rT} \) and \( P_R[K] = E\left[ (K - S)^1_{S \in [K_{-l}, K_{l}]} \right] e^{-rT} \); apparently, \( C_R[K] \) and \( P_R[K] \) are not options in the real market, so they cannot use equation (22) to approximate \( E\left[ f\left( \frac{S}{E[S]} \right) - f(1) \right] 1_{S \in [K_{-l}, K_{l}]} \) with the real market prices of options. The real reason of this problem is that \( 1_{S \in [K_{-l}, K_{l}]} \) is stochastic. To solve this problem, they substitute it with \( 1_{K \in [K_{-l}, K_{l}]} \), which is not stochastic. This is an approximation method. Then, equation (29) transforms to
If by call options which are out of the money. We derivates for $VaR$ and $TVaR$ are zero, we cannot replicate which will be not forward-looking. Because the second calculations need to estimate the risk-neutral density (RND), risk, especially for the left tail; many risk measures are situation.

Andersen et al. [14] is one of the special cases of this and can be replicated both by call and put options which are out of them by options prices as the method we proposed here. We only consider the right tail and can be replicated only by call options. Using equation (21), we have

$$E\left[ f\left( \frac{S}{E[S]} \right) - f(1) \right]_{K_e[k_a,k_b]} = \frac{e^{rT}}{(E[S])^2} \left( \int_{K_a}^{K_b} f''(\frac{K}{E[S]}) C[K] dK + \int_{0}^{E[S]} f''(\frac{K}{E[S]}) P[K] dK \right)_{K_e[k_a,k_b]},$$

(31)

where $C[K] = E[(S-K)^+]e^{-rT}$ and $P[K] = E[(K-S)^+]e^{-rT}$. If $K_a < E[S] < K_b$, then equation (31) will be

$$E\left[ f\left( \frac{S}{E[S]} \right) - f(1) \right]_{K_e[k_a,k_b]} = \frac{e^{rT}}{(E[S])^2} \left( \int_{K_a}^{K_b} f''(\frac{K}{E[S]}) C[K] dK + \int_{K_a}^{E[S]} f''(\frac{K}{E[S]}) P[K] dK \right).$$

(32)

It only considers the range without the left and right tails and can be replicated both by call and put options which are out of the money. The Corridor Implied Volatility Index of Andersen et al. [14] is one of the special cases of this situation.

In most of situations, people only consider the downside risk, especially for the left tail; many risk measures are created to quantify the risk of this tail, such as Value at Risk (VaR) and Average Value at Risk (TVaR). Similarly, their calculations need to estimate the risk-neutral density (RND), which will be not forward-looking. Because the second derivatives for VaR and TVaR are zero, we cannot replicate them by options prices as the method we proposed here. We can consider the moments higher than first order, such as second and third moments of tail, to quantify the risk of this tail. Using equation (32), moments of tails can be quantified by replicating with options. If $K_a < K_b < E[S]$, then equation (32) will be

$$E\left[ f\left( \frac{S}{E[S]} \right) - f(1) \right]_{K_e[k_a,k_b]} = \frac{e^{rT}}{(E[S])^2} \int_{K_a}^{K_b} f''(\frac{K}{E[S]}) P[K] dK.$$

(33)

It only considers the left tail and can be replicated only by put options which are out of the money. If $E[S] < K_a < K_b$, then equation (31) will be

$$E\left[ f\left( \frac{S}{E[S]} \right) - f(1) \right]_{K_e[k_a,k_b]} = \frac{e^{rT}}{(E[S])^2} \int_{K_a}^{K_b} f''(\frac{K}{E[S]}) C[K] dK.$$

(34)

It only considers the right tail and can be replicated only by call options which are out of the money.

For these three kinds of risk measures, we can compute arbitrary moments of them, except the first-order moment.

For the same reason, to calculate the moments through real options market, approximation by discretization must be done. Then using equation (22), equation (31) can be approximated by discretization as follows:

$$E\left[ f\left( \frac{S}{E[S]} \right) - f(1) \right]_{K_e[k_a,k_b]} \approx 2\frac{e^{rT}}{(E[S])^2} \left( \sum_{i=0}^{h_y} f''(\frac{K_i}{E[S]}) Q[K_i] \Delta K_i - \left( E[S] - K_0 \right)^2 \right),$$

(35)

where $K_{i+1} \geq K_a$ and $K_i \leq K_b$. $K_a = F^{-1}_S(\alpha)$, and $K_b = F^{-1}_S(\beta)$; we only need to choose proper $\alpha$ and $\beta$ to make the results stable or robust. That is a problem, not only because of selection of $\alpha$ and $\beta$, but also because of the calculation of the quantile of $S$, which means we need to estimate the risk-neutral density (RND) of $S$, which is not easy. How to solve this problem? Firstly, we try to determine $K_n$ by a fixed proportion of $E[S]$ under the standard log-normal assumption simply, such as $K_n = k_s E[S]$. And $k_s$ can be chosen based on

$$k_s = \frac{F^{-1}_S(\alpha)}{E[S]},$$

(36)

where $B$ refers to geometrical Brownian motion.

But, we find it is not good. Figure 13 shows $k_s$ with the confidence $\alpha$. From this figure, we can find that even when confidence is approaching zero, $k_s$ is still higher than 0.8. From Table 2, we can see that the median value of $\lambda = 0.81$, which means this method will lose a large part of information of the options market.
To avoid this difficulty, Andersen et al. [14] proposed a good method. They adopted the approach of Andersen and Bonareno [27]. It builds on the fact that option prices reflect tail moments. The left and right tail moments of a positive random variable, $x$ with strictly positive density $f(x)$ for all $x > 0$, are given by

$$LT(K) = \int_{0}^{K} (K - x) f(x) dx,$$

$$RT(K) = \int_{K}^{\infty} (x - K) f(x) dx,$$

and then they define the ratio statistic, $R(K)$, as an indicator of how far in the tail a given point, $K$, is located within the support of $x$:

$$R(K) = \frac{LT(K)}{LT(K) + RT(K)}$$

$R(K)$ is akin to a cumulative density function or CDF as it is increasing on $(0, \infty)$ with $R(0) = 0$ and $R(\infty) = 1$. So, for a given percentile, $q$, in the range of the $R(K)$ function, we can get the quotient, $K_q$, as

$$K_q = R^{-1}(q), \text{ for any } q \in [0, 1].$$

Let $f(x)$ denote the risk-neutral density for SP500 forward price at maturity $T = 1/12$, or one month, and $K$ be the strike price of European style put and call options; we have

$$R(K) = \frac{P(K)}{P(K) + C(K)}$$

The ratio statistic is computed only from put and call prices. Hence, if the range is located where option prices may be extracted reliably, $R(K)$ can be computed without estimation of risk-neutral density. Here, we can borrow this method to determine the $K_a, K_b$. After considering the market liquidity, we can settle the 2% truncation level for the tails. That is,

$$K_a \geq R^{-1}(0.02),$$

$$K_b \leq R^{-1}(0.98).$$

Utilizing this method, we can compute the moments via replicating options prices in the real market. However, the options traded are limited, and these computations may not be so good. And all of the range moments are just approximations as shown in equation (30); it will have big error. To solve these two problems, we propose the alternative approach in the next section.

3.2. Alternative Approach. As mentioned above, the options traded in the real market are limited; especially, for some extreme strikes, it will be sparse. Using the limited options to replicate the risk measures will lead to errors. And when calculating the range moments, the method of approximations as shown in equation (30) is not convincing. In this section, we will propose some methods to solve these two problems.

In the first section, we propose a novel method to evaluate the options well. It shows that the weight function can be fitted very well. So, for the first problem, we can use the method of distorted lognormal distribution to get all $m(\lambda)$ and $q(\lambda)$, from the limited options traded in the real market. Then, we use interpolation to get enough points of $m(\lambda)$ and $q(\lambda)$ between the minimum and maximum of $\lambda$. Based on this, we can get all the prices of call and put options with these strikes ($\lambda$). Then, we can use equation (21) to compute the risk measures precisely.

In fact, in equation (21), the options $Q[K]$ are supposed to be existing in the real market with price $\bar{Q}[K]$, and by using the implied volatility equation, we have

$$\bar{Q}[K] = Q[\bar{\sigma}, K],$$

where $Q[\bar{\sigma}, K]$ is $C[\bar{\sigma}, K]$ for call options and $P[\bar{\sigma}, K]$ for put options. Then, equation (20) transforms into

$$E \left[ f \left( \frac{S}{E[S]} \right) - f(1) \right] = e^{\alpha T} E[S] \left( \int_{E[S]}^{\infty} f'' \left( \frac{K}{E[S]} \right) C[\bar{\sigma}, K] dK + \int_{0}^{E[S]} f'' \left( \frac{K}{E[S]} \right) P[\bar{\sigma}, K] dK \right).$$
This means that equation (43) is the function of implied volatility of different options with all the strikes. Because we do not have the real traded prices for options with all the strikes, we can approximate equation (43) well. Here, we remember that if \( m(\lambda) \) and \( q(\lambda) \) can be obtained, we can use them to price all the options with all the strikes. Namely, we assume that the price of our new price model can fit the market prices well, so we can offer the option prices for strikes that are not traded in the real market:

\[
\overline{Q}[K] = Q[\bar{\sigma}, K] = Q^*[\sigma, K].
\]  

(44)

So, equation (43) can be rewritten as

\[
E^2 \left[ E \left( \frac{S}{E[S]} \right) - f(1) \right] = \frac{e^{\theta T}}{(E[S])^2} \left( \int_{E[S]}^{\infty} f'' \left( \frac{K}{E[S]} \right) C^*[\sigma, K] dK + \int_{0}^{E[S]} f'' \left( \frac{K}{E[S]} \right) P^*[\sigma, K] dK \right). 
\]  

(45)

Moreover, denote \( K = \lambda S_0 \); then, we have

\[
E^2 \left[ E \left( \frac{S}{E[S]} \right) - f(1) \right] = \frac{e^{\theta T}}{(E[S])^2} \left( \int_{E[S]0}^{\infty} f'' \left( \frac{\lambda S_0}{E[S]} \right) C^*[\sigma, \lambda S_0] d\lambda S_0 + \int_{0}^{E[S]0} f'' \left( \frac{\lambda S_0}{E[S]} \right) P^*[\sigma, \lambda S_0] d\lambda S_0 \right) 
\]  

(47)

\[
= \frac{e^{\theta T S_0^2}}{(E[S])^2} \left( \int_{E[S]0}^{\infty} f'' \left( \frac{\lambda S_0}{E[S]} \right) C^*[\sigma, \lambda] d\lambda + \int_{0}^{E[S]0} f'' \left( \frac{\lambda S_0}{E[S]} \right) P^*[\sigma, \lambda] d\lambda \right),
\]  

where

\[
C^*_1[\sigma, \lambda] = e^{-\lambda^2 T} N(d_1^*) - \lambda e^{-\lambda T} N(d_2^*),
\]  

(48)

\[
P^*_1[\sigma, \lambda] = \lambda e^{-\lambda T} N(-d_2^*) - e^{\lambda^2 T} N(-d_1^*),
\]  

(49)

and the definitions of \( d_1^* \) and \( d_2^* \) are

\[
d_1^* = \frac{\log(\lambda) - (r - \sigma^2/2)T}{\sigma \sqrt{T}} + \sigma \sqrt{T} + q,
\]  

(50)

\[
d_2^* = d_1^* - \sigma \sqrt{T}.
\]  

We can see that equations (48) and (49) are the functions of \( \lambda, m(\lambda) \) and \( q(\lambda) \).

For the second problem, we can still solve it using our novel method of distorted lognormal distribution. In equation (29),

\[
E^2 \left( \frac{f(S)}{E[S]} - f(1) \right) 1_{Se[K, K_i]} = -\frac{e^{\theta T}}{(E[S])^2} \left( \int_{E[S]}^{\infty} f'' \left( \frac{K}{E[S]} \right) C_R[K] dK + \int_{0}^{E[S]} f'' \left( \frac{K}{E[S]} \right) P_R[K] dK \right),
\]  

(51)

where \( C_R[K] = E[(S - K)^+ 1_{Se[K, K_i]}] e^{-\lambda T} \) and \( P_R[K] = E[(K - S)^+ 1_{Se[K, K_i]}] e^{-\lambda T}. \)

Because \( C_R[K] \) and \( P_R[K] \) are not options in the real market, we cannot use equation (22) to approximate \( E[(f(S/E[S])) - f(1)]] 1_{Se[K, K_i]} \) with the real market prices of options. The papers before \([14, 24]\) use a nonstochastic indicator \( 1_{Ke[K, K_i]} \) to take the place of the stochastic indicator \( 1_{Se[K, K_i]} \). Using our method, we can calculate them easily:
\[ E \left[ \left( f \left( \frac{S}{E[S]} \right) - f(1) \right) 1_{S \leq K_a, K_b} \right] = \frac{e^{rT}}{(E[S])^2} \left( \int_{E[S]}^{\infty} f''(K) C_R[K] dK + \int_{E[S]}^{E[S]} f''(K) P_R[K] dK \right), \]

where

\[ C_R[K] = E^{Q^*} \left[ (S - K)^{1_{S \leq K_a, K_b}} e^{-rT} \right] = E^{Q^*} \left[ (S - K)^{1_{S \leq K_a, K_b}} 1_{K_a < K_b} e^{-rT} \right] + E^{Q^*} \left[ (S - K)^{1_{S \leq K_a, K_b}} 1_{K_b < K_a} e^{-rT} \right] \]

\[ = E^{Q^*} \left[ (S - K)^{1_{S \leq K_a, K_b}} 1_{K_a < K_b} e^{-rT} \right] + E^{Q^*} \left[ (S - K)^{1_{S \leq K_a, K_b}} 1_{K_b < K_a} e^{-rT} \right] \]

\[ = E^{Q^*} \left[ (S - K)^{1_{S \leq K_a, K_b}} 1_{K_a < K_b} e^{-rT} \right] + E^{Q^*} \left[ (S - K)^{1_{S \leq K_a, K_b}} 1_{K_b < K_a} e^{-rT} \right] \]

\[ P_R[K] = E^{Q^*} \left[ (K - S)^{1_{S \leq K_a, K_b}} e^{-rT} \right] = E^{Q^*} \left[ (K - S)^{1_{S \leq K_a, K_b}} 1_{K_a < K_b} e^{-rT} \right] + E^{Q^*} \left[ (K - S)^{1_{S \leq K_a, K_b}} 1_{K_b < K_a} e^{-rT} \right] \]

\[ = E^{Q^*} \left[ (K - S)^{1_{S \leq K_a, K_b}} 1_{K_a < K_b} e^{-rT} \right] + E^{Q^*} \left[ (K - S)^{1_{S \leq K_a, K_b}} 1_{K_b < K_a} e^{-rT} \right] \]

\[ = E^{Q^*} \left[ (K - S)^{1_{S \leq K_a, K_b}} 1_{K_a < K_b} e^{-rT} \right] + E^{Q^*} \left[ (K - S)^{1_{S \leq K_a, K_b}} 1_{K_b < K_a} e^{-rT} \right] \]

As we show that our novel distorted method can price the options very well, all \( C^*[K] \), \( P^*[K] \), \( d(m,K) \), and \( d(q,K) \) at arbitrary strikes between \( K_{\text{min}} \) and \( K_{\text{max}} \) of the real market (which corresponds to \( \lambda_{\text{min}}, \lambda_{\text{max}} \)) can be calculated precisely. We do not need to use the approximation equation (30) anymore. Similarly, we can get different situations of range moments, including tail moments as equations (32) and (34):

If \( K_a < K_b < E[S] \), then equation (52) transforms to

\[ E \left[ \left( f \left( \frac{S}{E[S]} \right) - f(1) \right) 1_{S \leq K_a, K_b} \right] = \frac{e^{rT}}{(E[S])^2} \left( \int_{K_a}^{E[S]} f''(K) P^*[K] dK \right) + \int_{K_a}^{E[S]} f''(K) dK \]

\[ - (K - K_a) e^{-rT} N(d(q, K_a)) dK. \]
This situation is to measure the risk of the left tail; it can be replicated by a series of put options out of the money.

\[
E\left[\left(\frac{S}{E[S]} - (1)\right)1_{k}\right] = \frac{e^{RT}}{(E[S])^2} \int_{K_a}^{K} f''\left(\frac{K}{E[S]}\right)\left(C^* [K] - C^* [K_a] - (K - K_a)e^{-RT}N[-d(m,K_a)]\right) dK + \int_{K_a}^{E[S]} f''\left(\frac{K}{E[S]}\right)\left(P^* [K] - P^* [K_a] - (K - K_a)e^{-RT}N[d(q,K_a)]\right) dK.
\]

(57)

This situation is to measure the risk of the middle without left and right tails; it can be replicated by a series of put and call options out of the money.

\[
E\left[\left(\frac{S}{E[S]} - (1)\right)1_{k}\right] = \frac{e^{RT}}{(E[S])^2} \int_{K_a}^{K} f''\left(\frac{K}{E[S]}\right)\left(C^* [K] - C^* [K_a] + (K_a - K)e^{-RT}N[-d(m,K_a)]\right) dK + \int_{K_a}^{E[S]} f''\left(\frac{K}{E[S]}\right)\left(C^* [K] - C^* [K_a]\right) dK.
\]

(58)

This situation is to measure the risk of the right tail; it can be replicated by a series of call options out of the money.

\[
E\left[\left(\frac{S}{E[S]} - (1)\right)1_{k}\right] = \frac{2e^{RT}}{(E[S])^2} \sum_{i=1}^{h} f''\left(\frac{K_i}{E[S]}\right)Q_R^*[K_i] \Delta K_i - \left(\frac{E[S] - K_0}{E[S]}\right)^2.
\]

(59)

where

\[
Q_R^*[K_i] = \begin{cases} 
P_R^*[K_i], & \text{if } K_i < K_0, \\
P_R^*[K_i] + \frac{C_R^*[K_i]}{2}, & \text{if } K_i = K_0, \\
C_R^*[K_i], & \text{if } K_i > K_0, 
\end{cases}
\]

and \(\Delta K_i\) and other variables have the same definitions as that of Section 3.1.

In next section, we will use the real data of SP options market to calculate the moments and tail risk and compare these different methods mentioned here.

3.3. Comparison. In this section, we will utilize the real data of SP options market to calculate the moments and tail risk and compare these different methods mentioned here. We compare three kinds of approaches to compute two moments of risks with 3 pairs of ranges.

The three kinds of approaches are as follows: the first approach is to replicate risk measures by limited options traded in the market, used by current papers (see [14, 24]); the second approach is to replicate risk measures by enough interpolation points through our novel distorted lognormal distribution; these two methods are based on the nonstochastic indicator \(I_{K_0(K_a,K_b)}\). Here, we call the first method NSO (nonstochastic indicator with original data) and call the second method NSDI (nonstochastic indicator with distorted and interpolation data). And the third approach is to
replicate risk measures by enough interpolation points through our novel distorted lognormal distribution based on the stochastic indicator \( I_{\text{SE}}[K_a, K_b] \). We call the third method SDI (stochastic indicator with distorted and interpolation data).

The two moments are for \((S/E[S])^\theta\) with \(\theta = 2, 3\). We choose three pairs of range \([K_a, K_b]\) by changing \([\alpha, \beta] = \{[0, 1], [0.02, 0.98]\}, \{[0.01, 0.02], 0.1\}, \{[0.90, 1.0], [0.90, 0.98]\}\). Here, \([\alpha, \beta] = [0, 1]\) means calculations will be done by all the options traded in the market without discarding the extreme strikes of the tails. This situation can be viewed as a special case of range risk measures. The range \([K_{\min}, K_{\max}]\) is the maximum range of strikes for the options traded in the real market. \([\alpha, \beta] = [0.02, 0.98]\) means calculations will be done by some of the options with discarding 2% of the extreme strikes of both tails. According to Andersen et al. [14], this range will make the risk measures more stable.

The second pair of range \([\alpha, \beta] = [[0, 0.1], [0.02, 0.1]]\) is for the left tail risk without and with discarding 2% of the extreme strikes of the left tail.

The third pair of range \([\alpha, \beta] = [[0.9, 1.0], [0.9, 0.98]]\) is for the right tail risk without and with discarding 2% of the extreme strikes of the right tail.

We want to check if discarding 2% of the extreme strikes can make the risk measure to be more stable. \([\alpha, \beta] = [0.02, 0.10]\) will be replicated by some of the put options out of the money with discarding 2% of the extreme strikes of the left tail to make it stable. The right tail risk will be replicated by some of the call options out of the money with discarding 2% of the extreme strikes of the right tail to make it stable.

Besides these situations, to show that our interpolation methods are good for markets with sparse options, we increase the number of the samples and then compute those risk measures again. The method to decrease is as follows: \(\text{mod}(N, k) = 0\). For example, there are \(N = 12\) samples; \(\text{mod}(12, 2) = 0\) means only 6 samples are left, and they are 2, 4, 6, 8, 10, 12; \(\text{mod}(12, 6) = 0\) means only 2 samples are left, and they are 6, 12. Here, we will take \(k = 1, 2, \ldots, 6\). We can see that the number of samples decreases to \(1/k\) of original data. With \(k\) increasing, when \(k = 6\) or \(k = 5\), it will be very sparse, and there are not enough samples to finish the calculations, especially for the tail risk.

The data are the same as those in Section 2.2, and we use the method of Section 2 to estimate \(m(\lambda)\) and \(q(\lambda)\), and then we use the interpolation method to get all the points of \(m(\lambda)\) and \(q(\lambda)\) between \(\lambda_{\min}\) and \(\lambda_{\max}\). We will get 3000 points, which is enough.

Firstly, we compute \((S/E[S])^\theta\) with \(\theta = 2\). To be convenient, we denote

\[
\rho_{[\alpha, \beta]} = \sqrt{E \left[ \left( \frac{S}{E[S]} \right)^2 - 1 \right] I_{\text{SE}}[K_a, K_b]} \cdot (61)
\]

We compute all \(\rho_{[\alpha, \beta]}\) with different situations for all the 11 maturities. Here, we only show results of \(T = 0.28\) as an example because the results for the other maturities are similar.

Table 4 shows the results of \(\rho_{[0,1]}\) and its truncated range \(\rho_{[0.02,0.98]}\) with three approaches NSO, NSDI, and SDI and 6 kinds of number of samples \((k = 1, 2, \ldots, 6)\). It shows that NSO > NSDI > SDI. NSO > NSDI is expected because the former is coarser than the latter. NSDI > SDI can be explained too. They have the same integral interval, but there is a negative adjustment of integral term for SDI because of the truncated tails. For \(\rho_{[0,1]}\), there are still truncated tails with the range of \([K_{\min}, K_{\max}]\). For \(\rho_{[0.02,0.98]}\), the range is smaller, so difference between NSDI and SDI is bigger (theoretically, with the narrowing of range, this difference will increase first and then decrease to zero from equation (52)). This means that the risks of this kind (with \(K_a < E[S] < K_b\)) can be overvalued by fixed indicator methods (NSO and NSDI), relative to the stochastic indicator method (SDI). From Table 4, it shows that the distorted and interpolation methods (NSDI and SDI) are more stable than the traditional method (NSO) through different samples from \(k = 1\) to \(k = 6\). And the range moments \(\rho_{[0.02,0.98]}\) are more stable than \(\rho_{[0,1]}\).

Table 5 shows the results of \(\rho_{[0,0.1]}\) and its truncated range \(\rho_{[0.02,0.1]}\) with three approaches NSO, NSDI, and SDI and 6 kinds of number of samples \((k = 1, 2, \ldots, 6)\). It shows that NSO > NSDI < SDI. NSO > NSDI is the same as Table 4. NSDI < SDI is different from Table 4. It may be explained by the bigger integral interval because of the truncated tail effect (see equation (56)). For \(\rho_{[0,0.1]}\), there are still truncated tails with range of \([K_{\min}, K_{0.1}]\). This means that the risks of this kind (with \(K_a < E[S] < K_b\)) can be undervalued by fixed indicator methods (NSO and NSDI), relative to the stochastic indicator method (SDI). From Table 5, it shows that the distorted and interpolation methods (NSDI and SDI) are more stable than the traditional method (NSO) through different samples from \(k = 1\) to \(k = 6\). For NSO, when \(k = 6\), its number of samples is too small, only one, so that \(\rho_{[a,b]}\) cannot be calculated well (we only show it with zero here). And the range moments \(\rho_{[0.02,0.01]}\) are more stable than \(\rho_{[0,0.01]}\).

Table 6 shows the results of \(\rho_{[0.9,1]}\) and its truncated range \(\rho_{[0.90,0.98]}\) with three approaches NSO, NSDI, and SDI and 6 kinds of number of samples \((k = 1, 2, \ldots, 6)\). The results are almost the same as Table 5. Table 5 shows the left tail, and Table 6 shows the right tail. Theoretically, they must be similar. But it shows differences and unstability for NSO; sometimes they are even bigger than NSDI. This is because of the number of samples. For maturity \(T = 0.28\), there are only 63 samples for strikes and only a small part of them at the right tail; for \(k = 1\), there are only 6 and 4 samples for \(\rho_{[0.9,1]}\) and \(\rho_{[0.90,0.98]}\), respectively. So, the results are unstable and incorrect. This shows that our distorted and interpolation method is better than the traditional method. The other results are similar to Table 5, such as NSDI < SDI, and the range moments \(\rho_{[0.90,0.98]}\) are more stable than \(\rho_{[0.9,1]}\).

The results for the other maturities \(T\) are similar to \(T = 0.28\). To show this, we present the results in Figures 14–16 for three pairs of ranges, respectively. These three figures are outcomes calculated from the original data without decreasing \((k = 1)\). We can see that for range
Table 4: Second range moment for $T = 0.28$.

| $k$ | NSO | $\rho_{[0,1]}$ | NSDI | SDI | NSO | $\rho_{[0.02,0.98]}$ | NSDI | SDI |
|-----|-----|----------------|------|-----|-----|---------------------|------|-----|
| 1   | 0.0762 | 0.0761 | 0.0760 | 0.0713 | 0.0707 | 0.0574 |
| 2   | 0.0765 | 0.0759 | 0.0757 | 0.0720 | 0.0706 | 0.0574 |
| 3   | 0.0773 | 0.0761 | 0.0759 | 0.0725 | 0.0707 | 0.0575 |
| 4   | 0.0780 | 0.0754 | 0.0748 | 0.0743 | 0.0704 | 0.0575 |
| 5   | 0.0907 | 0.0877 | 0.0877 | 0.0766 | 0.0709 | 0.0579 |
| 6   | 0.0806 | 0.0755 | 0.0744 | 0.0777 | 0.0707 | 0.0575 |

Table 5: Second range moment for $T = 0.28$; left tail.

| $k$ | NSO | $\rho_{[0,0.1]}$ | NSDI | SDI | NSO | $\rho_{[0.02,0.1]}$ | NSDI | SDI |
|-----|-----|----------------|------|-----|-----|---------------------|------|-----|
| 1   | 0.0410 | 0.0409 | 0.0586 | 0.0331 | 0.0321 | 0.0459 |
| 2   | 0.0423 | 0.0407 | 0.0584 | 0.0351 | 0.0321 | 0.0459 |
| 3   | 0.0365 | 0.0404 | 0.0582 | 0.0275 | 0.0317 | 0.0457 |
| 4   | 0.0334 | 0.0405 | 0.0584 | 0.0252 | 0.0319 | 0.0459 |
| 5   | 0.0428 | 0.0407 | 0.0583 | 0.0385 | 0.0324 | 0.0461 |
| 6   | 0   | 0.0404 | 0.0581 | 0    | 0.0317 | 0.0457 |

Table 6: Second range moment for $T = 0.28$; right tail.

| $k$ | NSO | $\rho_{[0,0.1]}$ | NSDI | SDI | NSO | $\rho_{[0.02,0.98]}$ | NSDI | SDI |
|-----|-----|----------------|------|-----|-----|---------------------|------|-----|
| 1   | 0.0181 | 0.0183 | 0.0387 | 0.0139 | 0.0136 | 0.0187 |
| 2   | 0.0157 | 0.0183 | 0.0386 | 0.0113 | 0.0137 | 0.0189 |
| 3   | 0.0142 | 0.0191 | 0.0390 | 0.0080 | 0.0139 | 0.0190 |
| 4   | 0   | 0.0174 | 0.0369 | 0    | 0.0141 | 0.0190 |
| 5   | 0.0283 | 0.0170 | 0.0365 | 0.0267 | 0.0144 | 0.0198 |
| 6   | 0   | 0.0164 | 0.0359 | 0    | 0.0140 | 0.0189 |

Figure 14: $\rho_{[0,1]}$ and $\rho_{[0.02,0.98]}$ with different $T$. 
moment of type of $K_\alpha < E[S] < K_\beta$, NSO > NSDI > SDI; however, for the other types of tail risk, NSO > NSDI < SDI. Because the difference between NSO and NSDI is small, the curves in the figures are overlapped to one curve. For right tail, there are errors because of the limited samples and the nonrational market data at the extreme tail, and it performs better after being truncated (see Figure 16).

To elaborate the influence of size of samples on $\rho_{[k,\beta]}$ with different $T$, we put all the results of different $k$ in the same figure to compare them. Figures 17 and 18 show the influence of size of samples on $\rho_{[0,1]}$ and its truncated $\rho_{[0.02,0.08]}$. We can see from $k = 1$ to $k = 6$ that the influence of size of samples is weak. We attribute this result to the relatively large sample size between this interval. And it is clear that using the distorted and interpolation method (DI) is more sensitive to the size $k$ because this method will enlarge the effect of some extreme extraordinary value when the size of the samples is small. After truncated, they are better (see Figure 18). However, there is still an abnormal point for distorted and interpolation method (DI) in the figure. It is the second point for maturity $T = 0.1$. And we will see that this strange point exists in all the range moments when
considering the right tail risk. Finally, we try to seek the reason for it. Figure 19 shows the right tail of call options price for $T = 0.1$. It shows that there are extraordinary points from points A to B, especially for point C. As mentioned before, we know that if $\rho_{[\alpha, \beta]}$ involved the right tail, it will depend on the call options out of the money heavily. If there are extraordinary points in some sparse area, interpolation method will amplify the negative effect of these points.
Figure 19: Extraordinary value of call options price for $T = 0.1$. 

Figure 20: Influence of size of samples on $\rho_{[0.01]}$ with different $T$. 

Figure 21: Continued.
Figure 21: Influence of size of samples on $\rho_{[0.02,0.1]}$ with different $T$.

Figure 22: Influence of size of samples on $\rho_{[0.9,1]}$ with different $T$.

Figure 23: Continued.
Figure 23: Influence of size of samples on $\rho_{[0.9,0.98]}$ with different $T$.

Figure 24: Influence of size of samples on $\rho^\ast_{[0,1]}$ with different $T$.

Figure 25: Continued.
Figure 25: Influence of size of samples on $\rho^*_{[0.02,0.98]}$ with different $T$.

Figure 26: Influence of size of samples on $\rho^*_{[0,0.1]}$ with different $T$.

Figure 27: Continued.
extraordinary points; especially, for the situation there are several samples. Here Figures 17 and 18 show consistent performance.

Figures 20 and 21 show the influence of size of samples on $\rho_{[0.01]}$ and its truncated $\rho_{[0.02,0.1]}$. We can see from $k = 1$ to $k = 6$ that the influence of size of samples is very small for the distorted and interpolation method (DI), especially for truncated situations. That means that the distorted and interpolation method (DI) is better and the truncated method is better.

Figures 22 and 23 show the influence of size of samples on $\rho_{[0.9,1]}$ and its truncated $\rho_{[0.9,0.98]}$. They show the same results, except for points of maturity $T = 0.1$ because of the same reason we have explained.

Besides observing the influences of size of samples on $\rho_{[0.01]}$ we check the influences of range length of $[\alpha, \beta]$ on $\rho_{[\alpha,\beta]}$ simultaneously. We narrow the size of samples by mod $(N, k) = 0$ while reserving the first and the last point of the original data to hold the length of the range of $[\alpha, \beta]$. Then, we calculate all the range moments, denoted by $\rho_{[\alpha,\beta]}$.

Figures 24 and 25 show the influence of size of samples on $\rho_{[0.1]}$ and its truncated $\rho_{[0.02,0.098]}$. Figures 26 and 27 show the influence of size of samples on $\rho_{[0.02]}$ and its truncated $\rho_{[0.02,0.01]}$. Figures 28 and 29 show the influence of size of samples on $\rho_{[0.9,1]}$ and its truncated $\rho_{[0.9,0.98]}$.

From these figures, we can conclude that (1) the distorted and interpolation method (DI) is better than the traditional one (NSO), and they are more stable while the size of samples is decreasing; (2) the truncated versions are more stable than their nontruncated versions; (3) the length of range of strikes has influence on moments, especially for tail moments; and (4) the effect of extraordinary value points at the extreme tail still exists, though it is smaller. To offset this effect, the extraordinary value points must be deleted before calculations.

Secondly, we compute $(S/E[S])^\theta$ with $\theta = 3$. To be convenient, we denote $\rho_{[\alpha,\beta]}^3 = \sqrt{\mathbb{E}[(S/E[S])^3 - 1]}1_{k \in [K_s,K_f]}$. We compute all $\rho_{[\alpha,\beta]}$ with different situations for all the 11 maturities and find that the results are similar to $\rho_{[\alpha,\beta]}$. The only difference is that the results of $\rho_{[\alpha,\beta]}$ are more sensitive to extraordinary value points because when $\theta = 3$, the weights
are not constant anymore; it depends on the strike $K$. Here, we do not show the details.

4. Conclusions

From the discussions, including the tables and figures, we can conclude that

1. The distorted and interpolation method (DI) is better than the traditional one (NSO) because they are more stable when the size of samples decreases through different sparse samples from $k = 1$ to $k = 6$. This result is compatible with our theoretical analysis. As the foundation of model-free estimation, precise option price is important, and distorted lognormal distribution makes an important contribution to it. Furthermore, we interpolate the strikes of market trading options and then increase the available data for model estimation in a reasonable way, so the model error is reduced and more stable.

2. The method to compute the moments with stochastic indicator is apparently different from that with nonstochastic indicator; the risks of this kind (with $K_e < E[S] < K_g$) can be overvalued by fixed indicator methods (NSO and NSDI), relative to the stochastic indicator method (SDI), and the risks of this kind (with $K_e < K_g < E[S]$) can be undervalued by fixed indicator methods (NSO and NSDI), relative to the stochastic indicator method (SDI). That is to say, the moments calculated by fixed indicator will always be underestimated or overestimated, and the stochastic indicator method is more appropriate.

3. The truncated version is more stable than nontruncated version. Truncation means that there is a market without serious extreme situations, so we can get stable and smooth computation result easily.

4. The length of range of strikes has influence on moments, especially for tail moments. As for the left tail risk and the right tail risk, we can get more stable results with the narrowing of range, which is consistent with truncated cases.

5. The effect of extraordinary value points at the extreme tail still exists, though it is smaller. It is evident that sparsity of data in extraordinary area will enlarge the negative effect of these extraordinary points. So, if we want to offset this effect, the extraordinary value points must be deleted before calculations. So, if we want to get a better calculation of risk measure, we must (1) have more samples in the real market; (2) have the length of range of strikes as long as possible; (3) delete the extraordinary value points before calculations; (4) use range moments by discarding some of strikes at the tail (in fact, if we have deleted the extraordinary value points before calculations, range moments are not required); (5) use a stochastic indicator to calculate the range moments of risk rather than a fixed one; and (6) use the distorted and interpolation method to estimate option prices of all the strikes between the range of $[K_{min}, K_{max}]$.

Our future work is to explore a better way to fit the distorted functions $m(\lambda)$ and $q(\lambda)$ well so that we can extend the
range of strikes to $(0, +\infty)$ and can compute the risk measures more precisely and then do more works based on this.

**Data Availability**

The data used to support the findings of this study are deposited at the Reuters database.

**Conflicts of Interest**

The authors declare that they have no conflicts of interest regarding the publication of this paper.

**Acknowledgments**

This study was supported by the Humanities and Social Science Planning Fund from Ministry of Education (16YJAZH078), the Fundamental Research Funds for the Central Universities of China (grant nos. CCNU19TS062, CCNU19A06043, and CCNU19TD006), and the Raising Initial Capital for High-Level Talents of Central China Normal University (3010190001). The authors also greatly appreciate Prof. Carole Bernard, from Grenoble Ecole de Management, France, for her useful discussions and comments.

**References**

[1] R. E. Whaley, “Derivatives on market volatility: hedging tools long overdue,” *The Journal of Derivatives*, vol. 1, no. 1, pp. 71–84, 1993.

[2] G. J. Jiang and Y. S. Tian, “The model-free implied volatility and its information content,” *Review of Financial Studies*, vol. 18, no. 4, pp. 1305–1342, 2005.

[3] O. Bondarenko, “Variance trading and market price of variance risk,” *Journal of Econometrics*, vol. 180, no. 1, pp. 81–97, 2014.

[4] T. Bollerslev, G. Tauchen, and H. Zhou, “Expected stock returns and variance risk premia,” *Review of Financial Studies*, vol. 22, no. 11, pp. 4463–4492, 2009.

[5] P. Carr and L. Wu, “Variance risk premiums,” *The Review of Financial Studies*, vol. 22, no. 3, pp. 1311–1341, 2008.

[6] V. Todorov, “Variance risk-premium dynamics: the role of jumps,” *Review of Financial Studies*, vol. 23, no. 1, pp. 345–383, 2010.

[7] V. Todorov and T. Bollerslev, “Jumps and betas: a new framework for disentangling and estimating systematic risks,” *Journal of Econometrics*, vol. 157, no. 2, pp. 220–235, 2010.

[8] V. Todorov and G. Tauchen, “Activity signature functions for high-frequency data analysis,” *Journal of Econometrics*, vol. 154, no. 2, pp. 125–138, 2010.

[9] G. J. Jiang and Y. S. Tian, “Extracting model-free volatility from option prices: an examination of the VIX index,” *The Journal of Derivatives*, vol. 14, no. 3, pp. 35–60, 2007.

[10] T. G. Andersen, O. Bondarenko, and M. T. Gonzalez-Perez, “Uncovering novel features of equity-index return dynamics via corridor implied volatility,” Available at SSRN 1787528, 2012.

[11] M. T. Gonzalez-Perez, “Model-free volatility indexes in the financial literature: a review,” *International Review of Economics & Finance*, vol. 40, pp. 141–159, 2015.