Research on the Distribution of Asset Prices Based on the Principle of Maximum Entropy

Hongying Jin *, Meiyun Liu, Kuo Yang, Yong Song, Jianlin Mo, Jiayong Chen

*439988699@qq.com; Liu Meiyun, 1789083918@qq.com; Yang Kuo, 65968631@qq.com;
Song Yong, 115207787@qq.com; Mo Jianlin, 86003830@qq.com; Chen Jiayong, 3051741145@qq.com

Institute of Applied Physics, School of Physics and Electronic and Electrical Engineering, Aba Normal University, Aba Sichuan 623002, China

Abstract: The maximum entropy principle is a fundamental principle in statistical physics. The principle is used to study the distribution law of asset prices in financial markets, and it is concluded that the maximum entropy distribution of asset prices follows a lognormal distribution. The daily data of four typical stock price indices, namely the Shanghai Securities Composite Index, Shenzhen Component Index, Hang Seng Index, and Shanghai Shenzhen 300 Index, are selected as empirical data, and MATLAB is used for programming. By comparing the actual frequency distribution curve with the theoretical frequency distribution curve, the logarithmic normal distribution law of asset prices is verified. It is shown that the logarithmic normal distribution can be used to fit the actual asset price distribution, and the maximum entropy principle is also suitable for studying the distribution law of asset prices.

Keywords: econophysics, maximum entropy principle, price, log-normal distribution

1. INTRODUCTION

Financial physics is a rising interdisciplinary field that primarily utilizes concepts, methods, and theories from physics (such as statistical physics, theoretical physics, nonlinear physics, complex system theory, etc.) to study data generated in financial markets and quantitatively analyze phenomena and patterns in financial markets [1-3].

The maximum entropy principle is a fundamental principle in statistical physics, which states that the information entropy (generalized entropy) of a macroscopic system tends to a constrained maximum under a set of constraint conditions [4,5]. According to this principle, for a system, if we choose appropriate constraint conditions and calculate the maximum value of its information entropy under the selected constraint conditions, in principle, we can calculate the distribution of the system [6]. The maximum entropy principle has been widely applied in many fields [7-10]. Zeng Xiaohua and Yuan Zhiping constructed a financial system risk analysis model based on the maximum entropy principle. Empirical results show that the model is an effective risk analysis method [11]. In modern financial markets, asset pricing and price distribution are very important issues [12-14]. This article uses the maximum entropy principle to study the distribution of asset prices in the financial market, and finds that its logarithm follows a normal distribution, indicating that the price of financial assets follows a logarithmic normal distribution. By using Matlab to plot theoretical frequency curves and frequency distribution

ICEMBDA 2023, October 27-29, Tianjin, People's Republic of China

DOI 10.4108/eai.27-10-2023.2342002

curves of actual data, the conclusion that asset prices follow a lognormal distribution was verified by comparison between the two curves.

2. RESEARCH ON THE DISTRIBUTION OF ASSET PRICES USING THE MAXIMUM ENTROPY PRINCIPLE

The financial market is a complex system with great uncertainty. For the convenience of comparative analysis, the asset prices in the financial market are dimensionless: *x* is defined as the ratio of the price of a certain asset to the average price of that asset, and *x* is a dimensionless quantity with a value range of 0~∞, while its natural logarithm ln *x* has a value range of -∞~+∞. Let *y*=ln *x*, if y follows a normal distribution, it indicates that asset prices follow a lognormal distribution. Taking y as a random variable, the corresponding information entropy is defined as:

$$S = -\int_{-\infty}^{+\infty} f(\mathrm{y})\ln f(\mathrm{y})\mathrm{d}(\mathrm{y}) \tag{1}$$

According to the probability normalization condition, it can be concluded that:

$$\int_{-\infty}^{+\infty} f(y)\mathrm{d}y = 1 \tag{2}$$

The selection of constraint conditions is:

$$\int_{-\infty}^{+\infty} yf(y)\mathrm{d}y = u \tag{3}$$

$$\int_{-\infty}^{+\infty} y^2 f(y)\mathrm{d}y = w \tag{4}$$

In the two equations, *u* represents the average value of the random variable *y*, and *w* represents the average value of *y* squared.

By taking the maximum value of the information entropy *S* under the constraint conditions (2), (3), and (4), the distribution of the random variable *y* (the logarithm of dimensionless asset prices) can be obtained. This is a constrained extremum problem that can be solved using the Lagrange multiplier method. Following the general steps of using the Lagrangian multiplier method to obtain the constrained extremum, the functional *F* is introduced:

$$\begin{aligned} F = &-\int_{-\infty}^{+\infty} f(y)\ln f(y)\mathrm{d}y - \alpha\left(\int_{-\infty}^{+\infty} f(y)\mathrm{d}y - 1\right) \\ &- \beta\left(\int_{0}^{\infty} yf(y)\mathrm{d}y - u\right) - \gamma\left(\int_{0}^{\infty} y^2 f(y)\mathrm{d}y - w\right) \end{aligned} \tag{5}$$

In the equation, *α* is the Lagrange multiplier introduced by the constraint condition equation (2), and *β* is the Lagrange multiplier introduced by the constraint condition equation (3), *γ* is a Lagrange multiplier introduced by the constraint condition (4).

By $\delta F(f(y)) = 0$, the expression for the probability distribution function *f*(*y*) that maximizes the information entropy *S* under constraint conditions (2), (3), and (4) can be calculated as follows:

$$f(y) = e^{-1-\alpha-\beta y-\gamma y^2} \tag{6}$$

Substituting equation (6) into equation (2) yields:

$$\int_{-\infty}^{+\infty} e^{-1-\alpha-\beta y-\gamma y^2}\,\mathrm{d}y = 1 \tag{7}$$

From equation (7), it can be concluded that:

$$e^{1+\alpha} = \frac{\sqrt{\pi}}{\sqrt{\gamma}} e^{\frac{\beta^2}{4\gamma}} \tag{8}$$

Substituting equation (8) into equation (6) yields:

$$f(y) = \frac{\sqrt{\gamma}}{\sqrt{\pi}} e^{-\gamma\left(y+\frac{\beta}{2\gamma}\right)^2} \tag{9}$$

By substituting equation (9) into equation (3), it can be calculated that:

$$-\frac{\beta}{2\gamma} = u \tag{10}$$

The density function is obtained by substituting equation (10) back into equation (9), that is:

$$f(y) = \frac{\sqrt{\gamma}}{\sqrt{\pi}} e^{-\gamma(y-u)^2} \tag{11}$$

By substituting equation (11) into equation (4), it can be calculated that:

$$\gamma = \frac{1}{2(w-u^2)} \tag{12}$$

Substituting equation (12) into equation (11) yields:

$$f(y) = \frac{1}{\sqrt{2\pi(w-u^2)}} e^{-\frac{(y-u)^2}{2(w-u^2)}} \tag{13}$$

In this way, we obtain the distribution density function expression of the random variable *y*, which shows that *y* follows a normal distribution, with an expected value of *u* and a variance of w-u^2. According to the definition of a logarithmic normal distribution, this indicates that asset price *x* follows a logarithmic normal distribution.

3. COMPARATIVE ANALYSIS OF THEORETICAL DISTRIBUTION AND ACTUAL DATA DISTRIBUTION

We first select the daily closing prices of the Shanghai Securities Composite Index from January 4, 2010 to September 29, 2023 as empirical data. MATLAB is used for programming to calculate. We obtain the average value of the closing price data, and then divide the daily closing price by the average value of the closing price to obtain the dimensionless relative price *x*. We take the natural logarithm of the obtained dimensionless relative price, and denote as *y*, i.e. $y=\ln x$, and then calculate the average value *u* and the mean-square value *w* of *y*. Next, y_{min} represents the minimum value of *y* and y_{max} represents the maximum value of *y*. The interval of $[y_{min}, y_{max}]$ is partitioned into *n* equidistant subintervals, and the number of times which empirical data falls within each subinterval is counted as the actual frequency of price distribution within each subinterval. The calculation method for the theoretical frequency of price distribution within each subinterval is as follows:

The k-th subinterval is denoted as $[y_{1k}, y_{2k}]$, where the left endpoint of the subinterval is y_{1k} and the right endpoint is y_{2k}. According to the Simpson formula for numerical integration [15], the definite integral within the subinterval is

$$\int_{y_{1k}}^{y_{2k}} f(y)dy = \frac{y_{2k}-y_{1k}}{6}\left(f(y_{1k}) + 4f\left(\frac{y_{1k}+y_{2k}}{2}\right) + f(y_{2k})\right) \tag{14}$$

Based on this, the probability of y value falling within the subinterval can be calculated, and the theoretical probability is multiplied by the total number of data to obtain the theoretical frequency of y value falling within the subinterval. The theoretical frequency within each subinterval is taken as the ordinate, and the median of each subinterval ($\frac{y_{1k}+y_{2k}}{2}$) is taken as the abscissa. The plot function is used to draw a curve, which shows the variation of the theoretical frequency with the y-value. At the same time, based on the actual frequency calculated, a curve of its variation with y-value is drawn to obtain the curve of the actual frequency changing with y-value. According to the above steps, using the Shanghai Composite Index data, the frequency distribution diagrams are shown in Figure 1.

From Figure 1, it can be seen that it is generally feasible to use the theoretical frequency curve to fit the actual frequency curve for the Shanghai Composite Index, and the trends of the two curves are basically consistent.

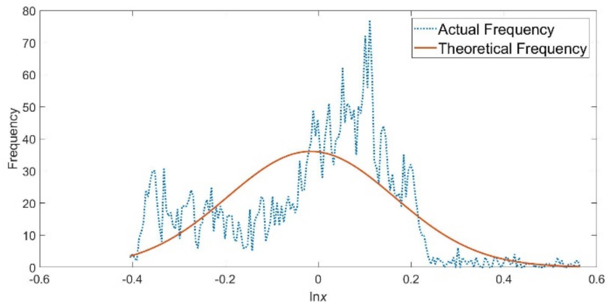

Figure 1. The frequency distribution of random variable y (based on the Shanghai Composite Index)

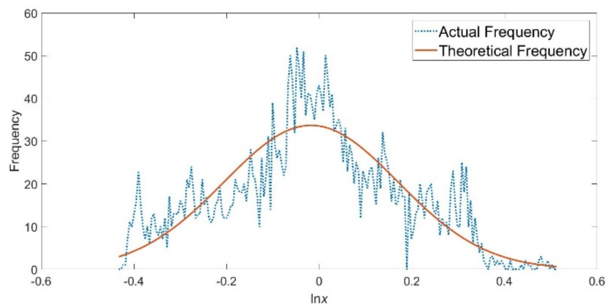

Figure 2. The frequency distribution of random variable y (based on Shenzhen Component Index)

In order to use more data for verification, we also select the daily closing prices of Shenzhen Component Index, Hang Seng Index, and Shanghai Shenzhen 300 Index from January 4, 2010 to September 29, 2023. Using the same method, we obtain the distribution curves of theoretical frequency and actual frequency, as shown in Figures 2-4.

From Figures 2-4, it can be seen that the theoretical and actual frequency distribution curve contours are basically consistent, and the two are in good agreement. It is generally feasible to use the theoretical frequency curve to fit the actual frequency distribution.

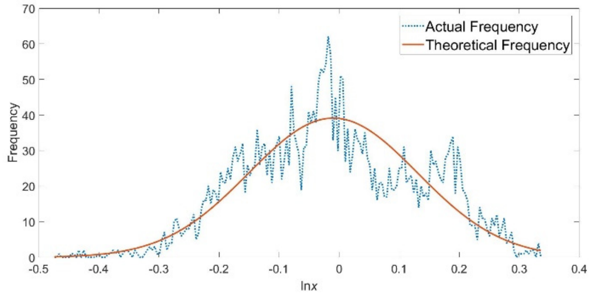

Figure 3. The frequency distribution of random variable y (based on the Hang Seng Index)

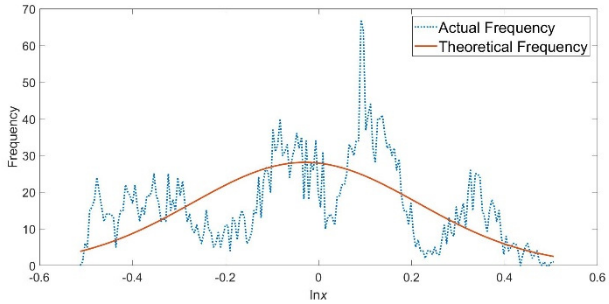

Figure 4. The frequency distribution of random variable y (based on the Shanghai Shenzhen 300 Index)

4. CONCLUSIONS

This article uses the maximum entropy principle to study the distribution law of the logarithm of dimensionless asset prices. The distribution function is derived, and it is found that is a normal distribution, indicating that asset prices follow a logarithmic normal distribution. In order to verify the consistency between the logarithmic normal distribution law and actual data, we select four typical daily data of financial market price indices, namely the Shanghai Securities Composite Index, Shenzhen Composite Index, Hang Seng Index, and Shanghai Shenzhen 300 Index, from January 4, 2010 to September 29, 2023. We compare the theoretical frequency distribution curves with the actual frequency distribution curves, and find that the two are in good agreement, which indicates it is more appropriate to fit the actual asset price distribution with a logarithmic normal distribution.

Fund Projects: Aba Normal University Campus Level Special Project; Project Number: AS-RCZX2023-02.

REFERENCES

[1] Zhao, L. F., "On the Cross-correlation of Financial Time Series," Central China Normal University, Wuhan (2018).

[2] Gao, Y. C., "Complex System Approach to Analysis and Modelling of Financial Market Dynamics," University of Science and Technology of China, Hefei (2013).

[3] Wang, Y., "Complexity Analysis of Time Series Based on Information Entropy Theory," Beijing Jiaotong University, Beijing (2021).

[4] Jin, H. Y., Luo, L. F. and Zhang, L.R., "Using estimative reaction free energy to predict splice sites and their flanking competitors," Gene, 424(1-2), 115-120 (2008).

[5] Jin, H. Y., "Using maximum entropy principle to study polarization of dielectric systems," Complex Systems and Complexity Science, 10(1), 83-88 (2013).

[6] Pressé, S., Ghosh, K., Lee, J. and Dill, K. A., "Principle of maximum entropy and maximum caliber in statistical physics," Rev. Mod. Phys., 85, 1115-1141 (2013).

[7] Golan, A. and Harte, J., "Information theory: A foundation for complexity science," Proceedings of the National Academy of Sciences of the United States of America, 119(33), e2119089119 (2022).

[8] Peng, Y.X. and Zhang, H., "Calculation of raindrop spectrum of flood discharge atomization based on maximum entropy principle," Journal of China Institute of Water Resources and Hydropower Research, 21(2), 148-156 (2023).

[9] Banavar, J. R., Maritan, A. and Volkov, I., "Applications of the principle of maximum entropy: from physics to ecology," Journal of Physics: Condensed Matter, 22, 063101 (2010).

[10] Martino, A. D. and Martino, D. D., "An introduction to the maximum entropy approach and its application to inference problems in biology," Heliyon, 4, e00596 (2018).

[11] Zeng, X. H. and Yuan, C. P., "Measurement and empirical study of financial system risk based on maximum entropy principle," Journal of DongGuan University of Technology, 26(3), 94-98 (2019).

[12] Sergey, I., "Transaction costs, frequent trading, and stock prices," Journal of Financial Markets, 64, 100775 (2023).

[13] Chen, X., Hyuk, C. J., Kasper, L. and Duane, J. S., "Price impact in Nash equilibria," Finance and Stochastics, 27(2), 305-340 (2023).

[14] Elena, M., Francisco, P. and Enrique, S., "Empirical evaluation of overspecified asset pricing models," Journal of Financial Economics, 147(2), 338-351 (2023).

[15] Peng, F. L., "Basics of Computational Physics," Higher Education Press, Beijing (2010).