

Research on Normality of Shanghai Composite Index

Caifeng Li

School of Mathematics and Statistics, Hechi University, Yizhou Guangxi
Email: licaif1103@163.com

Received: Mar. 7th, 2020; accepted: Mar. 23rd, 2020; published: Mar. 30th, 2020

Abstract

This paper studies the normality of Shanghai composite index because options pricing and other financial studies are often based on the normal distribution of the logarithm return of stock price. It is found that for the same group of Shanghai index, its return and logarithm return have the same results about normal distribution, and the rate of return obeys the normal distribution, which is more consistent with the actual index than the logarithmic return. The research results of this paper expand the conclusions of the existing literature, and provide the theoretical basis for the research of financial market.

Keywords

Shanghai Composite Index, Rate of Return, Logarithmic Return, Normal Distribution

上证指数的正态性研究

李彩凤

河池学院, 数学与统计学院, 广西 宜州
Email: licaif1103@163.com

收稿日期: 2020年3月7日; 录用日期: 2020年3月23日; 发布日期: 2020年3月30日

摘要

期权定价等金融研究常以股价的对数收益服从正态分布为基础, 而上证指数是A股股指的代表, 因此本文研究上证指数的正态性。通过研究发现, 对于同一组上证指数, 它的收益率与对数收益是同时服从或

同时不服从正态分布；并且上证指数的收益率服从正态分布比上证指数的对数收益服从正态分布更吻合实际指数。本文的研究结果扩充了已有文献的结论，为金融市场研究提供理论基础。

关键词

上证指数，收益率，对数收益，正态分布

Copyright © 2020 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

本模板股价变化的正态分布模型是很多金融理论研究的基础，也是股票市场研究基本内容。现有的我国股价正态性研究只有两个结论。一个结论是股价 p_t 用随机微分方程 $dp_t = \mu p_t dt + \sigma p_t dB_t$ 描述 [1]-[7]，这是由于著名的 Black-Scholes 期权定价公式以及金融应用中，股价常用此随机微分方程描述。股价用此微分方程描述则股价的对数收益服从正态分布，这在下文论述。另一个结论是股价收益率不服从正态分布 [8] [9] [10]。上海证券交易所综合股价指数(简称上证指数)当之无愧是我国 A 股股指的代表，因此本文以上证指数作为研究对象，研究上证指数对数收益是否服从正态分布以及收益率是否服从正态分布，在本文中称之为研究上证指数的正态性。通过研究，发现上证指数的正态性不完全如已有文献所描述。

2. 相关基本理论

定义 1 [11]: 若随机过程 $\{p_t, t \geq 0\}$ 满足:

- 1) $\forall 0 \leq k < t, P_t - P_k \sim N(0, \sigma^2)$,
- 2) $\forall 0 \leq k < t, P_t - P_k$ 独立于过去状态 $P_u (0 \leq u \leq k)$,

则称随机过程 $\{p_t, t \geq 0\}$ 为布朗运动。

如果 $\sigma = 1$ ，称为标准布朗运动，记为 B_t ，有 $dB_t = \varepsilon \sqrt{dt}$ ，其中 $\varepsilon \sim N(0, 1)$ 。

定义 2: 设 p_t 是 t 时的股指，称 $\frac{P_t - P_{t-\Delta t}}{P_t}$ 为股指在 t 时 Δt 时间段上的收益率，记为 r_t ，即 $r_t = \frac{P_t - P_{t-\Delta t}}{P_t}$ ；

称 $\ln \frac{P_t}{P_{t-\Delta t}} = \ln p_t - \ln p_{t-\Delta t} = \Delta p_t$ 为股指在 t 时 Δt 时间段上的对数收益，记为 l_t ，即 $l_t = \ln \frac{P_t}{P_{t-\Delta t}}$ 。

定理 1: 若股指 P_t 满足随机微分方程 $dP_t = \mu P_t dt + \sigma P_t dB_t$ ，其中 B_t 是标准的布朗运动，则股指 P_t 的对数收益服从正态分布。

证明: 考虑随机过程 $P_t = P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t}$ 。对于 $f(t, x) = P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma x}$ ，有 $\frac{\partial f}{\partial t} = \mu P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t}$ ，

$\frac{\partial f}{\partial x} = \sigma P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t}$ ， $\frac{\partial^2 f}{\partial x^2} = \sigma^2 P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t}$ ，由 ITO 公式得

$dP_t = \left(\mu - \frac{\sigma^2}{2}\right) P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t} dt + \sigma P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t} dB_t + \frac{1}{2} \sigma^2 P_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma B_t} (dB_t)^2$ 。因为 $(dB_t)^2 = dt$ ，所以

$$dP_t = \left(\mu - \frac{\sigma^2}{2} \right) P_0 e^{\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B_t} dt + \sigma P_0 e^{\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B_t} dB_t + \frac{1}{2} \sigma^2 P_0 e^{\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B_t} dt$$

也即随机过程 $P_t = P_0 e^{\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B_t}$ 是

$$= \mu P_0 e^{\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B_t} dt + \sigma P_0 e^{\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B_t} dB_t = \mu P_t dt + \sigma P_t dB_t$$

随机微分方程 $dP_t = \mu P_t dt + \sigma P_t dB_t$ 的解。

在任意时间段 $[t, t + \Delta t]$ 上, 由 $P_t = P_0 e^{\left(\mu - \frac{\sigma^2}{2} \right) t + \sigma B_t}$ 得 $\Delta \ln(P_t) = \left(\mu - \frac{\sigma^2}{2} \right) \cdot \Delta t + \sigma \cdot \Delta B_t$, 又因 $dB_t = \varepsilon \sqrt{dt}$,

$\varepsilon \sim N(0,1)$, 所以 P_t 在 $[t, t + \Delta t]$ 上有 $\Delta \ln(P_t) \sim N\left(\left(\mu - \frac{\sigma^2}{2} \right) \cdot \Delta t, \sigma^2 \cdot \Delta t \right)$ 。

即股指 P_t 的对数收益服从正态分布。证毕。

3. 上证指数正态性统计研究

本文通过研究发现, 时间长度一年以上的和时间长度为几个月的上证指数的正态性不同, 因此本文按时间段的长度对上证指数正态性进行研究。由于文章篇幅有限, 只选取研究中的一部分统计分析图及表格到文中, 其余的用文字说明。文中所说的时间长度均指自然日的而非交易日的时间长度, 所说的上证指数均指上证指数每个交易日的收盘价。

3.1. 一年以上时间段的上证指数正态性统计研究

上海证券交易所于 1990 年 12 月 19 日正式营业, 现 A 股市场已建立约三十年。为了考察 A 股市场建立后, 时间长度一年以上不同时间长度的上证指数正态性, 在这部分内容随机取了连续十年的数据三组、连续五年的数据五组、连续两年的数据五组、连续一年的数据五组, 分别对这二十组数据进行上证指数正态性研究。

由于篇幅原因, 只分别选取了三组不同时间段数据的偏度峰度描述, 并且略去了正态检验的具体结果。由图 1~4 易得出峰度系数的绝对值都远远大于 1.96, 峰度都是高峰, 所统计的数据均不服从正态分布。关于正态检验结果, 除了时间长度为一年的数据显著性为 0.001, 其它的显著性均为 0.000, 就是由正态检验同样得出时间长度大于一年的上证指数收益率和对数收益均不服从正态分布。

描述性统计资料

	N	偏斜度		峰度	
	统计资料	统计资料	标准错误	统计资料	标准错误
收益率1	2414	2.556	0.050	43.681	0.100
对数收益1	2414	5.499	0.050	111.830	0.100
收益率2	2414	-0.231	0.050	3.867	0.100
对数收益2	2414	-0.079	0.050	3.909	0.100
收益率3	2414	-1.042	0.050	6.345	0.100
对数收益3	2414	-0.896	0.050	5.770	0.100
有效的 N (listwise)	2414				

Figure 1. Description of skewness kurtosis of return rate and logarithm return for ten years

图 1. 三组连续十年的上证指数收益率和对数收益的偏度峰度描述

描述性统计资料

	N	偏斜度		峰度	
	统计资料	统计资料	标准错误	统计资料	标准错误
收益率1	1215	0.113	0.070	16.725	0.140
对数收益1	1215	0.822	0.070	21.196	0.140
收益率2	1215	-0.472	0.070	2.476	0.140
对数收益2	1215	-0.342	0.070	2.414	0.140
收益率3	1215	-1.428	0.070	8.085	0.140
对数收益3	1215	-1.252	0.070	7.381	0.140
有效的 N (listwise)	1215				

Figure 2. Description of skewness kurtosis of return rate and logarithm return for three years
图 2. 三组连续五年的上证指数收益率和对数收益的偏度峰度描述

描述性统计资料

	N	偏斜度		峰度	
	统计资料	统计资料	标准错误	统计资料	标准错误
收益率1	483	0.532	0.111	1.226	0.222
对数收益1	483	0.586	0.111	1.329	0.222
收益率2	483	-0.181	0.111	1.195	0.222
对数收益2	483	-0.063	0.111	1.262	0.222
收益率3	483	-0.400	0.111	3.300	0.222
对数收益3	483	-0.316	0.111	3.086	0.222
有效的 N (listwise)	483				

Figure 3. Description of skewness kurtosis of return rate and logarithm return for two years
图 3. 三组连续两年的上证指数收益率和对数收益的偏度峰度描述

描述性统计资料

	N	偏斜度		峰度	
	统计资料	统计资料	标准错误	统计资料	标准错误
收益率1	238	-0.456	0.158	2.598	0.314
对数收益1	238	-0.382	0.158	2.422	0.314
收益率2	244	-0.340	0.156	1.586	0.310
对数收益2	244	-0.312	0.156	1.543	0.310
收益率3	237	0.870	0.158	6.919	0.315
对数收益3	237	1.059	0.158	7.522	0.315
有效的 N (listwise)	237				

Figure 4. Description of skewness kurtosis of return rate and logarithm return for one year
图 4. 三组连续一年的上证指数收益率和对数收益的偏度峰度描述

3.2. 时间长度为几个月的上证指数正态性统计研究

为了考察 A 股市场建立后, 时间长度为几个月的上证指数正态性, 在这部分内容随机取了连续八个月的数据五组、连续六七个月的数据五组、连续五六个月的数据五组、连续三四个月的数据五组、连续两三个月的上证指数数据五组, 连续一个月的数据五组, 分别对这 30 组数据进行正态性研究。

在随机选取的 30 组数据中, 发现时间长度为两三个月的上证指数的收益率及对数收益, 随着 A 股市场的发展逐渐由不服从正态分布变成基本上近似服从正态分布。而其它时间长度的上证指数收益率及对数收益并没有随着时间的发展逐渐变成服从正态分布这样的规律。因此, 这部分主要研究时间长度为两三个月的上证指数正态性。为了方便统计与文字描述, 这部分的两三个月的上证指数全部取连续六十几个交易日的上证指数。

图 5~8 是 1990 年到 2003 年间随机取的五组连续 66 个交易日的收益率及对数收益直方图。通过直方图易看出, 这些收益率及对数收益都不服从正态分布, 并且同一时间段的上证指数的收益率与对数收益的直方图非常相像。对这些数据的收益率及对数收益进行正态检验, 基本上都不服从正态分布。

接下来用直方图、Q-Q 图、正态检验对 2004 年至 2019 年期间的的时间长度为两三月的上证指数正态性进行研究。同一时间段的上证指数的收益率及对数收益的直方图也是非常相像, 且分布曲线是正态曲线。由 Q-Q 图可看出, 这些数据的收益率及对数收益基本上服从正态分布。由于篇幅原因文中没给出直方图、Q-Q 图的图形, 只给出相应的正态检验结果。

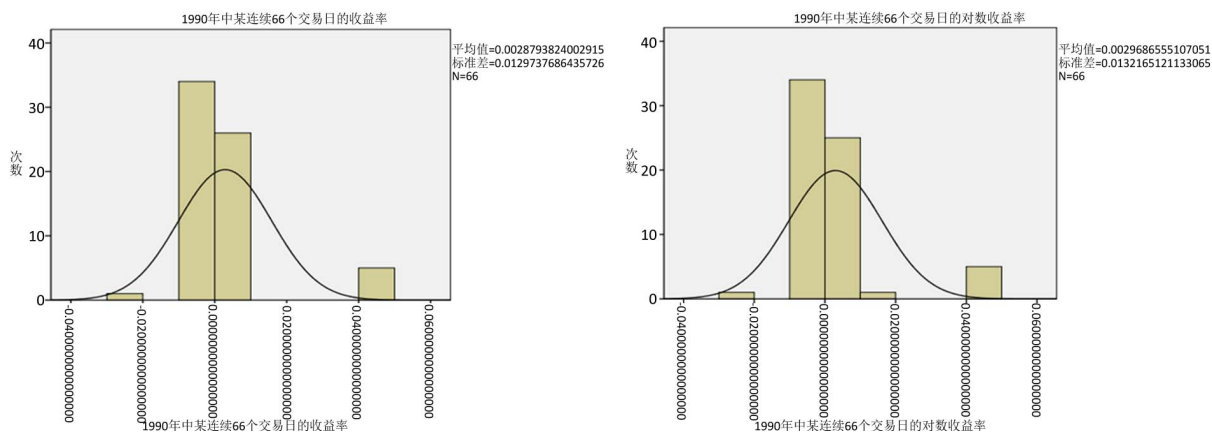


Figure 5. Histogram of the return and logarithm return of Shanghai stock index in the same period
图 5. 同一时间段上证指数的数收益率和对数收益的直方图

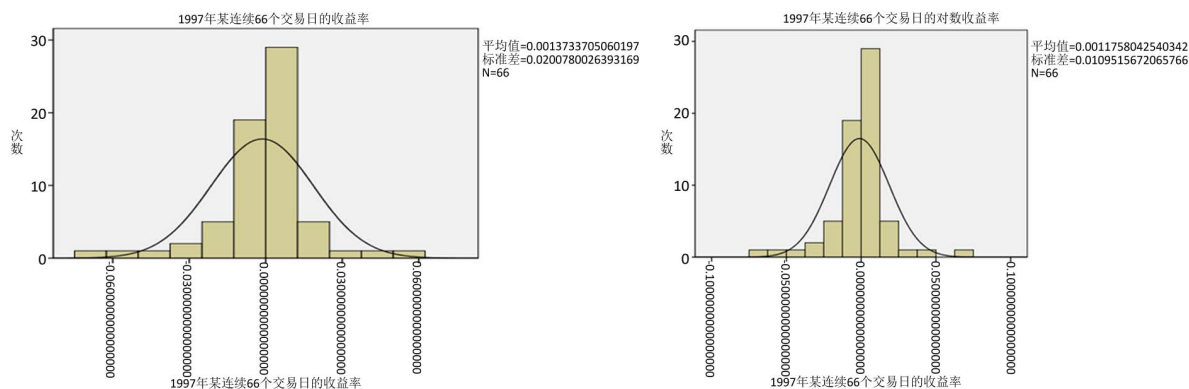


Figure 6. Histogram of the return and logarithm return of Shanghai stock index in the same period
图 6. 同一时间段上证指数的数收益率和对数收益的直方图

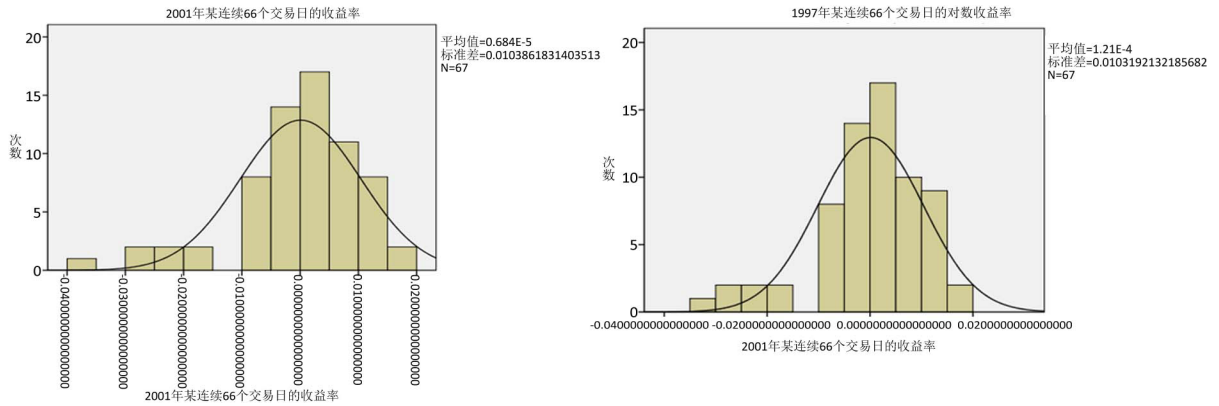


Figure 7. Histogram of the return and logarithm return of Shanghai stock index in the same period
图 7. 同一时间段上证指数的数收益率和对数收益的直方图

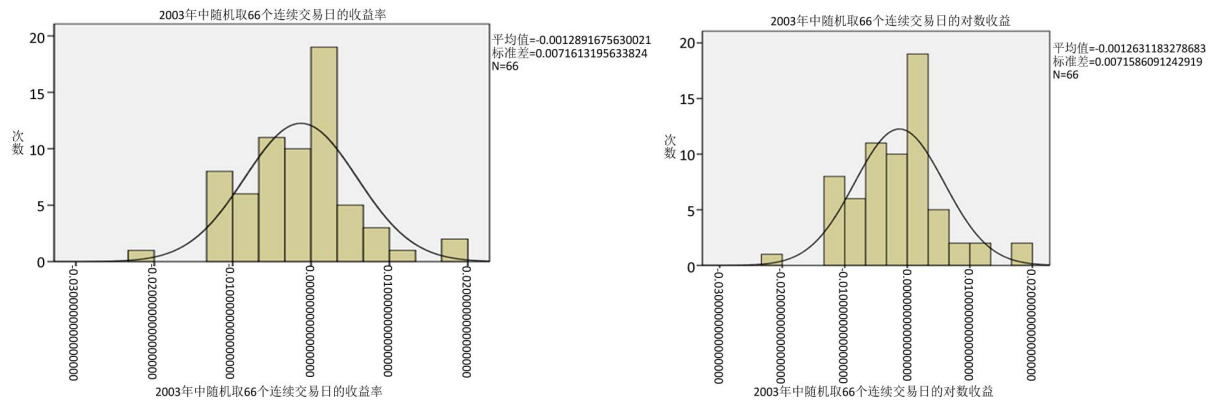


Figure 8. Histogram of the return and logarithm return of Shanghai stock index in the same period
图 8. 同一时间段上证指数的数收益率和对数收益的直方图

单一样本 Kolmogorov-Smirnov 检定

		03年1月到4月收益 率	03年1月到4月对数 收益	07年7月到07年10月 收益率	07年7月到07年10月 对数收益
N		69	69	69	69
常态参数 ^{a,b}	平均数	0.000822842257903	0.000902354810065	0.004897894844252	0.005092258130770
	标准偏差	0.012603941239256	0.012707372030074	0.019227672057601	0.019193968828807
最极端差异	绝对	0.100	0.103	0.120	0.116
	正	0.100	0.103	0.075	0.073
	负	-0.077	-0.077	-0.120	-0.116
测试统计资料		0.100	0.103	0.120	0.116
渐近显著性 (双尾)		0.087 ^c	0.069 ^c	0.016 ^c	0.023 ^c

单一样本 Kolmogorov-Smirnov 检定

		08年6月到9月收益 率	08年6月到9月对数 收益	13年1月到5月收益 率	13年1月到5月对数 收益率
N		69	69	69	69
常态参数 ^{a,b}	平均数	-0.006685292020251	-0.006318716838998	-0.000319204201848	-0.000246988667724
	标准偏差	0.026547772164121	0.026468867467627	0.012127577321850	0.012091093457363
最极端差异	绝对	0.077	0.081	0.088	0.089
	正	0.077	0.081	0.088	0.089
	负	-0.056	-0.051	-0.085	-0.083
测试统计资料		0.077	0.081	0.088	0.089
渐近显著性 (双尾)		0.200 ^{c,d}	0.200 ^{c,d}	0.200 ^{c,d}	0.200 ^{c,d}

单一样本 Kolmogorov-Smirnov 检定					
		18年9月到12月收益	18年9月到12月对数	19年1月到5月收益	19年1月到5月对数
		率	收益	率	收益
N		69	69	69	69
常态参数 ^{ab}	平均数	-0.000784544601895	-0.000669128345147	0.001381490002292	0.001525958391819
	标准偏差	0.015337795014540	0.015264112237785	0.017076612739967	0.017050358132067
最极端差异	绝对	0.098	0.095	0.068	0.066
	正	0.089	0.089	0.056	0.059
	负	-0.098	-0.095	-0.068	-0.066
测试统计资料		0.098	0.095	0.068	0.066
渐进显著性 (双尾)		0.095 ^c	0.199 ^c	0.200 ^{c,d}	0.200 ^{c,d}

Figure 9. Normal test results of Shanghai stock index of six random groups for two or three consecutive months

图 9. 2003.01.04 至 2019.9.30 间随机六组连续两三个月上证指数正态检验结果

通过图 9 正态检验结果可看出：同一时间段的上证指数的收益率及对数收益检验结果基本是一致的，都是同时通过或者同时不通过，渐进显著性也比较接近；这期间两三个月时间长度的收益率及对数收益基本上服从正态分布。通过实证研究得出两个结论。一是时间长度为两三个月的上证指数的收益率及对数收益的正态性是一致的，都是同时服从正态分布或者都不服从正态分布，且直方图形状非常相像以及正态检验的渐进显著性较接近；二是时间长度为两三个月的上证指数的收益率及对数收益，由 A 股市场建立初期的不服从正态分布，随着时间的推移，到约 2003 年后逐渐变成基本上都服从正态分布。基于第二第三部分得出的结论：2003 年以后时间长度为两三个月的上证指数的收益率及对数收益基本上都服从正态分布，在第四部分进行了 2018 年 1 月 2 日到 2019 年 9 月 30 的上证指数图像模拟。

4. 图像模拟

这部分中，先假设 2018 年和 2019 年长度为两三个月的上证指数的收益率及对数收益都服从正态分

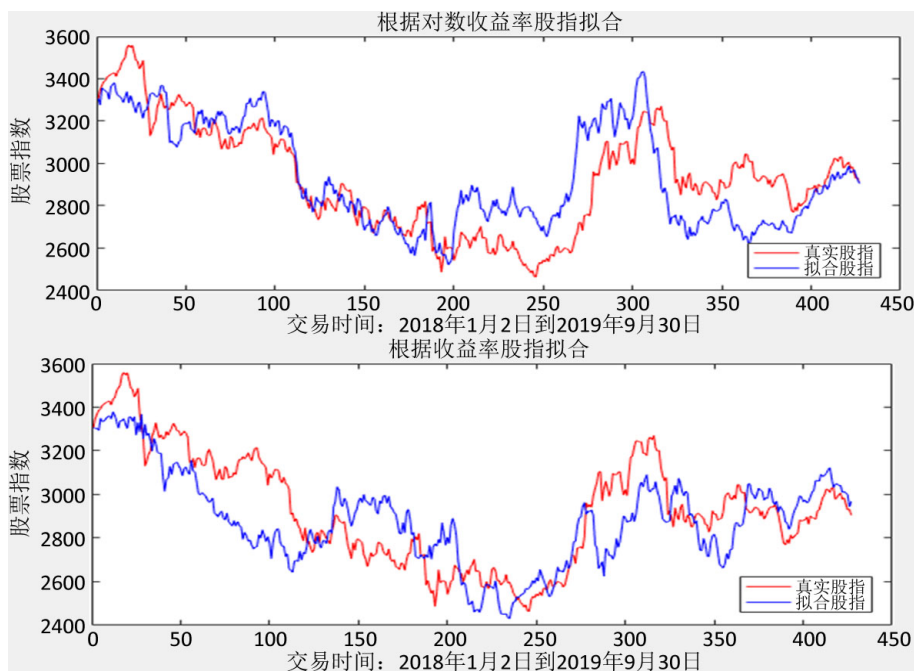


Figure 10. Simulation chart based on the normal distribution of Shanghai Stock Index Returns

图 10. 上证指数收益率服从正态分布得出的模拟图

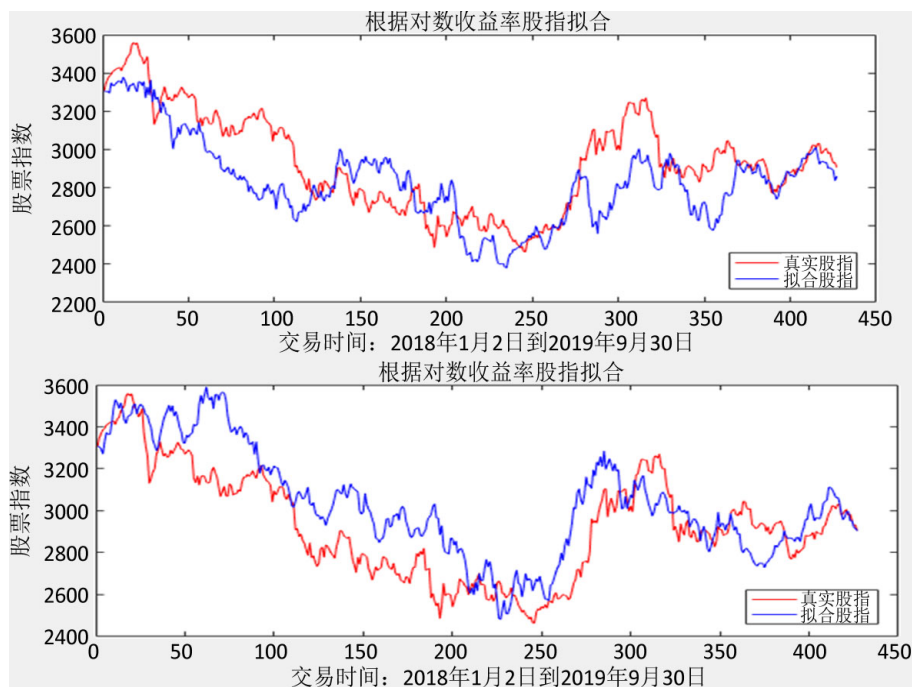


Figure 11. Simulation chart based on the normal distribution of logarithmic return of Shanghai Stock Index

图 11. 上证指数对数收益服从正态分布得出的模拟图

	N	最小值	最大值	平均数
模拟1对数收益	426	-0.040797826	0.0438105957	-0.000725398
模拟2对数收益	426	-0.040797229	0.0439105938	-0.000736892
模拟1收益率	426	-0.053708072	0.0384901773	-0.000242626
模拟2收益率	426	-0.047961784	0.0467538837	-0.000341364
原对数收益	426	-0.0575000	0.0545000	-0.000305338
原收益率	426	-0.0591000	0.053000	-0.000381640
有效的N (listwise)	426			

Figure 12. Statistical results of simulated yield and logarithmic yield

图 12. 模拟的上证指数收益率及对数收益统计结果

布, 然后由实际上证指数的收益率及对数收益分别求出每时间段正态分布的两个参数, 分别模拟 2018 年 1 月 2 日到 2019 年 9 月 30 的上证指数的曲线图, 若模拟的图形与实际上证指数曲线图偏差较小, 则假设是成立的。具体如下: 从大智慧下载 2018 年 1 月 2 日到 2019 年 9 月 30 的上证指数每个交易日的收盘价, 求出这些上证指数的收益率及对数收益; 把 426 个收益率(或对数收益)数据等分为 n 部分, 假设每一部分都服从正态分布, 求出每一部分的均值与方差 (μ_i, σ_i^2) , $i=1, 2, \dots, n$; 用 matlab 产生服从参数为 (μ_i, σ_i^2) 的正态分布的随机数, 用产生的随机数分别作为上证指数的收益率及对数收益求出对应时间的模拟上证指数; 分别由收益率及对数收益服从正态分布得出的模拟上证指数画出模拟上证指数曲线图, 并将模拟图与实际曲线图进行比较。对 n 分别取 1, 2, ... 进行模拟实验, 发现 n 取 6 或 7 时得到的模拟效果最好。也就是时间长度为两三个月时模拟效果最好, 这跟前面的实证结果一致。接着对 n 取 6 或 7 进行了十次模拟实验, 得出图 11 和图 12 是其中模拟效果较好的两次结果。

由图 10 和图 11 看出, 模拟效果还是比较理想的。由图 12 统计结果得出模拟收益率及模拟对数收益的最值与实际结果相差不大。求出这十次模拟收益率平均数的均值为 -0.000335985 与实际结果相差不大, 求出这十次模拟对数收益平均数的均值为 -0.000725409 与实际结果相差稍大。从第四部分的图像模拟进一步确定了这个结论: 时间长度为两三个月的上证指数收益率及对数收益 2003 年以后基本上都服从正态分布, 并且收益率服从正态分布比对数收益服从正态分布与实际股指更加吻合。

5. 结论

本文利用统计中能最大程度保留原始信息的峰度偏度系数、直方图和 KS 检验, 对上证指数的收益率及对数收益的正态性进行研究, 最大程度地避免了统计中造成的信息丢失。研究得出: 同一时间段的上证指数的收益率与它的对数收益, 它们的直方图非常相像、它们的峰度偏度比较接近, 并且它们的正态检验的渐进显著性接近, 它们的正态检验结果相同。上证指数的收益率与对数收益的正态性与时间段长度及时间有关: 时间长度一年以上的上证指数收益率及对数收益均不服从正态分布; 时间长度为两三个月的上证指数收益率及对数收益在 2003 年以后基本上都服从正态分布, 并且收益率服从正态分布比对数收益服从正态分布与实际股指更加吻合。

常说“多因素小影响综合成正态”, 股价差是多因素造成, 除了特殊情况存在某因素造成较大影响外, 一般情况下按道理各元素都是小影响。为了去掉量纲将股价差除以股价得到收益率。按正态分布原理来说收益率应该是“多因素小影响综合成正态”, 也就是上证指数收益率应该服从正态分布。但已有文献中实证得出股价收益率不服从正态分布, 是由于 A 股市场建立初期的不完善等因素造成, 现随着我国证券发行制度的变迁、交易制度的建设、上市公司治理机制的优化、证券市场信息披露制度的规范以及投资者结构的改善, 我国上证指数的股价正态性也有着相应的变化, 在本文中得到了实证。这为期权定价等方面金融市场研究方面提供了新的理论基础。

基金项目

国家自然科学基金项目(项目编号: 11961021)。

参考文献

- [1] 徐景昭. 上证指数关于几何布朗运动的实证分析[J]. 市场论坛, 2017(5): 64-65+69.
- [2] 刘洋, 张悦, 朱丽芳. 基于几何布朗运动的房地产价格模拟预测[J]. 统计决策, 2018, 34(9): 86-89.
- [3] 张峻华, 戴伦, 刘文浩. 股票价格行为关于几何布朗运动的模拟——基于中国上证综指的实证研究[J]. 财经界(学术版), 2014(19): 31-33.
- [4] 韩婵, 陈东立. 非线性 Black-Scholes 模型下利差期权定价[J]. 西南师范大学学报(自然科学版), 2019, 44(7): 110-116.
- [5] 陆士杰, 杨朝军. 基于股价服从对数正态分布的凯利投资策略[J]. 经济数学, 2013, 30(3): 40-45.
- [6] 瞿慧, 陈静雯. 考虑跳跃波动与符号跳跃的 50ETF 期权定价研究[J]. 管理评论, 2019, 31(9): 28-36.
- [7] 高宏. 股票价格几何布朗运动模型的理论错误及纠正[J]. 时代金融, 2019(11): 50-51.
- [8] 王娟. 我国股票指数收益的正态性分析[J]. 江西教育学院学报(综合), 2004, 25(6): 62-65.
- [9] 金俊, 李俊. 上海股票市场收益率正态性分析[J]. 沿海企业与科技, 2005(11): 60.
- [10] 倪洪燕, 王勇. 股票市场价格序列正态性检验——基于上海股票市场 2007-2013 年数据[J]. 财务与金融, 2014(1): 33-37.
- [11] 张波, 商豪. 应用随机过程[M]. 北京: 中国人民大学出版社, 2001: 136-137.