Stock price return distribution of Hong Kong Stock Exchange

DING MENG

1 Introduction

Research shows that stock prices return distribution follow the power law behavior which is typical in finance market, and the time scale is a significant factor. Hong Kong Stock Exchange (HKEX) is a historic and mature stock market. It is a perfect choice to verify the financial theories of stock market.

2 Data Description

To accurately reflect the stock market changes of Hong Kong Stock Exchange (HKEX), we use the 50 model stocks of Hang Seng Index (HSI)[1] as research samples. HSI also divides the stocks into 10 types according to their industry classification, such as Financials, Information Technology, Telecommunications, Energy, Consumer Discretionary, Healthcare, Properties & Construction, Utilities, Conglomerates, and Industrials. Stock price data with different time scales (1 day, 1 hour, 30min, 15min and 1min) is downloaded from Yahoo Finance (https://finance.yahoo.com). The specific information of data is in appendix.

3 Theory

3.1 Stock Price Return

Financial market is always fluctuating over different time scales. To study these fluctuations such that the result is independent of the scale of measurement, logarithmic return of stock i for a time scale Δt is defined as

$$R_i(t, \Delta t) \equiv \ln P_i(t + \Delta t) - \ln P_i(t),$$

where $P_i(t)$ is the stock price at time t. In convenience of the comparisons of different stocks owning different return widths, normalized return is defined as

$$r_i(t, \Delta t) \equiv \frac{R_i(t) - \langle R_i(t) \rangle}{\sigma_i(t)}$$

where $\langle R_i(t) \rangle$ denotes the time average return over the given period and $\sigma_i(t)$ denotes the volatility of stock i.

3.2 Fat Tail of Stock Price Return Distribution

In financial market, probability distribution usually exhibits a large skewness from the normal distribution at the tail of the distribution, which is called a fat tail.

Researches[2] show that the tails of the cumulative return distribution for stock prices actually follow a power law

$$P_c(r > x) \sim x^{-\alpha}$$

with $\alpha \approx 3$, which is also called the inverse cubic law, and outside the stable levy regime $0 < \alpha < 2$, consistent with the fact that at larger scale the distribution tends to converge to a normal distribution.

3.3 TP Statistic and TE Statistic

3.3.1 TP Statistics

TP statistics is a powerful method to determine whether the tail of the distribution obeys the power law and where the power law tail start from (lower cut-off u). The raw data of a cumulative distribution is denoted by D, and a new set $X(x_1, x_2 \ldots x_n)$ is extracted from D by $x \geq u$. Define $TP(X)$ as

$$TP = \left[\frac{1}{n}\sum_{i=1}^{n} log\frac{x_i}{u}\right]^2 - \frac{1}{2}\left[\frac{1}{n}\sum_{i=1}^{n} log^2\frac{x_i}{u}\right]$$

If the underlying distribution obeys power law, the value of TP converges to zero as $n \to \infty$, and vice versa.

3.3.2 TE Statistics

Likewise, TE Statistics is a tool for the determination of exponential distribution, a distribution easily confused with the power law distribution in appearance. Its formula is as follows.

$$TE = -\frac{1}{n}\sum_{i=1}^{n} log^2\left(\frac{x_i}{u} - 1\right) - \left[\frac{1}{n}\sum_{i=1}^{n} \log\left(\frac{x_i}{u} - 1\right)\right]^2 - \frac{\pi^2}{6}$$

The value of TE tends to zero as $n \to \infty$ when the tail obeys an exponential distribution, and the value deviates from zero and fluctuates dramatically when it does not.

3.4 Tail Exponent α

3.4.1 Method: Curve Fitting

It is logical and meaningful to get the exponent α of the power law distribution tails, and one intuitive method is the curve fitting. Through TP statistics, it is easy to get the start point when distribution behaves in power law, i.e., the value of u satisfies $TP(u) = 0.05\ and\ TP(u') \to 0\ with\ u' \geq u$. Cut off the cumulative distribution of stock daily returns at u to get the tail and fit the power law formula $y = cx^{\alpha}$ to the tail data by methods like ordinary least squares.

3.4.2 Verification: Hill Estimator

The hill estimator is one of the best-known tail estimators, giving a consistent estimate of the tail exponent α from random samples of a distribution with an asymptotic power law form[3]. It is achieved by the following steps.

(1) Arrange all the returns in a decreasing order, i.e., $r_1 \geq r_2 \geq \cdots \geq r_n$.

(2) Extract the first $k\ (k = 1,2 \ldots n-1)$ returns from the decreasing statistics, i.e., $r_1,\ r_2 \ldots r_k$.

(3) $\gamma_{k,n} = \frac{1}{k}\sum_{i=1}^{k} log\frac{r_i}{r_{k+1}}$

Usually, $\gamma_{k,n} \to \frac{1}{\alpha}\ when\ k \to \infty\ and\ \frac{k}{n} \to 0$. Nevertheless, using such a k is literately impossible with finite returns. So, it is essential to choose an optimal $\bar{k}$, and subsample bootstrap method[4] provides a suitable solution.

We first choose a $k_0\ (e.g.\ 0.5\%\ of\ n)$ and get the γ_0. Then, choose various subsamples of size $n_s\ (e.g.\ \frac{n}{40})$ for cross validation. For each subsample, calculate $\gamma_{ks,ns}$ with $k_s = 1, 2 \ldots n_s - 1$. In the subsample, the best k_s owns the $\gamma_{ks,ns}$ minimizing the deviation from γ_0, the corresponding best k for the full size is $\bar{k} = \bar{k}_s(\frac{n}{ns})^{2/3}$.

After the calculations of all the subsamples, take the average value of the $\bar{k}$ as the final k_{best}. The estimated exponent $\alpha = \frac{1}{\gamma_{k_{best},n}}$. The 95% confidence interval is given by $\pm 1.96\left[1/\alpha^2\bar{k}\right]^{1/2}$.

4 Results

4.1 Stock Price Return

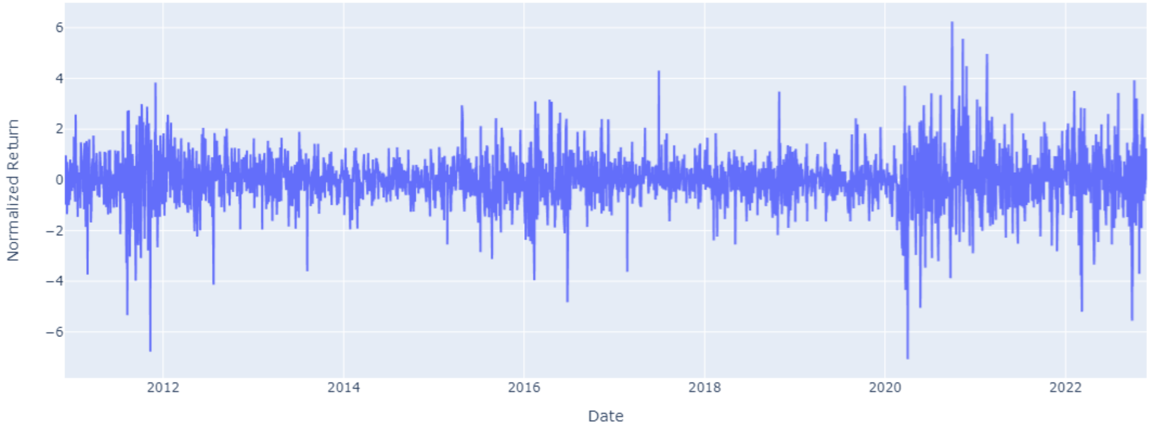


Fig.4-1. Normalized Daily Return of HSBC (0005.HK)

4.2 Fat Tail of Stock Price Return Distribution

Fig 4-2 shows the fat-tail phenomenon in stock market, using HSBC returns at the time scale of 1 day.

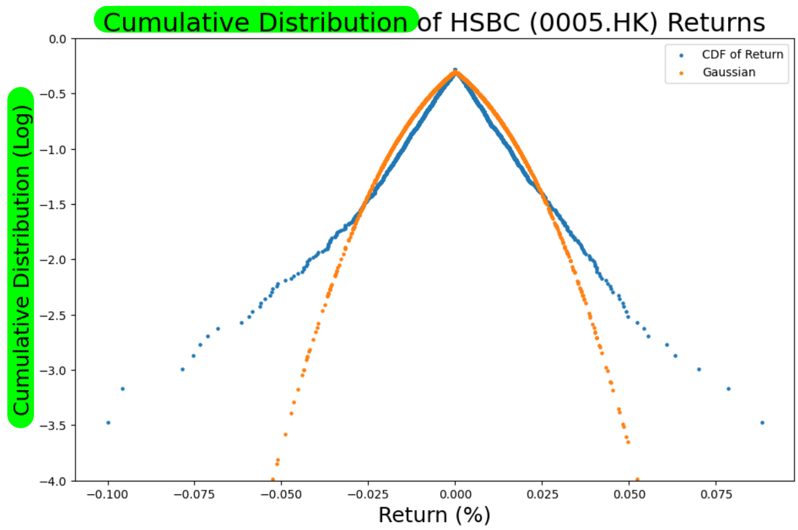


Fig.4-2. Cumulative Distribution of HSBC (0005.HK) Daily Returns

Choose one typical stock of each industry classification from HSI model stocks and plot their positive and negative normalized return cumulative distribution respectively (Fig 4-3). It roughly verifies the inverse cubic law hypothesis. To furtherly determine the type of the distribution tails, statistical methods like TP statistics and TE statistics should be applied.

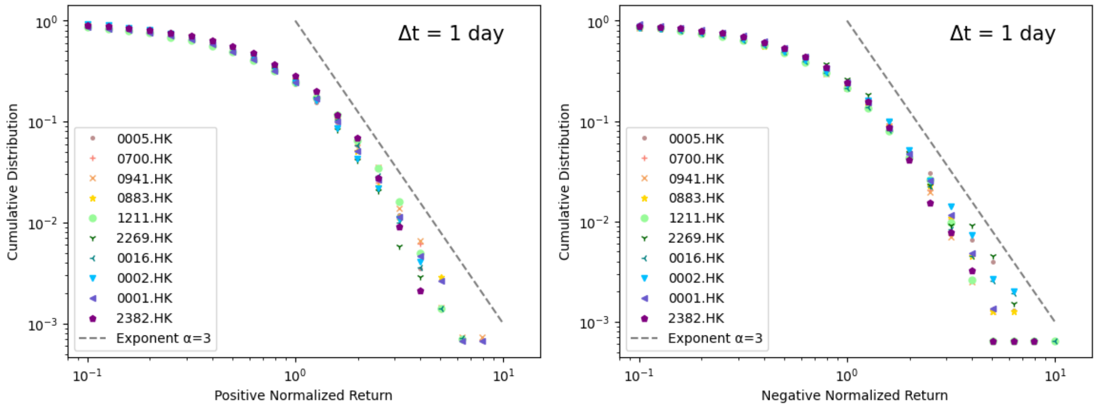


Fig.4-3 10 Typical Stocks' Returns Distribution

4.3 TP and TE Statistics

4.3.1 Explanation of TP and TE Statistics

According to Fig 4-4, with the increase of u, TP decreases continuously and converges to zero when $u \geq 0.9$, and the fluctuations occurring when u is relatively large might be the result of sample size effect. This illustrates that the cumulative distribution becomes a power law distribution when normalized return is larger than 0.9 (the tail part).

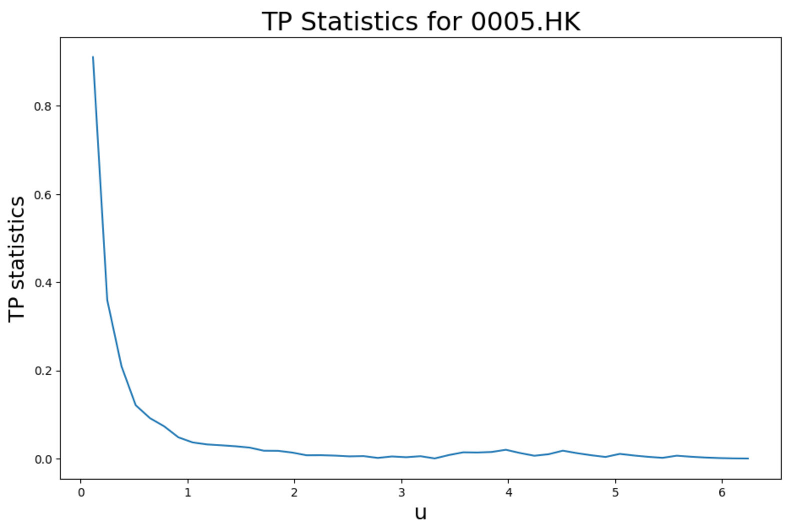


Fig.4-4. TP Statistics for 0005.HK

Fig 4-5 shows a typical situation of TE statistics of a non-exponential distribution with u increasing. The result of TP statistics and TE statistics together prove the power law property of the daily return of stock 0005.HK.

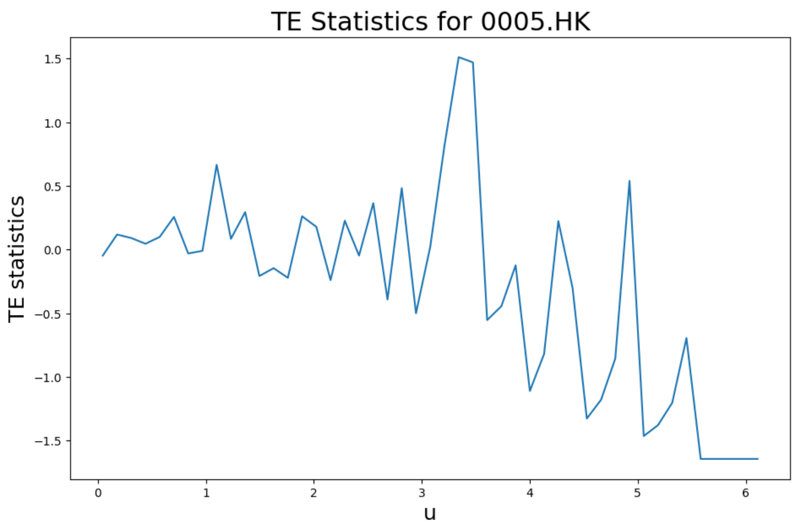
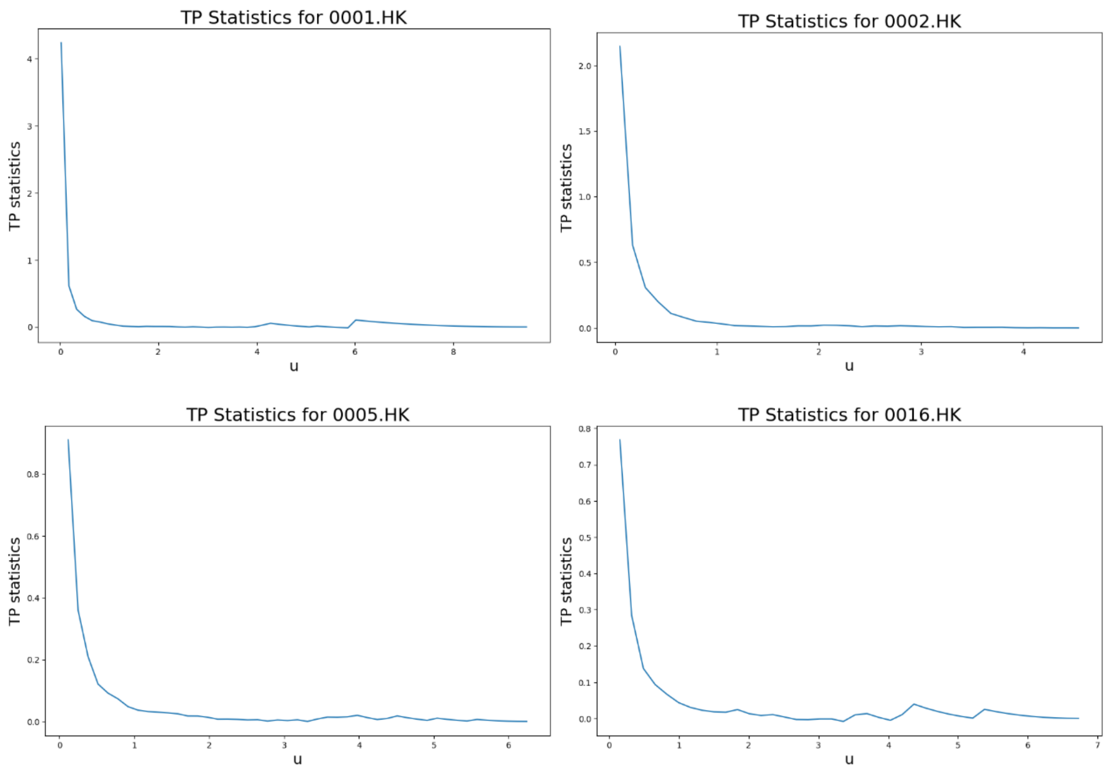


Fig.4-5. TE Statistics for 0005.HK

4.3.2 Result of TP and TE Statistics

However, one stock cannot represent all the stocks in the market, hence we apply the TP and TE statistics to all the 50 stocks and show the results of the 10 typical stocks in Fig 4-6, 4-7 below. The results illustrate the fact that the daily return distribution tails all follow power law behaviors.



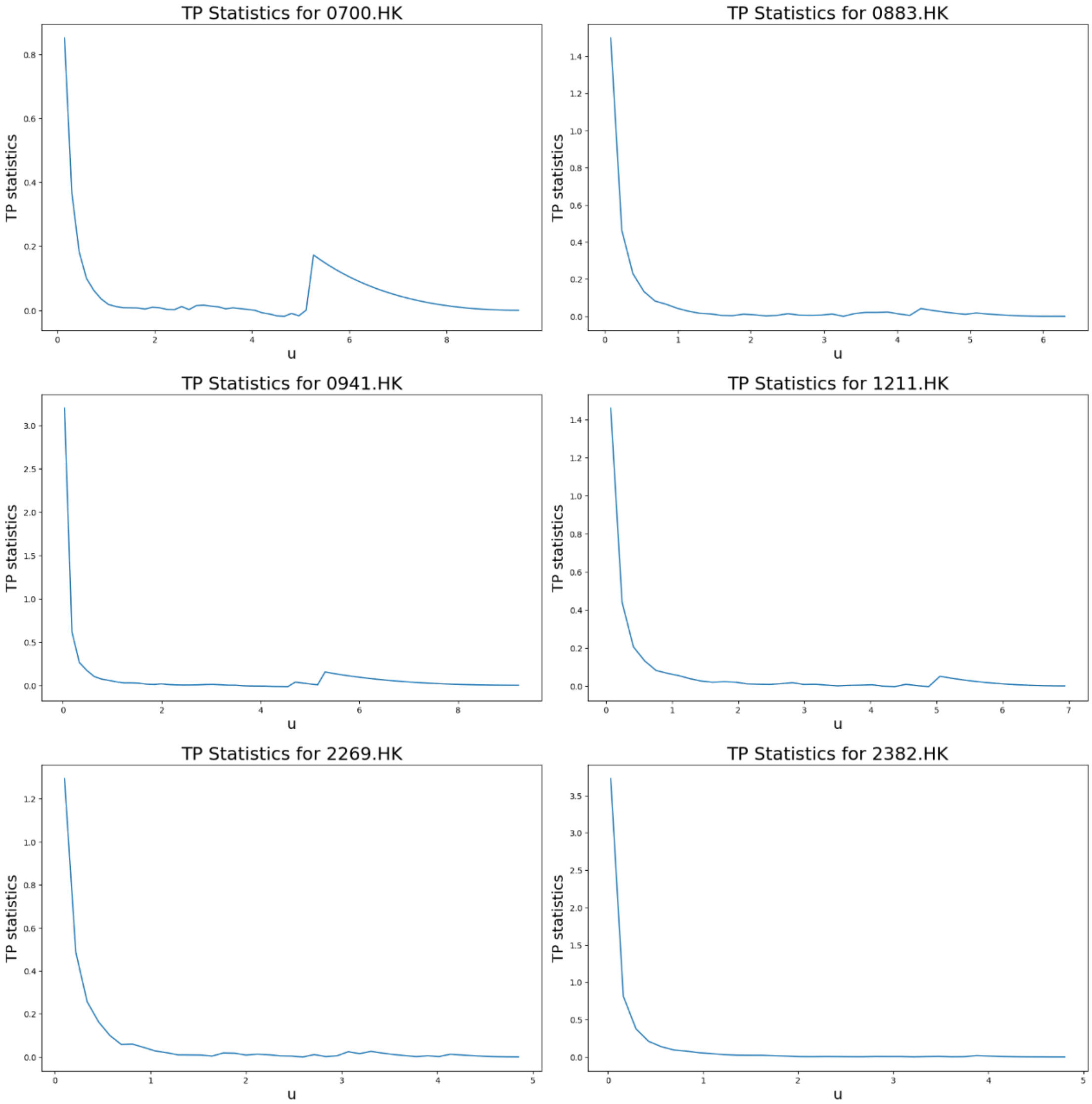
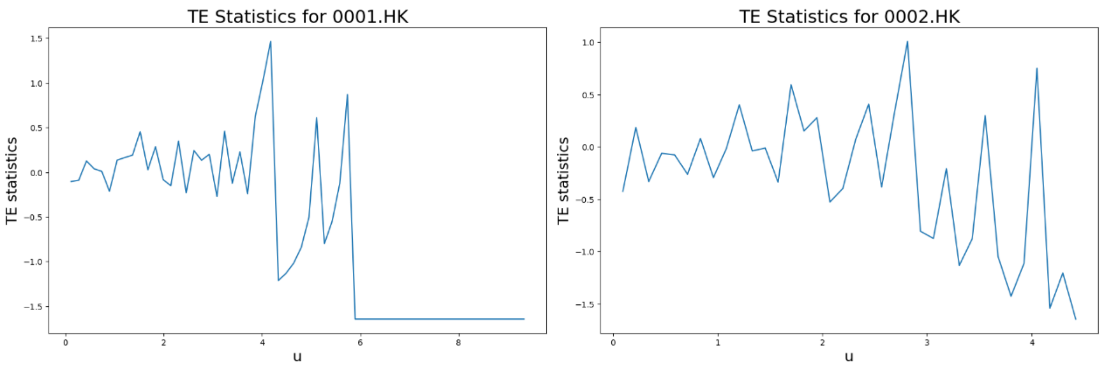


Fig.4-6. TP Statistics for 10 Typical Stocks



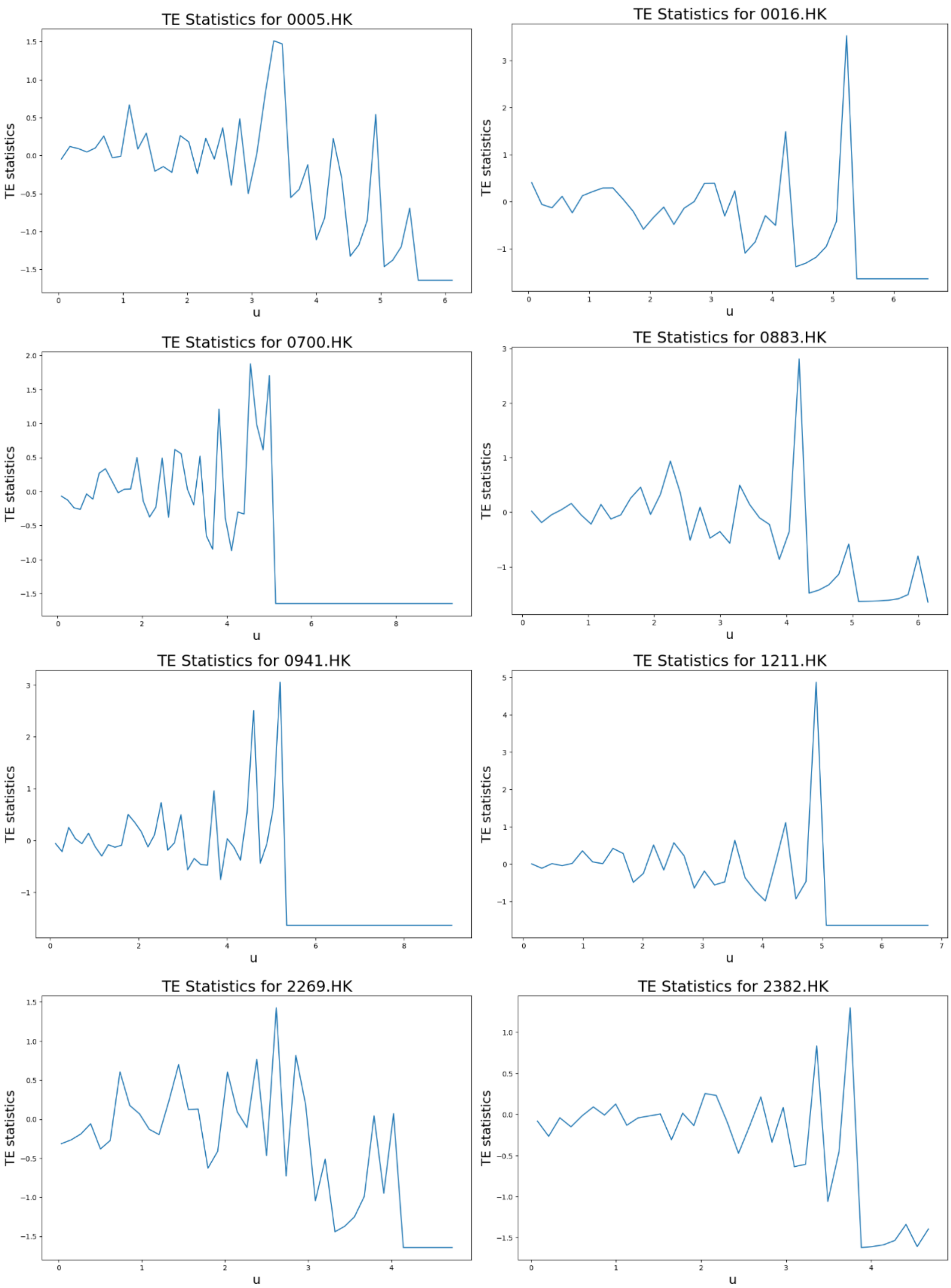


Fig.4-7. TE Statistics for 10 Typical Stocks

4.4 Tai Exponent α

4.4.1 Obtain the Value of Exponent α

Fig 4-8 shows the curve fitting result of the positive and negative tails of stock 0005.HK respectively. The fitting results match the real data greatly. Also, it is obvious that fitted exponent α of the negative tail is notably smaller than that of the positive tail.

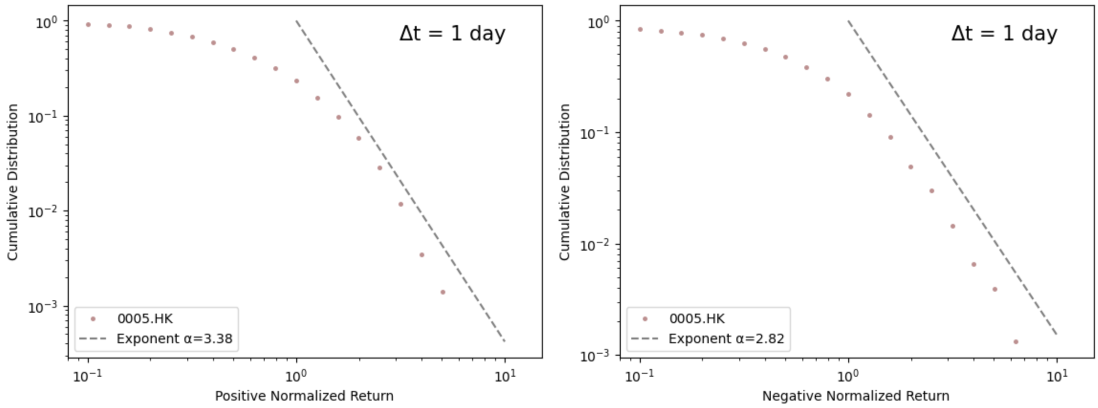


Fig.4-8. Curve Fitting Results for 0005.HK

Table 4-1 shows the estimated results of exponent α using the aforementioned two methods, i.e., curve fitting and hill estimator. It is obvious that the corresponding results are close and consistent, verifying the accuracy of the estimated exponent α got by curve fitting.

Table.4-1. Estimated Exponents of 10 Typical Stocks

	Positive Tail			Negative Tail		
Stock	α_1	α_2	**Confidence**	α_1	α_2	**Confidence**
0005.HK	3.38	3.45	0.06	2.82	2.75	0.07
0700.HK	3.02	3.14	0.06	3.39	3.70	0.06
0941.HK	3.25	3.51	0.07	3.56	4.23	0.05
0883.HK	3.02	3.29	0.06	3.00	3.27	0.06
1211.HK	3.80	3.57	0.07	2.88	2.97	0.07
2269.HK	3.48	3.55	0.09	2.56	2.83	0.10
0016.HK	3.62	3.52	0.06	2.60	2.89	0.07
0002.HK	3.15	3.32	0.06	2.70	3.08	0.07
0001.HK	3.05	3.10	0.07	3.34	2.88	0.07
2382.HK	3.77	4.01	0.05	3.22	3.41	0.05

α_1 is the estimated result of curve fitting, α_2 is the estimated result of hill estimator.

4.4.2 Exponent α with Variation of Stock Classification

According to Fig 4-3, the daily return of different stocks follows very similar distributions, so we merge the absolute returns of the 50 stocks to obtain a single normalized absolute return distribution. Get the estimated exponent α by curve fitting, the result is shown in Fig 4-9. The estimated exponent $\alpha = 3.83$, relatively larger than the theoretical value 3, and larger than most of the exponent values got by single stock.

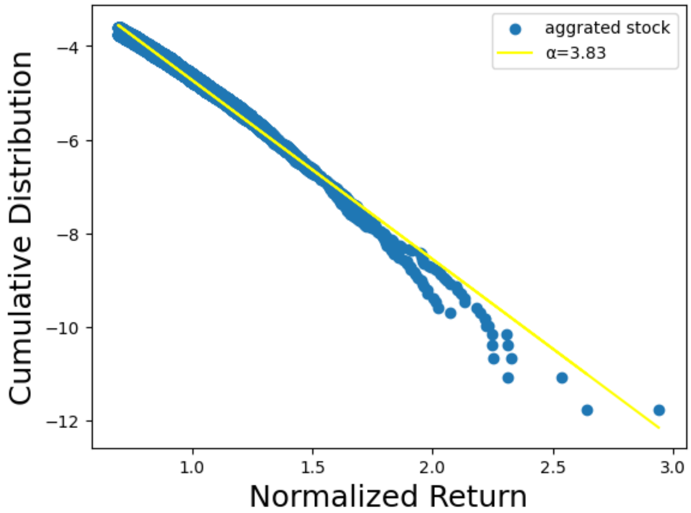


Fig.4-9. Curve Fitting Results for Aggregated Stock

To research the difference of the exponent between stocks of different classifications, divide the 50 stocks into 6 classes based on their industry classification in HIS, i.e., consumer, finance, energy and utility, IT, property, and other stocks. In accordance with Fig 4-10, all the exponents are close to 3.7 except consumer stocks, whose estimated exponent is up to 3.9.

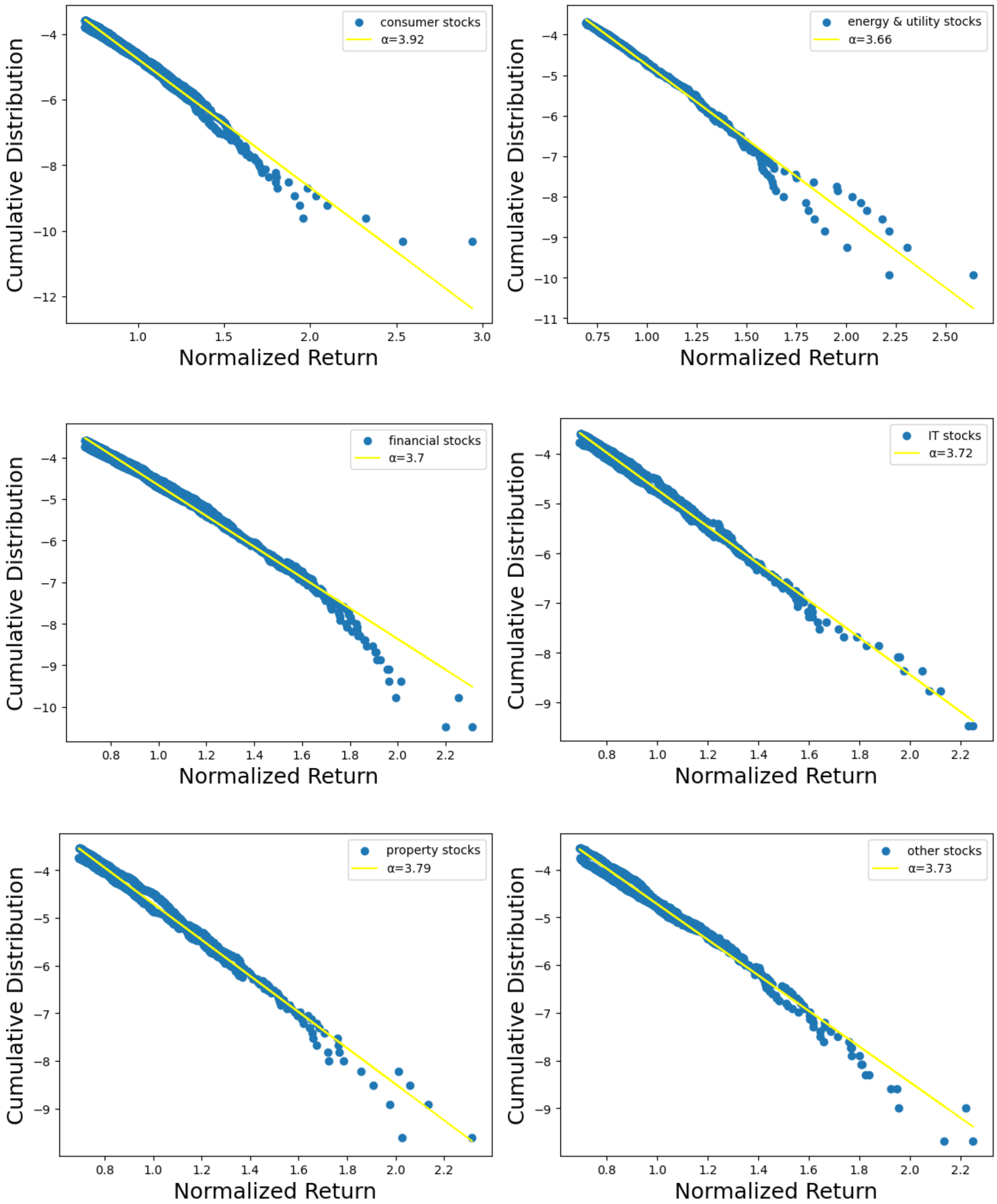


Fig.4-10. Curve Fitting Results for Stocks of Different Industries

4.4.3 Exponent α with Variation of Interval Δt

Considering daily changes in stock return, we also explore the return distribution as well as their exponents at high frequency time scale, like 1 hour, 30 min, 15 min and 1 min. Through TP and TE statistics, all the stocks follow power law distribution behaviors. Fig 4-11 to Fig 4-15 show the exponent distribution of the 50 stocks at different time scales.

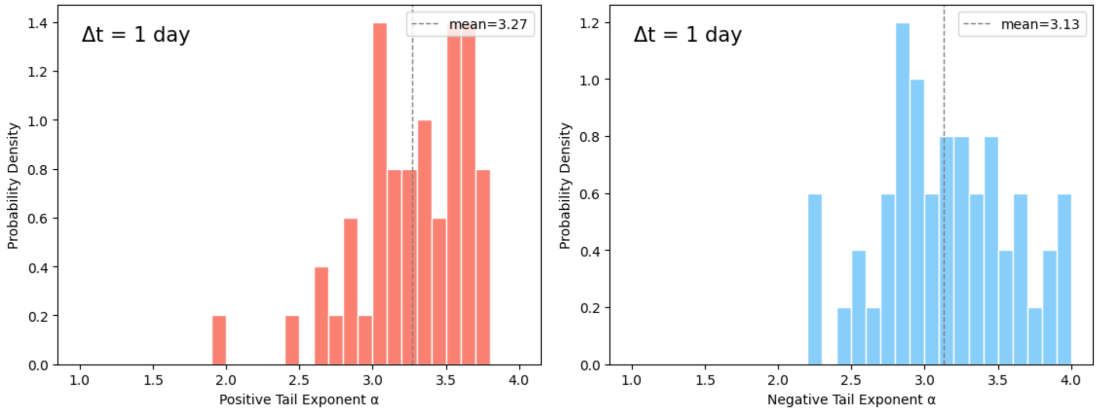


Fig.4-11. Exponents Distribution at $\Delta t = 1\ day$

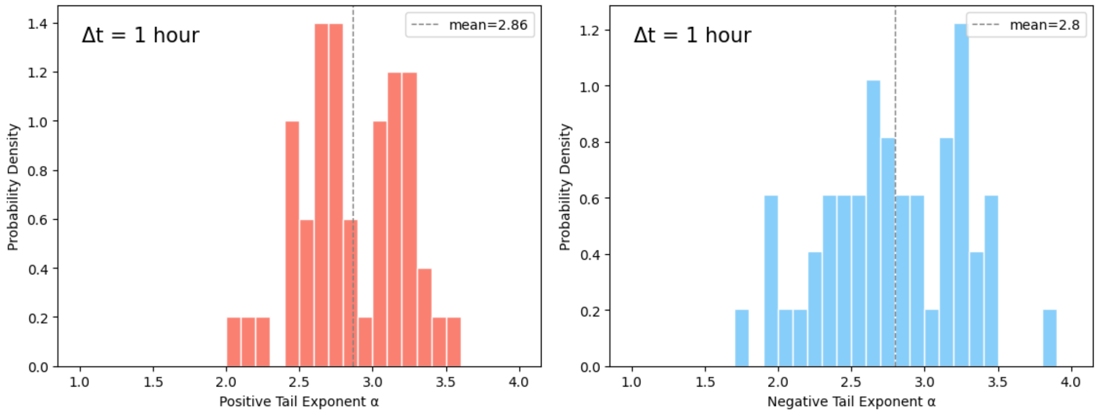


Fig.4-12. Exponents Distribution at $\Delta t = 1\ hour$

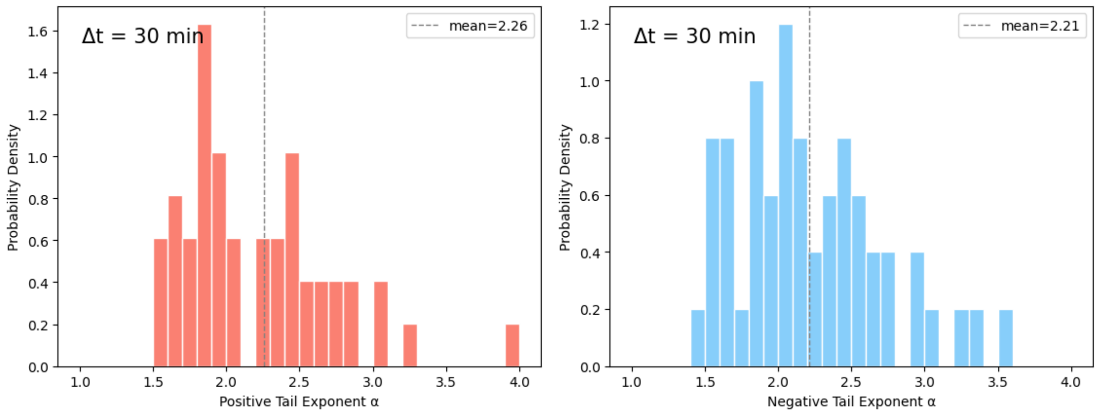


Fig.4-13. Exponents Distribution at $\Delta t = 30\ min$

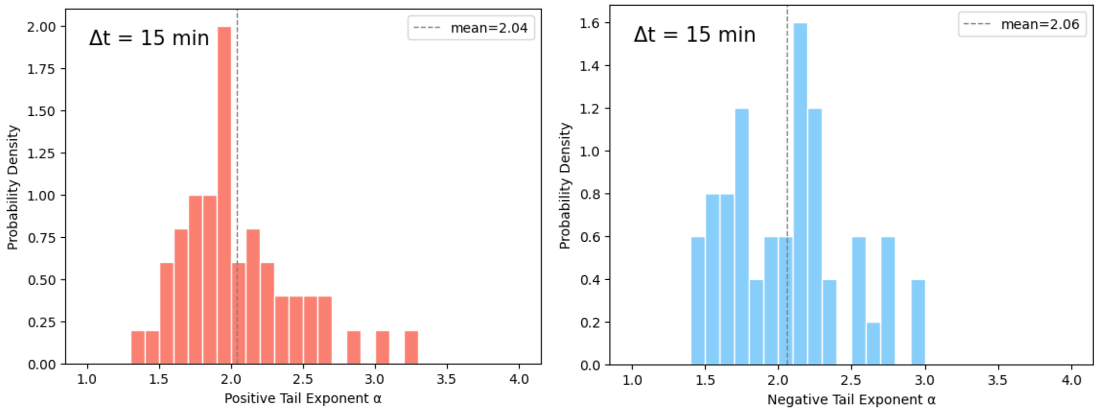


Fig.4-14. Exponents Distribution at $\Delta t = 15\ min$

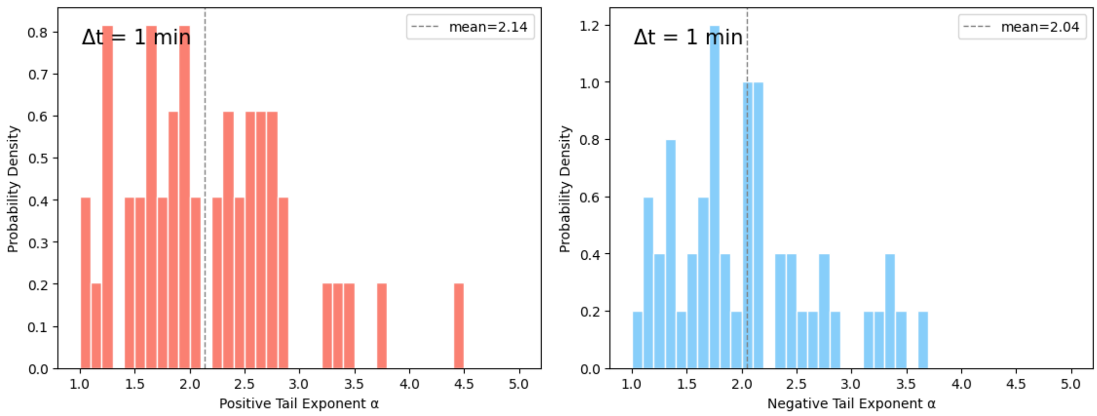


Fig.4-15. Exponents Distribution at $\Delta t = 1\ min$

Compare vertically, whether positive tail or negative tail, the average exponent as well as the range of exponents tend to move left with the decrease of Δt. This means higher-frequency data usually owns smaller exponents. Compare horizontally, at the same time scale, positive tail's power law exponent is prone to be relatively larger than that of negative tail. Also, positive tail exponents seem to be more concentrated than negative tail exponents.

Table.4-2. Average Exponent

Time Scale	**Positive Tail α**	**Negative Tail α**
1 day	3.27	3.13
1 hour	2.86	2.80
30 min	2.26	2.21
15 min	2.04	2.06
1 min	2.14	2.04

5 Conclusion

(1) Power law is a common phenomenon in return distribution of stock price, a truth that is verified many times.

(2) The inverse cubic law is not always the case, especially to the time scale shorter than 1 hour. The decrease of time scale tends to result in a decrease of power law exponent.

(3) The return distribution is an asymmetric distribution, the power law exponent of positive tail is usually larger than that of negative tail.

(4) Aggregation is not a precise way to reflect the overall situation of the stock market, market indexes are better choices.

(5) Different stock classes do not show significant differences in exponent which might because of the classification method and lack of samples.

6 Appendix

Table.6-1. Time Information of Data

Time Scale	Start Date	End Date
1 day	2010-11-28	2022-11-27
1 hour	2021-01-01	2022-12-01
30 min	2022-10-15	2022-12-10
15 min	2022-10-15	2022-12-10
1 min	2022-12-04	2022-12-11

Table.6-2. Specific Information of Stocks

恒生指數
HANG SENG INDEXES

CONSTITUENTS (TOP 50)

Stock Code	ISIN CODE	Company Name	Industry Classification	Share Type	Weighting (%)
5	GB0005405286	HSBC HOLDINGS	Financials	HK Ordinary	8.90
1299	HK0000069689	AIA	Financials	HK Ordinary	8.04
700	KYG875721634	TENCENT	Information Technology	Other HK-listed Mainland Co.	6.90
9988	KYG017191142	BABA - SW	Information Technology	Other HK-listed Mainland Co.	6.75
3690	KYG596691041	MEITUAN - W	Information Technology	Other HK-listed Mainland Co.	6.52
939	CNE1000002H1	CCB	Financials	H Share	5.08
941	HK0941009539	CHINA MOBILE	Telecommunications	Red Chip	3.27
388	HK0388045442	HKEX	Financials	HK Ordinary	2.83
1398	CNE1000003G1	ICBC	Financials	H Share	2.83
9618	KYG8208B1014	JD - SW	Information Technology	Other HK-listed Mainland Co.	2.69
3988	CNE1000001Z5	BANK OF CHINA	Financials	H Share	2.26
2318	CNE1000003X6	PING AN	Financials	H Share	2.24
883	HK0883013259	CNOOC	Energy	Red Chip	1.88
1211	CNE100000296	BYD COMPANY	Consumer Discretionary	H Share	1.74
1810	KYG9830T1067	XIAOMI - W	Information Technology	Other HK-listed Mainland Co.	1.73
2269	KYG970081173	WUXI BIO	Healthcare	HK Ordinary	1.43
16	HK0016000132	SHK PPT	Properties & Construction	HK Ordinary	1.38
669	HK0669013440	TECHTRONIC IND	Consumer Discretionary	HK Ordinary	1.31
2	HK0002007356	CLP HOLDINGS	Utilities	HK Ordinary	1.20
1	KYG217651051	CKH HOLDINGS	Conglomerates	HK Ordinary	1.18
3968	CNE1000002M1	CM BANK	Financials	H Share	1.13
823	HK0823032773	LINK REIT	Properties & Construction	HK Ordinary	1.10
2331	KYG5496K1242	LI NING	Consumer Discretionary	Other HK-listed Mainland Co.	1.08
2020	KYG040111059	ANTA SPORTS	Consumer Discretionary	Other HK-listed Mainland Co.	1.05
2388	HK2388011192	BOC HONG KONG	Financials	HK Ordinary	1.02
1113	KYG2177B1014	CK ASSET	Properties & Construction	HK Ordinary	0.98
11	HK0011000095	HANG SENG BANK	Financials	HK Ordinary	0.95
2319	KYG210961051	MENGNIU DAIRY	Consumer Staples	Other HK-listed Mainland Co.	0.90
386	CNE1000002Q2	SINOPEC CORP	Energy	H Share	0.89
1109	KYG2108Y1052	CHINA RES LAND	Properties & Construction	Red Chip	0.89
9633	CNE100004272	NONGFU SPRING	Consumer Staples	H Share	0.89
27	HK0027032686	GALAXY ENT	Consumer Discretionary	HK Ordinary	0.88
981	KYG8020E1199	SMIC	Information Technology	Other HK-listed Mainland Co.	0.83
1093	HK1093012172	CSPC PHARMA	Healthcare	Other HK-listed Mainland Co.	0.81
1088	CNE1000002R0	CHINA SHENHUA	Energy	H Share	0.79
3	HK0003000038	HK & CHINA GAS	Utilities	HK Ordinary	0.76
66	HK0066009694	MTR CORPORATION	Consumer Discretionary	HK Ordinary	0.72
2628	CNE1000002L3	CHINA LIFE	Financials	H Share	0.72
857	CNE1000003W8	PETROCHINA	Energy	H Share	0.71
2688	KYG3066L1014	ENN ENERGY	Utilities	Other HK-listed Mainland Co.	0.70
9999	KYG6427A1022	NTES - S	Information Technology	Other HK-listed Mainland Co.	0.69
291	HK0291001490	CHINA RES BEER	Consumer Staples	Red Chip	0.68
688	HK0688002218	CHINA OVERSEAS	Properties & Construction	Red Chip	0.65
6	HK0006000050	POWER ASSETS	Utilities	HK Ordinary	0.59
267	HK0267001375	CITIC	Conglomerates	Red Chip	0.58
1997	KYG9593A1040	WHARF REIC	Properties & Construction	HK Ordinary	0.58
175	KYG3777B1032	GEELY AUTO	Consumer Discretionary	Other HK-listed Mainland Co.	0.57
2382	KYG8586D1097	SUNNY OPTICAL	Industrials	Other HK-listed Mainland Co.	0.55
992	HK0992009065	LENOVO GROUP	Information Technology	Other HK-listed Mainland Co.	0.51
2313	KYG8087W1015	SHENZHOU INTL	Consumer Discretionary	Other HK-listed Mainland Co.	0.51

Reference

1. Hang Seng Indexes
   https://www.hsi.com.hk/eng
2. Gopikrishnan, P., Meyer, M., Amaral, L. et al. Inverse cubic law for the distribution of stock price variations. Eur. Phys. J. B 3, 139–140 (1998).
3. Statistics of Extreme Events
   http://sfb649.wiwi.huberlin.de/fedc_homepage/xplore/tutorials/sfehtmlnode91.html#mex
4. Olivier V. Pictet & Michel M. Dacorogna & Ulrich A. Muller, 1996. "Hill, Bootstrap and Jackknife Estimators for Heavy Tails," Working Papers 1996-12-10, Olsen and Associates.