Explore Not Independence and Correlation of Random Variables and Methods to Judge Them between Stocks

Han Han[1], Zhehan Wang[2]*, Qixun Dong[3], Zugen Liu[4]

[1]*School of Guangxi University, Nanning, Guangxi 530000,China*
[2]*University of Liverpool, Liverpool, Merseyside L69 3BX,United Kingdom*
[3]*Beilin Tieyi Middle School, Xian, Shanxi,710000, China*
[4]*Chengdu Foreign Languages School, Chengdu, Sichuan,610000, China*
**Corresponding author. Email: sgzwan93@liverpool.ac.uk*

ABSTRACT
In recent years, whether there is independence and correlation between stocks and how to use research results to minimize risks and maximize investment returns in the stock market have attracted the attention of many economists, mathematicians, and the public. It is worthy of discussion about how to correctly study the correlation and independence between stocks to form an optimized investment thinking and investment mode. This work mainly combs the meaning of independence and correlation and uses various proof methods to distinguish some interdependent relations of independence and correlation in some cases. This work uses mathematical proof and statistical software to verify whether there is independence and correlation between stocks. This paper uses mathematical proof to verify the independence of stocks, uses STATA to estimate the statistical conclusion between stocks, shows that the positive or negative correlation between stocks is significant, and studies its economic significance.

Keywords: *Stock, Independence, Correlation, Correlation Coefficient, STATA*

1. INTRODUCTION

1.1.Research background, purpose, and significance of the work

In random variables, correlation and independence have always been a topic worth discussing, and there are many ways to distinguish them. It is vital to explore whether stocks have a correlation and the correlation between stocks to study the changes of the entire stock market and the investment choices of shareholders. Through the correlation of stocks, this work can explore the degree of mutual influence and interaction between different stocks, which is of constructive significance for shareholders to choose stocks for investment. The independence of stocks is also proved in this work. It is of innovative significance to better control market changes and the relationship between stocks by exploring the independence of stocks.

Stock correlations for investment decisions have significant reference value. The stock of different industries in the stock market in our country shows excessive correlation, and the correlation of industry stocks is likely to be affected by the excess linkage in the stock market, which will lead to producing inaccurate correlation between stocks and has a certain degree of influence on investors' decisions. In the early stage of stock market research, foreign scholars believed that the stock market had moderate fluctuation state law and created a classical financial theory to explain the reasons for stock price fluctuation. The theory proposes that based on market efficiency, some mathematical formulas can calculate the prices of financial assets and their derivatives. Under the theoretical framework, the change of stock price has a certain regularity, and stock prices can fully reflect the value of stocks. That is why stock price volatility is mainly composed of stock transaction costs, liquidity, and other essential factors; it can also further explain that the stock fundamental information changes will fully embody a state of share price volatility.

1.2.Main research contents of the work

In this work, the independence and correlation of random variables are discriminated from the perspective

Copyright © 2022 The Authors. Published by Atlantis Press International B.V.
This is an open access article distributed under the CC BY-NC 4.0 license -http://creativecommons.org/licenses/by-nc/4.0/.

of multiple mathematical proofs through the definition of independence and correlation. Moreover, through the mathematical proof method to verify that the stock is not independent, and the use of statistical software STATA to analyze the correlation of the stock, through the scatter diagram and correlation coefficient to judge the correlation between stocks, and from the correlation coefficient to draw a significant conclusion.

1.2.1.Research ideas and methods

Correlation analysis is used to study the relationship between quantitative data, including whether there is a relationship and how close the relationship is. If significant (* in the upper right corner of the result, there is a relationship; Otherwise, it does not matter); After the relationship, the closeness of the relationship directly depends on the size of the correlation coefficient. Generally, above 0.7, the relationship is very close; The range from 0.4 to 0.7 indicates that the relationship is close. 0.2 to 0.4 Indicates that the relationship is normal. If the correlation value is less than 0.2, it still presents significance (* sign in the upper right corner, 1 * signal is significant at 0.05 level, 2 * signal is significant at 0.01 level; Significant means that the occurrence of the correlation coefficient is statistically significant and common, rather than accidental), indicating that the relationship is weak, but there is still a correlation. Correlation analysis is the premise of regression analysis. First, the correlation should be guaranteed, and then the regression influence relationship can be studied. Because if they all show no correlation, it is impossible to affect the relationship. If there is a correlation, it is not necessarily regressive.

1.3.Based on some research results

We have learned experience from a paper called 'Forecasting International Equity Correlations' authored by Claude B. Erb, Campbell R. Harvey, and Tadas E. Viskant [1]. It concerns the international stocks level in terms of change in correlation. Research shows that the relationship is related to the economic cycle. The innovation of this paper lies in the use of out-of-sample global portfolio allocation and derivatives. Then, concerning the correlation between stocks, we have collected material on 'Studying the Correlation of Stocks via Copula Function' by Yishuai, T., Boying, L., Botao, L [2]. It focused on the determination that affects the stock value; the most critical part is the correlation between stock and stock index research. This paper chose the closing prices of 4 industries and 15 industries related to the market index for correlation analysis based on the Copula function, which is the creative side of this paper.

Furthermore, we have gathered the paper which investigated the correlation of stock market among China Mainland, China Hongkong, and America by Wenrong P [3]. It is generally discussed that traditional linear models cannot accurately and reliably measure the co-activity of the international stock market due to the uncertainty of stock market data. Fourthly but lastly, to get results from a broader range, we reviewed the work from Monica S and Shachi P, 'Volatility and cross-correlations of stock markets in SAARC nations'[4]. This paper aims to examine the relationship of stocks between SAARC countries. The results show that the stock market returns are sequentially auto correlated, which indicates that the current stock price is dependent on the previous stock price and leads to the rejection of the efficient market hypothesis.

1.4.Difficulties encountered and some problems

This work focuses not only on the correlation between stocks but also on the independence of stocks. Therefore, proving the non-independence of the stock and explaining the significance of the non-independence of the stock become the main problems encountered in this work.

In the process of using STATA to verify stock correlation, the author has been struggling with whether to construct the regression between stocks. However, since there is only one explained variable, it is impossible to use a scatter chart to reflect the correlation of all stocks and impossible to use regression analysis to conveniently explore the correlation of all stocks in the data given in this work. Therefore, the regression analysis is abandoned, and only correlation analysis is carried out, and better results are obtained.

1.4.1.innovations

This work is the first to use mathematical definitions and theorems after analyzing the relationship between independence and correlation and then bring into the analysis of stocks to carry out the organic combination of mathematical concept discrimination and economical application. This work focuses on the correlation proof and application analysis between stocks and chooses to verify the non-independence of stocks and carries on the interpretation analysis.

2. INDEPENDENCE AND CORRELATION OF TWO RANDOM VARIABLES

2.1.Approach to discriminating the independence and correlation of two-dimensional discrete random variables X and Y

2.1.1.Approach to discriminating the independence

Definition: For all values of x and y (x_i, y_j), $i = 1,2,3,\ldots, j = 1,2,3,\ldots$, there is $P(X = x_i, Y = y_j) =$

$P\big(X = x_i\big)P\big(Y = y_j\big)$. Let $P_X(x_i) = \sum_j P_{X,Y}(x_i, y_j)$, $P_Y\big(y_j\big) = \sum_i P_{X,Y}(x_i, y_j)$, that is $P_{X,Y}\big(x_i, y_j\big) = P_X(x_i)P_Y\big(y_j\big)$ for all i, j.

Definition: Considering conditional probability, $P_{X|Y}\big(x_i, y_j\big) = P_X(x_i)$ and $P_{Y|X}\big(x_i, y_j\big) = P_Y\big(y_j\big)$ for all i, j.

Deduction: Any elements of two lines of the joint distribution matrix of X, Y are proportional, namely the rank of the joint distribution matrix equals 1, then X, Y are independent [5].

Proof: Assuming the joint distribution matrix of X, Y is $A = \begin{pmatrix} P_{x_1y_1} & P_{x_1y_2} & \cdots & P_{x_1y_j} \\ P_{x_2y_1} & P_{x_2y_2} & \cdots & P_{x_2y_j} \\ \vdots & \vdots & \ddots & \vdots \\ P_{x_iy_1} & P_{x_iy_2} & \cdots & P_{x_iy_j} \end{pmatrix}$, where $P_{x_iy_j} = P_{X,Y}(x_i, y_j)$ and the margin distribution matrices of X, Y are $(P_X(x_1)P_X(x_2)\ldots P_X(x_i))$ and $(P_Y(y_1)P_Y(y_2)\ldots P_Y(y_j))$ respectively.

Sufficiency: If two arbitrary lines of the disjoint distribution matrix are proportional, that is

$$\frac{P_{x_my_1}}{P_{x_ny_1}} = \frac{P_{x_my_2}}{P_{x_ny_2}} = \ldots = \frac{P_{x_my_j}}{P_{x_ny_j}} \quad (1)$$

According to the geometric property,

$$\frac{P_{x_my_1}}{P_{x_ny_1}} = \frac{P_{x_my_1}+P_{x_my_2}+\cdots+P_{x_my_j}}{P_{x_ny_1}+P_{x_ny_2}+\cdots+P_{x_ny_j}} = \frac{\sum_j P_{x_my_j}}{\sum_j P_{x_ny_j}} = \frac{P_X(x_m)}{P_X(x_n)} \quad (2)$$

From formula (2), which can also be written as

$$\frac{P_{x_my_j}}{P_X(x_m)} = \frac{P_{x_ny_j}}{P_X(x_n)} \quad (3)$$

Since n is arbitrary,

$$\frac{P_{x_my_j}}{P_X(x_m)} = \frac{P_{x_ny_1} + P_{x_ny_2} + \cdots + P_{x_ny_j}}{P_X(x_1) + P_X(x_2) + \cdots + P_X\big(x_j\big)} = \frac{\sum_n P_{x_ny_j}}{1} = \sum_n P_{x_ny_j} \quad (4)$$

So $P_{x_my_j} = P_X(x_m)\sum_n P_{x_ny_j} = P_X(x_m)P_Y(y_j)$. According to definition 1, QED.

Necessity: Take two arbitary lines, the mth line and the n th line. There is $P_{x_iy_j} = P_{X,Y}\big(x_i, y_j\big) = P_X(x_i)P_Y(y_j)$ since X, Y are independent. Then

$$\frac{P_{x_my_k}}{P_{x_ny_k}} = \frac{P_X(x_m)P_Y(y_j)}{P_X(x_n)P_Y(y_j)} = \frac{P_X(x_m)}{P_X(x_n)} = c \text{ (} c \text{ is a constant)} \quad (5)$$

Therefore, for the corresponding columns of the two lines,

$$\frac{P_{x_my_1}}{P_{x_ny_1}} = \frac{P_{x_my_2}}{P_{x_ny_2}} = \cdots = \frac{P_{x_my_j}}{P_{x_ny_j}} \quad (6)$$

which means the m th line and the nth line are proportional, and according to the theory of matrices, the rank (A)=1.

Theorem: If the joint distribution of X abd Y is normal, then independence can be derived from irrelevance since independence and irrelevance are equivalent under this situation. See 2.2.3.

2.1.2.Approach to discriminating the correlation

Correlation coefficient is often used to judge whether two random variables are linear correlated. It is denoted by r,

$$r = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\sum_{i=1}^{n}(y_i-\bar{y})^2}} = \frac{COV[X,Y]}{STED[X]STED[Y]}, |r| < 1 \ (7)$$

If correlation coefficient r is near 0, then X and Y are not linear correlated, but it does not mean that X and Y are not correlated since there maybe exist a nonlinear correlation relationship. If r is near 1, then X and Y are linear correlated to some extent. To further study the linear relationship between X and Y, the least-square approximation will be introduced.

Given a group of experimentally determined point $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, roughly estimate that this set of data is linearly dependent. To fit the linear equation $y = a + bx$ to the data, calculate the error

$$E = \sum_{i=1}^{n}(y_i - bx_i - a)^2 \quad (8)$$

When E gets the minimum value, then $y = ax + b$ is called the least squares line which can represent the best possible fit to the data. Let matrix $A = \begin{pmatrix} x_1 & 1 \\ x_2 & 1 \\ \ldots & \ldots \\ x_m & 1 \end{pmatrix}$, $\overrightarrow{x_0} = \binom{a}{b}$, $\vec{y} = \begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_m \end{pmatrix}$, then

$$E = \|y - Ax\|^2 \quad (9)$$

To find a and b, which minimize E, x_0 should be found firstly such that $\|y - Ax_0\|^2 \le \|y - Ax\|^2$ for all vectors x.

By Approximation Theorem: If W is a finite-dimensional subspace of an inner product space V, and if $\vec{b}$ is a vector in V, then $proj_w\vec{b}$ is the best approximation to $\vec{b}$ from W in the sense that $\|\vec{b} - proj_w\vec{b}\| < \|\vec{b} - \vec{w}\|$ for every vector $\vec{w}$ in W that is different from $proj_w\vec{b}$ [2].

Define $W = A\vec{x}{:}\vec{x}F^n$, then $A\overrightarrow{x_0} = proj_w\vec{b}, A\overrightarrow{x_0} - \vec{y}$ belong to the vertical plane of W, so $(A\vec{x}, A\overrightarrow{x_0} - \vec{y})_m = 0$ for all $x \in F^n$.

By the corollary theorem of matrix [6]: $A^{**} - A, (AB)^* = B^*A^*$. There is $(A\vec{x}, A\overrightarrow{x_0} - \vec{y})_m = (A\overrightarrow{x_0} - \vec{y})^*(A\vec{x}) = ((A\overrightarrow{x_0} - \vec{y})^*A)\vec{x} = (A^*(A\overrightarrow{x_0} - \vec{y}))^*\vec{x} =$

$(\vec{x}, A^*(A\overrightarrow{x_0} - \vec{y}))_n$ for all $x \in F^n$, which means $A^*(A\overrightarrow{x_0} - \vec{y}) = 0$, namely $A^*A\overrightarrow{x_0} = A^*\vec{y}$. By the theorem [6]: If A is an $m \times n$ matrix, then the following are equivalent:

<1>The column vectors of A are linearly independent

<2>A^TA is invertible

So $x_0 = (A^*A)^{-1}A^*y$. (Note that if A has real entries, A^* is simply the transpose of A). Finally, $y = a + bx$ is obtained.

2.2. The relationship between independence and irrelevance of two-dimensional random variables

2.2.1. Irrelevance can be derived from independence in any case

Proof: $COV[X,Y] = E[(X - E[X])(Y - E[Y])] = E[XY - XE[Y] - YE[X] + E[X]E[Y]] = E[XY] - E[X]E[Y] - E[Y]E[X] + E[X]E[Y] = E[XY] - E[X]E[Y]$. Since X and Y are independent, there is $E[XY] = E[X]E[Y]$, so $COV[X,Y] = 0$, they are irrelevant.

2.2.2. Independence cannot be derived from irrelevance

Counterexample: Given the probability distribution of X:

Table 1 Probability distribution of X

X	-1	0	1
P	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

Table 1 shows the probability distribution of X. Define $Y = X^2$, then the probability distribution of Y is:

Table 2 Probability distribution of Y

Y	0	1
P	$\frac{1}{3}$	$\frac{2}{3}$

Table 2 shows the probability distribution of Y.

Calculate $E[XY] - E[X]E[Y] = (-1) \times \frac{1}{3} + 0 \times \frac{1}{3} + 1 \times \frac{1}{3} = 0$, so $COV[X,Y] = E[(X - E[X])(Y - E[Y])] = E[XY] - -2E[X]E[Y] + E[X]E[Y]$, X, Y are not correlated. However, X, Y are not independent too, since the joint probability distribution is:

Table 3 Joint probability distribution of X and Y

Y/X	-1	0	1
1	$\frac{1}{3}$	0	$\frac{1}{3}$
0	0	$\frac{1}{3}$	0

Table 3 shows the joint probability distribution of X and Y. And $P(X = X_i, Y = Y_i) \neq P(X = X_i)P(Y = Y_i)$.

2.2.3. Independence and irrelevance are equivalent under the two-dimensional normal distribution

Proof: Necessity: It has been known that if the probability density function of X, Y is $f(X,Y) = (2\pi\delta_1\ \delta_2\sqrt{(1 - P^2\ )})^{-1}\exp[-\frac{1}{2(1-P^2)}\left(\frac{(X-\mu_1)^2}{\delta_1^2} - \frac{2\rho(X-\mu_1)(Y-\mu_2)}{\delta_1\delta_2}\right)]$ (10)

Then X, Y satisfy two-dimensional normal distribution. If $P = 0$, then $f(X,Y) = \frac{1}{2\pi\delta_1\delta_2}\exp[-\frac{1}{2}\left(\frac{(X-\mu_1)^2}{\delta_1^2} + \frac{(Y-\mu_2)^2}{\delta_2^2}\right)]$ and $f(X) = \frac{1}{2\pi\delta_1}\exp\left[\frac{(X-\mu_1)^2}{-2\delta_1^2}\right], f(Y) = \frac{1}{2\pi\delta_2}\exp[-\frac{1}{2}\left(\frac{(X-\mu_1)^2}{\delta_1^2} + \frac{(Y-\mu_2)^2}{\delta_2^2}\right)]$, then $f(X,Y) = f(X)f(Y). X, Y$ are independent.

Sufficiency: X, Y are independent, then $f(X,Y) = f(X)f(Y)$. Especially, let $X = \mu_1, Y = \mu_2$, there is $(2\pi\delta_1\ \delta_2\sqrt{(1 - P^2\ )})^{-1} = \frac{1}{2\pi\delta_1\delta_2}$, so $P = 0$, X, Y are not correlated.

3. VERIFIES THE NON-INDEPENDENCE OF STOCKS

3.1. Data

Table 4 the one-day return data from September 30, 2015, to September 2, 2021 (examples)

Date	AAPL	BABA	AMZN	TCEHY	HSBC	JPM
1-Jan-18	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
2-Jan-18	1.76%	6.30%	1.66%	4.85%	1.48%	0.94%
3-Jan-18	0.00%	0.19%	1.27%	0.53%	-0.50%	0.10%
4-Jan-18	0.46%	0.93%	0.45%	1.36%	0.21%	0.90%

5-Jan-18	1.13%	2.65%	0.60%	0.59%	-0.17%	-0.64%
……………						
31-Jan-18	0.29%	2.29%	0.90%	1.26%	-0.54%	0.49%
1-Feb-18	0.19%	-6.09%	-4.29%	-2.15%	0.41%	1.30%
2-Feb-18	-4.41%	-2.59%	2.83%	-3.17%	-1.56%	-2.24%
5-Feb-18	-2.55%	-3.69%	-2.83%	-3.28%	-3.26	-4.19%
6-Feb-18	4.11%	2.54%	3.73%	2.22%	1.48%	3.00%
7-Feb-18	-2.18%	-2.67%	-1.82%	-3.33%	-1.17%	0.68%
…………….						
28-Feb-18	-0.16%	-1.13%	0.03%	-2.27%	-1.06%	-1.60%
1-Mar-18	-1.77%	-2.25%	-1.26%	1.55%	-0.06%	-1.81%
2-Mar-18	0.68%	-1.23%	0.45%	0.25%	-0.24%	-0.10%
3-Mar-18	0.36%	1.02%	1.55%	-0.52%	-0.34%	1.52%
4-Mar-18	-0.09%	3.13%	0.92%	1.23%	0.14%	0.09%
……………						

Table 4 shows a part of the one-day return data of six companies: Apple, Alibaba, Amazon, Tencent, HSBC UC Equity and JPMorgan.

3.2.Mathematical verification process and method processing

How to judge the independence between the two stocks is still a complex problem. Four methods have been introduced to judge the independence of random variables above. However, the stock price cannot be regarded as a random variable when studying stock independence. According to the definition of a random process, a random variable is a function with different states as variables defined at a fixed time point. Although the change of stock price is random, it is a function with time as variable. There is a unique determining value corresponding to it at each time point. It is generally be called time-series data, that is, the sequence of random variables arranged in time order.

Figure 1 shows the relationship between random process and random variables and sample function.

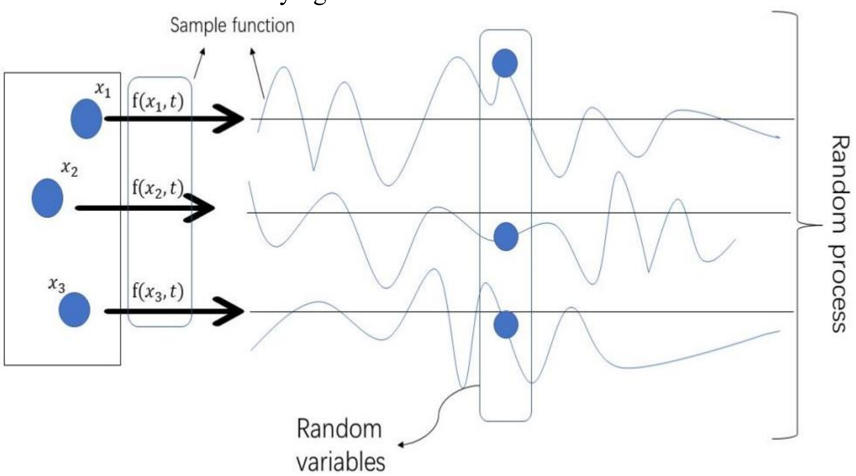

Figure 1 Representation of a stochastic process

In addition, the joint distribution function of the two stocks is still unknown. Even if their respective marginal distribution can be found to fit the normal distribution, it does not mean that their joint distribution is like the normal distribution. Therefore, whether their independence is equivalent to irrelevance is uncertain, and then the independence cannot be derived from irrelevance.

For example, suppose two independent random variables $X$and Y, where $X{\sim}(0,1)$ and the distribution of Y is

Table 5 Probability distribution of Y

Y	-1	1
P	1/2	1/2

Table 5 shows the probability distribution of Y.

Let $Z = XY$, then it can be proved that $Z{\sim}(0,1)$ but

(X, Z) does not satisfy two-dimensional normal distribution.

Therefore, in this paper, a simple method is used. It has been known that events A, B are independent if $P(A|B) = P(A)$. For the daily returns of two stocks, if event A is that the daily return of the first stock is less than one measure and event B is that the daily return of the second stock is less than the other measure, then verify whether $P(A|B)$ is equal to $P(A)$.

In this work, the daily return rate data of Apple, Tencent and Alibaba in 2018 is collected. Suppose events A, B and C represent the one-day returns of Apple, Tencent and Alibaba are larger than their own measure respectively. For different value ranges of these three random variables, take the value of top 5%, top 15%, top20%…top 50% respectively as the measure of each event. Table 6 shows the statistical result of the measure of each event:

Table 6 The measure of events A, B and C

Measure of	AAPL	TCEHY	BABA
Top 5%	2.88	3.98	3.73
Top 10%	1.83	2.74	2.87
Top 15%	1.42	2.28	2.25
Top 20%	1.09	1.82	1.82
Top 25%	0.91	1.41	1.34
Top 30%	0.74	1.17	0.92
Top 35%	0.60	0.80	0.64
Top 40%	0.25	0.39	0.37
Top 45%	0.19	0.08	0.23
Top 50%	0.00	0.00	0.00

Assume A_1 represents that the daily return of AAPL is larger than 2.88 and B_1 represents that the daily return of TCEHY is larger than 3.98, A_2 represents that the daily return of AAPL is larger than 1.83 and B_2 represents that the daily return of TCEHY is larger than 2.74, A_3….B_3…by analogy. Then calculate their probability and conditional probability respectively.

Table 7 Probability and conditional probability of events A and B

The measure	$P(A)$	$P(A|B)$	A&B
Top 5%	0.05	0.23	NO
Top 10%	0.10	0.46	NO
Top 15%	0.15	0.62	NO
Top 20%	0.20	0.51	NO
Top 25%	0.25	0.53	NO
Top 30%	0.30	0.51	NO
Top 35%	0.35	0.54	NO
Top 40%	0.40	0.46	NO
Top 45%	0.45	0.66	NO
Top 50%	0.50	0.72	NO

As shown in Table 7, events A and B are not independent in each measure.

Then, assume C_1 represents that the daily return of BABA is larger than 3.73, C_2 represents that the daily return of BABA is larger than 2.87, C_3, C_4…by analogy. Then calculate their probability and conditional probability of A respectively.

Table 8 Probability and conditional probability of events A and B

The measure of event	$P(A)$	$P(A|C)$	A&C independent
Top 5%	0.05	0.23	NO
Top 10%	0.10	0.42	NO
Top 15%	0.15	0.44	NO
Top 20%	0.20	0.42	NO
Top 25%	0.25	0.43	NO
Top 30%	0.30	0.47	NO
Top 35%	0.35	0.54	NO
Top 40%	0.40	0.63	NO
Top 45%	0.45	0.66	NO
Top 50%	0.50	0.70	NO

As shown in Table 8, events A and C are not independent in each measure too.

From table 7 and 8, the events A_i, B_i and A_i, B_i are not independent. The conditional probability is always larger than the true probability, which means the event A is more likely to happen if B and C happened. Moreover, Tencent has a greater impact on the stock movements of Apple than Alibaba since $P(A|B)$ is always larger than $P(A|C)$.

4. verifies the correlation of stocks

4.1.Scatter diagram

Simulate the scatter plot between AAPL and TCEHY to intuitively see the relationship between them.

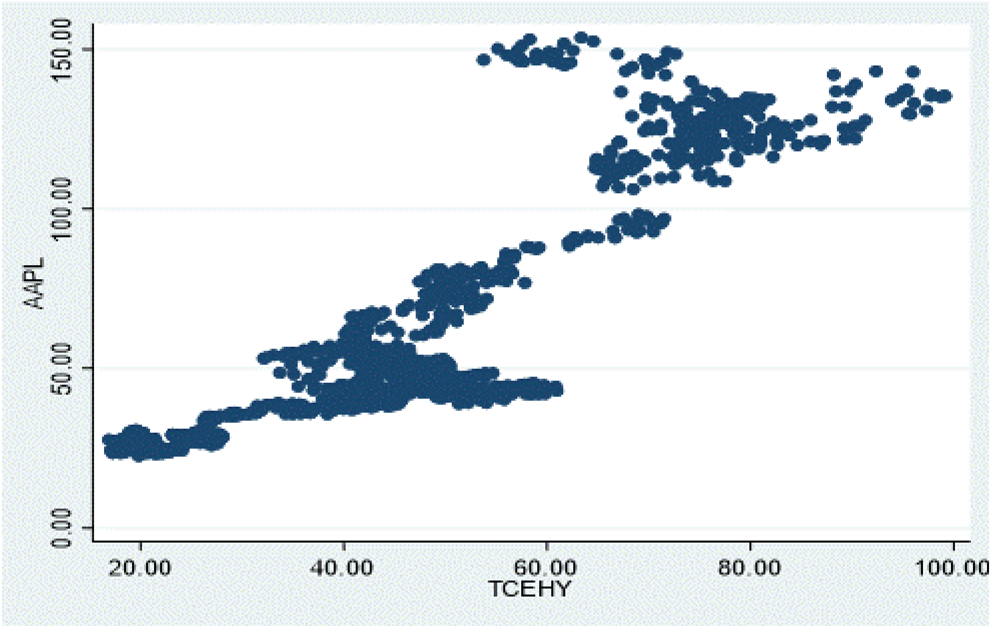

Figure 2 Scatter diagram between APPL and TCEHY

As shown in Figure 2, With the increase of TCEHY, the trend of AAPL also increases. Therefore, it can be initially seen from the scatter chart that there is a positive correlation between the two stocks.

4.2.Correlation Analysis

The absolute value of the correlation coefficient represents the magnitude of the correlation, and the positive and negative represent the direction of the correlation. After studying whether the correlation between variables is significant, the correlation between variables can be understood.

Table 9 Correlation analysis

Variables	AAPL	BABA	AMZN	TCEHY	HSBC	JPM
AAPL	1					
BABA	0.7993***	1				
AMZN	0.9586***	0.8715***	1			
TCEHY	0.8791***	0.9277***	0.9112***	1		
HSBC	-0.6312***	-0.3283***	-0.5799***	-0.3543***	1	
JPM	0.7609***	0.7266***	0.7796***	0.7974***	-0.0845***	1

Note: The correlation coefficients in the table are ***, ** means there is more than 99% probability of correlation, ** means there is more than 95% probability of correlation, * means there is more than 90% probability of correlation, no asterisk means there is not enough probability of correlation.

As shown in Table 9, the correlation coefficients between AAPL and BABA, AMZN, TCEHY, HSBC and JPM are 0.7993, 0.9586, 0.8791, -0.6312 and 0.7609, respectively. There is a significant correlation between AAPL and BABA, AMZN, TCEHY, HSBC and JPM, and there is a significant positive correlation between AAPL and BABA, AMZN, TCEHY and JPM, that is, all change in the same direction. AAPL is significantly negatively correlated with HSBC, that is, there is a reverse change. The correlation coefficients of BABA with AMZN, TCEHY, HSBC and JPM were 0.8715, 0.9277, -0.3283 and 0.7266, respectively. In addition to the significant negative correlation with HSBC, BABA was significantly positively correlated with AMZN, TCEHY and JPM. The correlation coefficients of AMZN with TCEHY, HSBC and JPM are 0.9112, -0.5799 and 0.7796, respectively. In addition to the significant negative correlation with HSBC, AMZN is significantly positively correlated with TCEHY and JPM. The correlation coefficient between TCEHY and HSBC is -0.3543, a significant negative correlation. The correlation coefficient between TCEHY and JPM is -0.0845, a significant negative correlation. Among them, the correlation coefficient is the largest between AAPL and AMZN, and the correlation coefficient is the smallest between HSBC and JPM.

5. CONCLUSION

The above analysis shows that the stock is independent and correlated, and the obtained data shows that the correlation is very significant. Every stock portfolio is exposed. If there is a significant correlation between stocks, is to put the eggs in one basket, so once the correlation of the portfolio is very strong, it is not a true market hot spot (even in the field of market, such as Japan's Fukushima nuclear blast, so will cause nuclear power company has very strong correlations with nuclear power equipment company). However, due to the adverse impact of nuclear power in the market, the purchased portfolio is facing greater risks, and it is difficult to achieve market gains or even losses. Even if the above data are stocks with a high beta coefficient, due to their correlation, it is likely that the strength of the market gives a weak stock, just because of its strong correlation, and at the same time by a factor interference.

REFERENCES

[1] Forecasting International Equity Correlations[J]. Financial Analysts Journal,1994,50(6):

[2] Yishuai Tian1, Boying Lv1, Botao Liu1. Studying the Correlation of Stocks via Copula Function[J]. Journal of Theoretical and Applied Economics,2019,4:

[3] Pan Wenrong. The Research on Correlation of Stock Markets among China Mainland,Hongkong and America Based on Wavelet Analysis[C]//Proceedings of 2012 International Conference on Education Reform and Management Innovation(ERMI 2012) Volume 2,2012:62-67.

[4] Monica Singhania,Shachi Prakash. Volatility and cross correlations of stock markets in SAARC nations[J]. South Asian journal of global business research,2014,3(2):

[5] Ma Shuanghong, Chen Jinshu. Journal of Lanzhou University of technology,2019,45(06):169-172.

[6] Friedberg S.H., Insel, A.J. Spence, L.E. (2018) Inner Product Spaces. In: Jeff, W. (Eds.), Linear Algebra. Pearson Education, Inc., New Jersey. pp358-360.