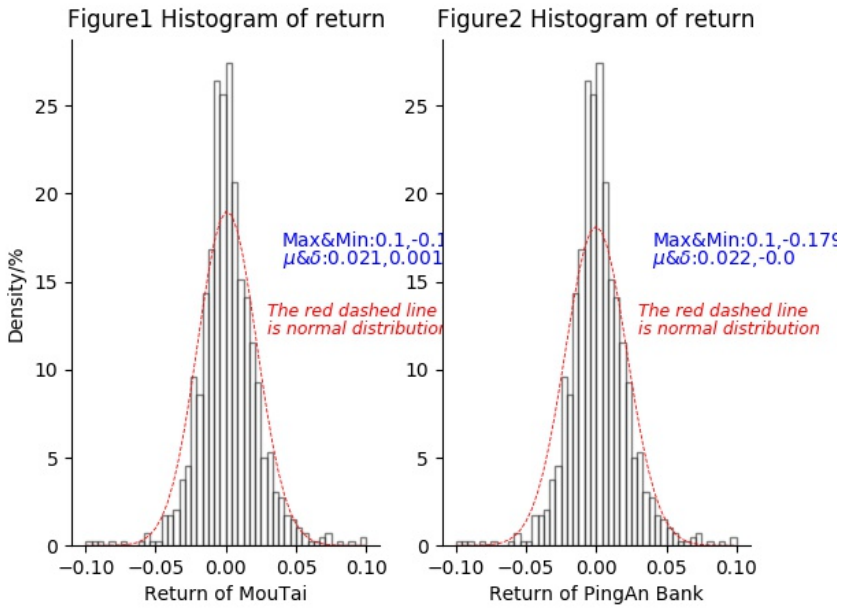
1. Description of data

This report aims to describe and analyze the return of MouTai (stock1)and PingAn Bank (stock2) collected from Tushare from year 2015 to 2018. The data contains columns regarding trade_date, ts_code, open, high, low, close,p ct_chg, change, volume, amout. There is no 'NA' value.Sample size of each stock is 955, so the number of valid return value is 954.The number of valid return values in 2017 and 2018 are 244 and 222 respectively, recoded
as sample0 and sample1.

a. Single variable analysis

Table 1. Summary statistics

	Count	mean	std	Min	25%	50%	75%	Max
MouTai	954	0.0013	0.02078	-0.09999	-0.00969	0.00020	0.01157	0.10000
PingAn Bank	954	-0.00021	0.02202	-0.17912	-0.00817	0	0.00775	0.10000



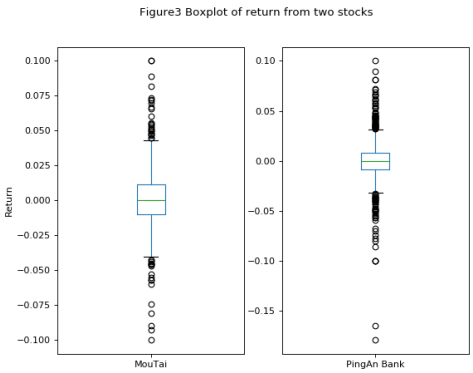
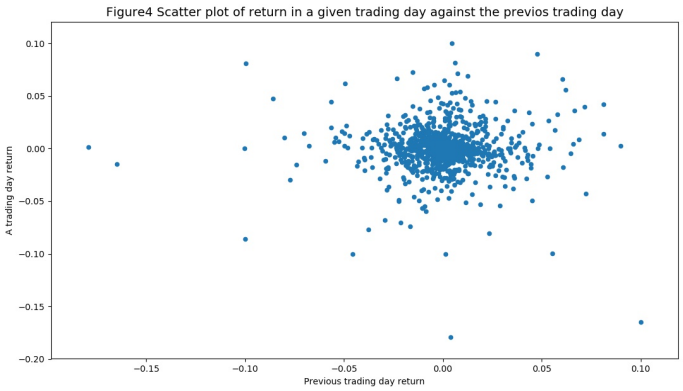


Table 1 suggests the standard deviations (0.0207 and 0.0220) and means are quite similar for both stocks. The IQRs value (0.0116+0.0097=0.0213 and 0.0077+0.0082=0.0159) of stock1 is larger ,as well as the range value (0.1000+0.1000=0.2 and 0.1000+0.1791=0.2791) of stock2 is bigger, which can be clearly seen in Figure3. There is a big difference of Kurtosis for the two stocks (3.371 and 10.663), so we conclude that the stock1 has a greater dispersion than stock2. The skewness for stock1 is 0.2240 and -0.9901 for stock2. In another words, stock1 is right-biased and stock2 left-biased.

This shows that at the same risk level,Investing in stock1 has more opportunity to get reward while investing in stock2 will have a greater probability of losing money in the statistical time periods.

b. Multivariable analysis



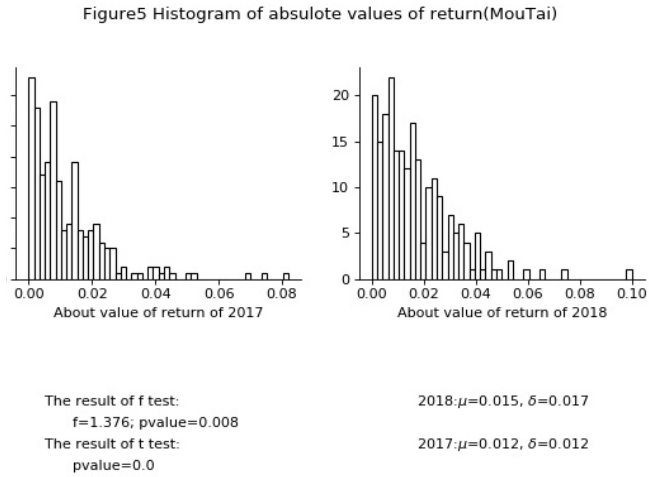


Figure4 shows the relationship of a trading day and the previous trading day.We use linear regression model(OLS) to fit it. The Mathematical expression is: y=ax+b, a is coefficient, b is interception.The coefficient of linear regression is -0.0053, whose absolute value approximately equals to 0. So the linear correlation between them is negligible.
Hence, it's unwise to make trading decisions only depending on the return of previous trading day .

2.Test

We take natural logarithm (data transformed = ln(data)) for both two samples.The Shapiro–Wilk tests return values approximately equals to 1. We hence assume both samples to be normally distributed.
The F-test(p=0.0079) is strongly against the H0 hypothesis in 95% confidence interval.
Then 2 sample T-test is used to test whether the mean of two samples are eqal. The return value(p=5.5e-5) strongly supports H1 hypothesis, that is to say the mean is not equal.
Figure5 shows the distribution of two samples. The mean and standard deviation of sample1 is greater than sample0. And based on the fact that as long as the the sample size is big enough, the sample can give enough information of population, so the conclusion is that market volatility is increased in 2018.