

Problem 1

Data Collection

The data are all downloaded from CSMAR as required by the problem, namely the closing index over 2003/12/31 to 2023/12/31 (the data is 2003 is for calculating the return in 2004/01).

Question (a)

Claim

The downloaded data are daily returns, but the question asks to derive monthly returns. Here, I use the daily returns in the last day of each month to represent the monthly returns. As this is not specifically identified in the question, I believe all methods including taking average and last-day-return work.

Steps

Data Manipulation. I convert the ‘date’ column from object to datetime format and then changing it to a monthly period, as it works better for time-series data.

Return Calculation. First the data with index code “000300” (standing for CSI 300) is taken out for calculation. The monthly return is derived by `.pct_change()` method (whose algorithm is exactly the same as the provided formula). Then the 2005-04 data is dropped as this is the first index data, with no return.

Results

Summary Statistics. By using library `scipy.stats`, we can get the data required.

Table 1: Summary Statistics

Mean	Standard deviation	Skewness	Kurtosis
0.0090864533	0.0807116815	0.0152174982	1.4498566009

Question (b)

The graph is shown in the next page.

Question (c)

By analyzing the figure data from (b), the skewness value is relatively large but near zero, suggesting that the data distribution is nearly symmetrical, which is in accordance with normal distribution. The kurtosis value is notably lower than that of a normal distribution, suggesting that the data exhibits a heavy-tailed distribution.

Being more rigorous, based on the Shapiro-Wilk test, the p-value is less than the typical signifi-

cance level of 0.05. This indicates strong evidence against the null hypothesis that the data follows a normal distribution. Therefore, returns of the CSI300 index is NOT normally distributed.

Table 2: Shapiro-Wilk Test Statistics

Shapiro-Wilk test statistic	p-value
0.971137	0.000155

Graph. Histogram as shown below, from 2005 – 2023.

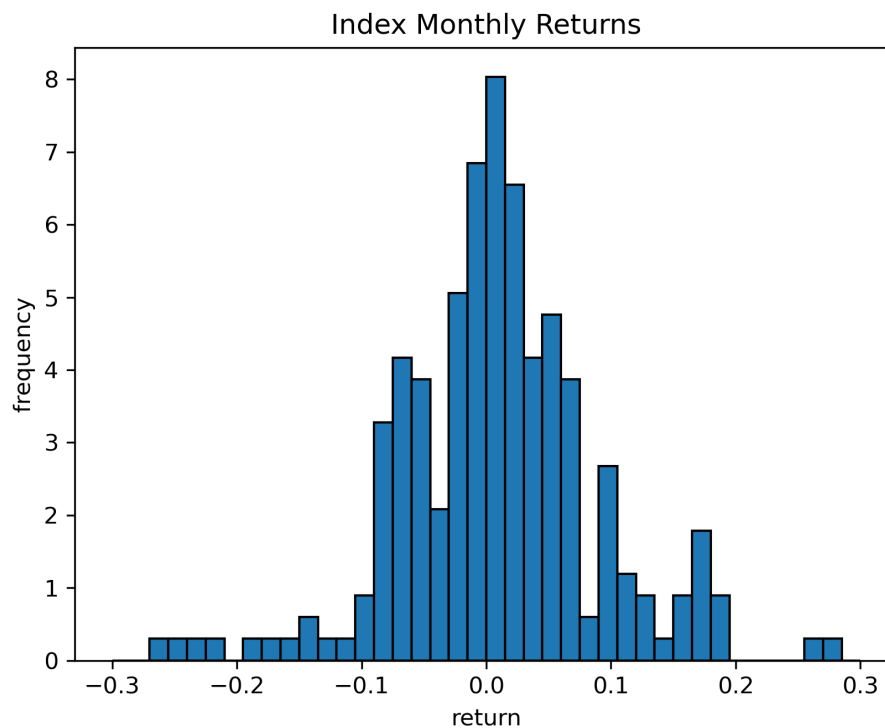


Figure 1: Histogram for CSI 300 monthly returns

Problem 2

Data Collection

The data are all downloaded from CSMAR as required, namely Returns Without Cash Dividend Reinvested from 2017-01 to 2022-53 (YYYY-WW).

As required, the time period should be divided to three parts. In this question, 2 years will be one regression period, namely 2017-2018 (first regression), 2019-2020 (second regression), 2021-2022 (third regression).

xf Steps

(a) *Data Manipulation*. We can simply use `.concat()` to connect the two .csv files downloaded. Besides, I take out market type 1, 4, 64, which are the codes for A-share mainboard as required.

(b) *Market Return*. Use `.groupby('week')` and `.mean()` to calculate the market return by covering all stock returns.

(c) *Data Manipulation*. Just load the data, and align the 'week' data type and structure using some string methods.

(d) *Process*.

Replicate Table 2 from the referenced paper.

Firstly, regress every stock return on the market return, so we can get the beta in single factor models in the first time period. Then, after sorting and ranking beta, we divide the stocks into 10 groups. We construct 10 portfolios by combining all the stocks in each group. Then we calculate the difference among return and risk-free rate. Finally we can get regression result in Table 2 using the second period data.

Table 3: (**Table 2 in paper**) Time series regression results of the second period of portfolio

Group	α_p	alpha t-value	alpha p-value	β_p	beta t-value	beta p-value	R^2
1.0	-0.00128	-1.98316	0.05009	0.79934	39.08258	0.0	0.93855
2.0	-5e-05	-0.09287	0.92619	0.88132	52.52708	0.0	0.96502
3.0	0.00021	0.49203	0.62378	0.89614	65.38391	0.0	0.97714
4.0	0.00043	0.99721	0.32107	0.94753	69.04839	0.0	0.97946
5.0	-1e-05	-0.03337	0.97345	1.00091	98.40713	0.0	0.98978
6.0	-7e-05	-0.16987	0.86545	1.04398	84.57895	0.0	0.98621
7.0	-0.0002	-0.57567	0.56613	1.04015	92.02092	0.0	0.98833
8.0	-0.00013	-0.27314	0.78531	1.10491	73.84576	0.0	0.98199
9.0	0.0003	0.67425	0.50171	1.08707	75.82531	0.0	0.9829
10.0	-0.00095	-1.66571	0.0989	1.15054	63.23312	0.0	0.9756

Note: all the data are round to five decimals.

Results. By comparing with the paper given, we find the results are a little different:

1. The Beta values across all portfolios are fairly consistent, typically around 1. Moreover, the p values are consistently low, showing the significant impact of market returns on stock returns. This aligns with the findings of the referenced paper.

2. Initially, the R-squared values are consistently high, indicating that our regression model effectively captures the variance in stock returns. Furthermore, there's no notable increase in R-squared values with higher Beta values, implying that factors beyond systematic risk influence stock

returns. This aligns with the referenced paper.

3. However, the alpha values we computed are notably small, approaching zero. Based on the t value and p value, we lack sufficient evidence to reject the null hypothesis that alpha equals zero. This does NOT align with the referenced paper.

Replicate Table 3 from the referenced paper.

I firstly take out the third period data from 2021-2022. Then calculate the average excess return, and merge corresponding beta to the dataframe. By repeating the regression in the paper, we can get:

Table 4: (**Table 3 in paper**) Cross-sectional regression results of the third period of portfolio

γ_0	γ_1	$\gamma_0 t - value$	$\gamma_1 t - value$	R^2	F-statistic	P
-0.00107	0.00266	-1.05138	2.60449	0.45885	6.78338	0.0314

Note: all the data are round to five decimals.

Results. Based on the table above, the following conclusions can be drawn:

1. The R-squared value stands at only 0.45885, similar to the explaining power in the paper.
2. The value of γ_1 is 0.00266, and t-statistics is greater than 2, p-value is 0.031 (smaller than 0.05) as well, showing statistically significant positive correlation between returns and systematic risk. This suggests that returns tend to rise with increasing systematic risk, consistent with the principles of the CAPM model, and aligns with the conclusions drawn in the reference paper.
3. However, the alpha values obtained are notably small, approaching zero. The t-value is rather close to 0 (greater than -2), p-value is 0.324, so we fail to find sufficient evidence to reject the null hypothesis that alpha equals zero. This contradicts the assertions in the reference paper.

Possible reason: Chinese stock market has undergone rapid development in these years after the paper published, becoming substantially more efficient over time. The establishment and enhancement of the multi-level market system, alongside reforms in the new stock issuance system, have contributed to the maturation of China's securities market. Consequently, the equilibrium relationship between risk and return in securities investment is increasingly manifested, leading to a diminishing impact of non-systematic risks on pricing. As a result, alpha is not significantly different from zero.