Problem 1

Monthly CSI 300 Index (i.e., Month_RT) = Closing index of the last trading day of the month (i.e., clsindex)/ Closing index of the last trading day of last month (i.e., L. clsindex) – 1

(a) Here we offer the summary statistics for CSI300 monthly index returns.

Summary statistics of Monthly CSI300 Index Returns

Observations	Mean	Standard deviation	Skewness	Kurtosis
3,310	0.0080958	0.0821718	0.0406479	4.520758

Table 1 Summary statistics for CSI 300 Monthly Returns

(b) Here we plot a histogram for CSI300 Index Monthly Returns, dividing the data we get into 100 intervals. Moreover, to facilitate comparison with the normal distribution, we also draw the figure of the pdf of a normal distribution with same mean and standard deviation of CSI300 Index Monthly Returns (i.e., the green line).

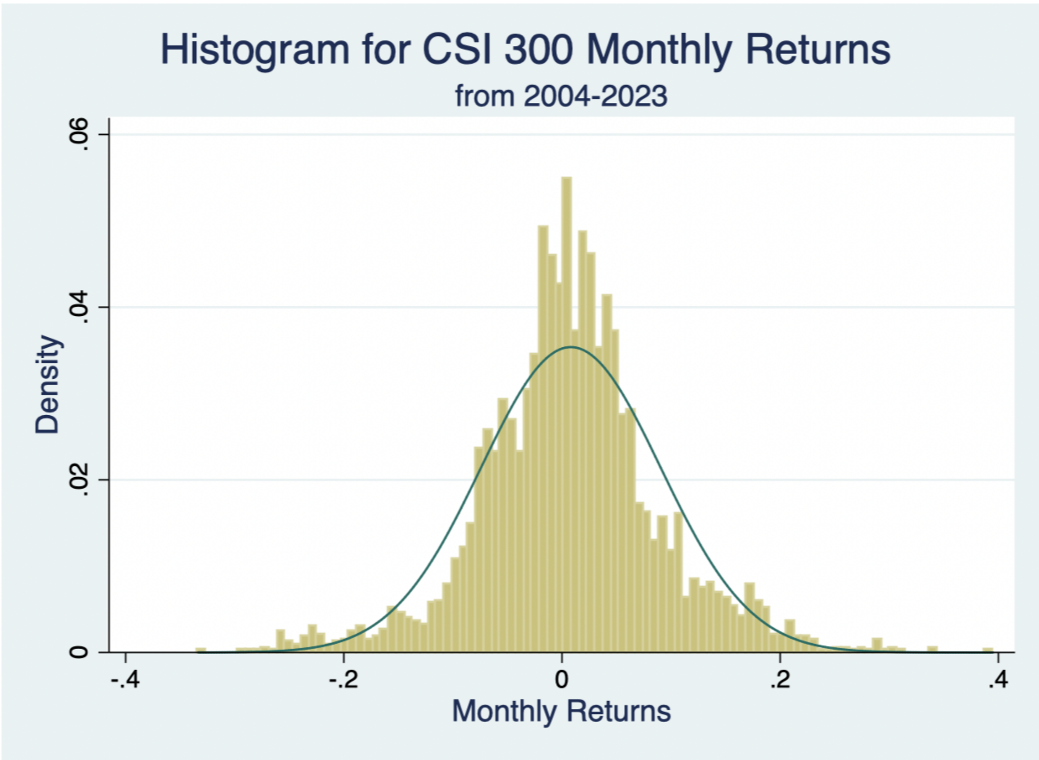

Figure 1. Histogram for CSI300 Monthly Return

(c) We use three different ways to compare the distribution of the monthly returns of CSI 300 index with normal distribution with same mean and standard deviation of CSI300 Index Monthly Returns.

i. Summary statistics

To get the summary statistics of normal distribution with same mean and standard deviation of CSI300 Index Monthly Returns, we first generate 3,310 random number which all follow this distribution. Here is the summary statistics we get.

normal_data_custom

	Percentiles	Smallest		
1%	-.1827076	-.2495025		
5%	-.1248079	-.2449654		
10%	-.1001332	-.2386767	Obs	3,310
25%	-.0467678	-.230902	Sum of Wgt.	3,310
50%	.0065595		Mean	.0066268
		Largest	Std. Dev.	.0808017
75%	.06338	.2421119		
90%	.108277	.2451802	Variance	.0065289
95%	.1355126	.2456086	Skewness	-.0295458
99%	.1929935	.3313658	Kurtosis	2.949786

Table 2. Summary statistics of the normal distribution

Compared with it, the Skewness and Kurtosis of Month_RT is much larger. So intuitively speaking, the probability density function CSI300 Index Monthly Returns may not follow a normal distribution.

ii. Plot

From the histogram plot we get in (b), we can clearly obvious that Month_RT is more concentrate around the mean compared to normal distribution.

iii. Normality testing

To give more persuasive evidence to our intuition, we conduct Shapiro-Wilk Test. Here is the result we get. Since the p-value < 0.05, we can draw significant verdict that the probability density function CSI300 Index Monthly Returns doesn't follow normal distribution.

```
. swilk Month_Rt
```

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
Month_Rt	3,310	0.97877	39.739	9.534	0.00000

Note: The normal approximation to the sampling distribution of W' is valid for 4<=n<=2000.

Table 3. The result of Shaprio-Swilk Test

In conclusion, returns of CSI300 index doesn't follow normal distribution.

Problem2

In this problem, we try to imitate Chen et al. (2019), using the weekly returns of all A-share mainboards stock in Chinese stock market from 2017 to2022 to conduct a simplified version of BJS test.

After concatenating all necessary data, we derive the market returns as the average returns of all stocks of the same time. We also also calculate the individual risk premium and market risk premium. Using 33 weeks as a period, we split them into three groups (i.e., P1, P2, P3) based on their different periods.

i. Time Series Tests for CAPM

1) Calculate Individual Stock βs

We first treat on the P1 panel data set.

<u>Regression Model:</u> $r_{i,t} = \alpha_i + \beta_i * r_{m,t} + \mu_i$ (i.e., assuming E(μi,t) = 0)

Using the above one-factor model, we conduct the 3,571 stock-level time series regressions for each stock. We totally get 3,571 estimated β_is.

2) Construct Stock Portfolios

We next switch to P2 panel data set. We merge the β_is we get in (1) into this data set. In an effort to diversify part of the unsystematic risks, we construct ten stock portfolios based on their β_i, and calculate the corresponding portfolio returns as the mean value of returns of all stocks in that portfolio. We calculate the risk premium of each portfolio :

port_prm = portfolio_return - risk_free_return

<u>Regression Model:</u> $r_{p,t} – r_{f,t} = \alpha_p + \beta_p * (r_{m,t} – r_{f,t}) + \mu_i$ (i.e., assuming E(μi,t) = 0)

Using the above model, we conduct 10 portfolio-level time-series regressions and get 10 estimated β_ps for each portfolio. Here we list 0the results of these regressions.

Time regression results of portfolio return premium						
Stock code	Obs	α_p	t-value	β_p	t-statistics	R^2
Portfolio 1	102	-0.00117	-1.68	0.7862***	36.56	0.9304
		(0.0007)		(.02151)		
Portfolio 2	102	-0.000464	-0.79	0.8744***	48.26	0.9588

		(0.00058)		(.01812)		
Portfolio 3	102	0.00037 (0.00047)	0.78	0.8982*** (0.01445)	62.17	0.9748
Portfolio 4	102	0.000144 (0.00046)	-0.31	0.9371*** (0.01419)	66.03	0.9776
Portfolio 5	102	-0.00019 (0.00036)	-0.35	1.007*** (0.0111)	90.81	0.9880
Portfolio 6	102	-0.00031 (0.00038)	-0.80	1.006*** (0.0118)	85.55	0.9865
Portfolio 7	102	-0.000214 (0.00041)	-0.53	1.057*** (0.0125)	84.79	0.9863
Portfolio 8	102	0.0003 (0.00042)	0.71	1.091*** (0.0129)	84.48	0.9862
Portfolio 9	102	-0.000163 (0.00049)	-0.33	1.098*** (0.0151)	72.92	0.9815
Portfolio 10	102	0.00029 (0.00075)	0.30	1.142*** (0.0231)	49.44	0.9607

Table 4. Time regression results of portfolio return premium. Standard errors are in parentheses under coefficients. The portfolio coefficients is statistically significant at the *5% level, **1% level or *** 0.1% level.

Table 4 shows that the β_p values of the portfolios are similar, mostly around 1, and their significance levels are basically small, which indicates that stock returns are significantly influenced by stock market returns. The null hypothesis is not rejected since none of the portfolios have significant α_p. The coefficient of determination R2 seems to increase with the increase of βp from portfolio 1 to portfolio 5, but the absence of this tendency in portfolio 6 to portfolio 10 indicates that stock returns are affected by other factors besides systematic risk.

ii. CAPM Cross-sectional Regression

We finally switch to P3 panel data set. Like what we have down in P2 data set, we merge the β_is we get in i.1) into this data set and construct 10 stock portfolios based on them. Next, we calculate the average portfolio returns and average portfolio premiums over the full P3 period. Then we merge the 10 β_ps we get in i.2) into this data set. After duplicating the repeated data, we get 10 observations listed in Table 5.

Portfolio	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Avg_rp_rf	0.0012	0.001	0.001	0.0015	0.00084	0.0014	0.0016	0.0012	0.0012	0.002
β_p	0.786	0.874	0.898	0.947	1.007	1.006	1.057	1.091	1.098	1.142

Table 5. 10 observations used for cross-section regression

<u>Regression Model</u>: $r_{p,t} - r_{f,t} = \gamma 0 + \gamma 1 * \beta_p + \mu_p$

Using the 10 observations in Table 5 and the model above, we conduct a cross-sectional regression. Here we offer the regression result in Table 6.

	γ0	*γ1*	*R^2*	*F-statistics*	*P-value*
Coefficients	0.0004863	0.0008082	0.1248	1.14	0.3167
t-statistics	0.65	1.07			

Table 6 Cross-sectional regression results of the third period of stock portfolios

From Table 6, the R2 is only 0.1248, which is an average fit while γ1 = 0.0008082 The statistics of γ1 shows that the return is significantly positively correlated with the systematic risk, indicating that the return increases with the risk, which is consistent with the CAPM model. The constant γ0 is very close to 0, indicating that factors other than the systematic risk are not significant.

iii. Conclusion

From above analysis, we can see a certain positive linear correlation between returns and risks, which is basically consistent with the conclusions of the CAPM model. However, we can’t exclude the possibility of the existence of unsystematic risk that may also influence the stock pricing and consequently lead to the inefficiency of the market.