**Assignment 2 – t54zheng**

**Summary of conclusions from the code**

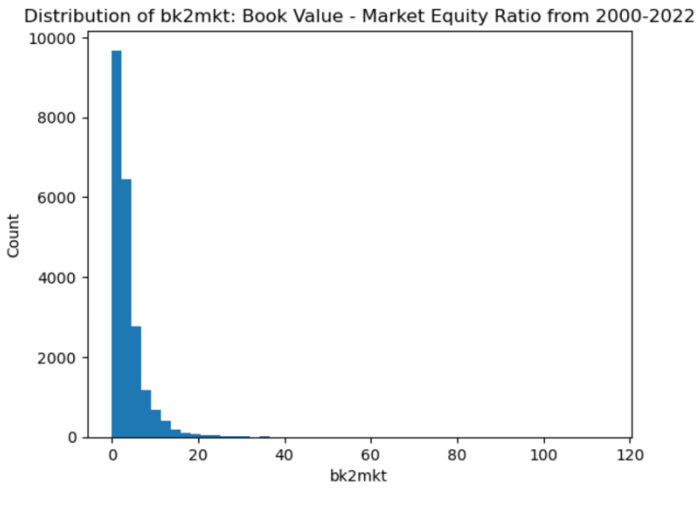
**Q4 – Creating our Additional Variables**

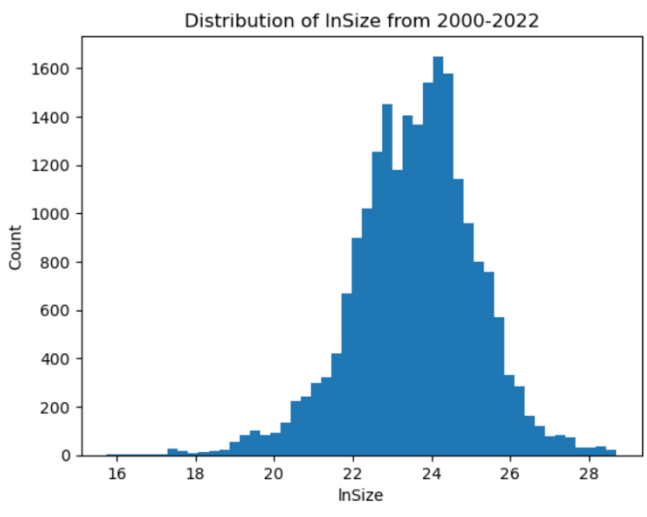
**Data Cleanup**

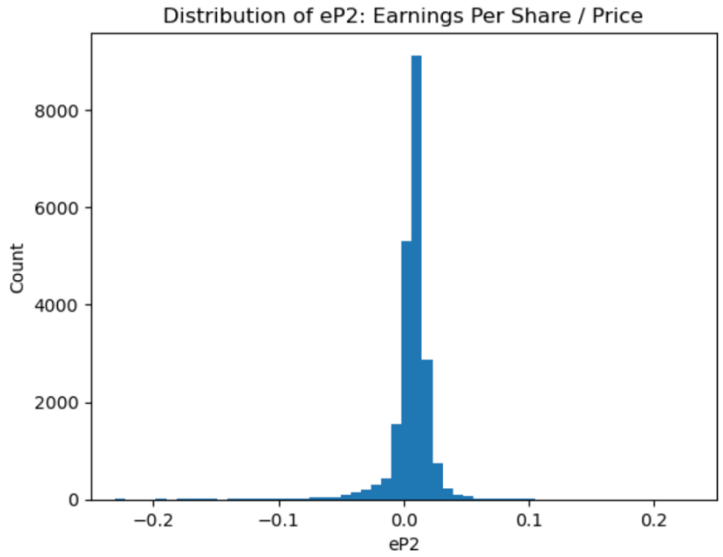
To clean up our merged data, the steps that were implemented in our code were:

1. Correcting negative stock prices by changing them to be equal to their absolute values.
2. Removing duplicate values
3. Deleting rows with null values
4. Correcting Unit Differences between CRSP/Compustat
    - ATQ, CEQQ, IBQ, SALEQ needed to be multiplied by one million
    - Shares Outstanding (SHROUT) needed to be multiplied by one thousand

**Visualizing our Distributions**



**eP1 vs eP2**

The two measures both illustrate how the market values the company's quarterly earnings. The difference is that eP1, which is calculated as total income / total market equity, represents the ratio at which the entire company is valued compared to its earnings. On the other hand, eP2 represents how the company's earnings is valued at a per-share basis (Earnings per Share / Price).

Statistically, we may say that the latter (eP2) is the better comparison because it has a smaller relative standard deviation, and because it describes an individual stock better (which is what we are trying to model)

**Question 5 – Summary Statistics for Our Variables**

| | lnSize | bk2mkt | eP1 | eP2 |
|---|---|---|---|---|
| **mean** | 23.5785 | 3.7413 | 0.0041 | 0.0026 |
| **median** | 23.6671 | 2.5860 | 0.0086 | 0.0085 |
| **variance** | 2.4713 | 17.0004 | 0.0126 | 0.0237 |
| **standard deviation** | 1.5720 | 4.1232 | 0.1121 | 0.1541 |
| **5th Percentile** | 20.8903 | 0.4671 | -0.0190 | -0.0189 |
| **25th Percentile** | 22.6312 | 1.4116 | 0.0034 | 0.0034 |
| **50th Percentile** | 23.6671 | 2.5860 | 0.0086 | 0.0085 |
| **75th Percentile** | 24.5674 | 4.6843 | 0.0136 | 0.0132 |
| **95th Percentile** | 25.9738 | 10.7655 | 0.0267 | 0.0242 |

**Q6 – Removing Outliers**

We will choose to winsorize our data, since the validity of our data should be good. I.e. the data that we have downloaded is not incorrect, but just reflects the notion that surprise stock events are a reality and should be reflected in the data. So, instead of truncating them to be removed, we just winsorize them so that they are included in the data, but we do not influence the skew of the data too much.

| | lnSize (winsorized) | bk2mkt (winsorized) | eP1 (winsorized) | eP2 (winsorized) |
|---|---|---|---|---|
| **mean** | 23.5775 | 3.6494 | 0.0059 | 0.0049 |
| **median** | 23.6671 | 2.5860 | 0.0086 | 0.0085 |
| **variance** | 2.4540 | 12.2252 | 0.0023 | 0.0039 |
| **standard deviation** | 1.5665 | 3.4965 | 0.0476 | 0.0621 |
| **5th Percentile** | 20.8910 | 0.4671 | -0.0190 | -0.0186 |
| **25th Percentile** | 22.6312 | 1.4116 | 0.0034 | 0.0034 |
| **50th Percentile** | 23.6671 | 2.5860 | 0.0086 | 0.0085 |
| **75th Percentile** | 24.5674 | 4.6843 | 0.0136 | 0.0132 |
| **95th Percentile** | 25.9738 | 10.6606 | 0.0267 | 0.0242 |

**Q7 – Testing Asset Pricing Model Validity**

a) Below are the respective t-statistics for each of the lambdas for our model betas

| | 0 |
|---|---|
| beta_t_stat | -1.15 |
| beta_p_value | 0.25 |
| ff3_beta_t_stat | -1.74 |
| ff3_beta_p_value | 0.08 |
| smb_beta_t_stat | -0.01 |
| smb_beta_p_value | 0.99 |
| hml_beta_t_stat | 1.31 |
| hml_beta_p_value | 0.19 |

- We see that for CAPM, we have a t-stat if 1.15, which is not enough to reject our null hypothesis of beta being significant in modelling returns
- However, for our FF values it is a different story.
- Our FF3 beta has a higher t-stat if 1.74, which is enough for a 90% confidence interval

Other statistics have weak levels of significance

From these interpretations, we can conclude that two factors: fama-french market and value premium (HML) to a lesser extent, are priced in the model at a significant level.

**b)** Here are our graphs that replicate the techniques shown in the class slides. We see that CAPM's predictions are slightly worse than the predictions from Fama-French 3-factor