HARVARD UNIVERSITY

STAT139 FINAL PROJECT REPORT

In-depth research on financial models in Chinese capital market

---

Author:
Chengcheng Qin
Haosu Tang
Chenchen Zhang

A project report submitted in fulfillment of the requirements of STAT 139: Statistical Sleuthing Through Linear Models

December 9, 2015

Abstract

In-depth research on financial models in Chinese capital market

by Chengcheng Qin
Haosu Tang
Chenchen Zhang

The required rate of return is an important factor for investors to benchmark the risk of a piece of asset. The celebrated CAPM model takes the sensitivity with regard to non-divisible risk and proposes a simple linear relationship. Fama-French model builds upon the original CAPM model by adding additional factors such as size, value, further profitability and investment patterns. The success implemention of these models in US market motivates us to extend the experiment to emerging market such as the Chinese market. We used multiple linear regression to fit CAPM, Fama-French 3-factor and 5-factor models. With all the constraints and strict regulations in Chinese market, as well as its inefficiency and high volatility, our analysis shows these models, in general, interpolate and explain the required rate return of portfolios in Chinese market fairly well. We further explored the difference between average return of portfolios constructed based on capital size and momentum through t-test, and found our strategies yield a return slightly higher than the market, but not significant.

Contents

List of Figures

Chapter 1

Introduction

In financial markets, a variety of models have been invented to determine the theoretically appropriate required rate of return for a capital asset. By using a proper model to help investors seeking arbitrage opportunities, capital pricing plays an important role as it sets up the minimum acceptable return for investors. One major class of pricing model we plan to investigate here is Capital Asset Pricing Model (CAPM), more specifically, this project focuses on exploring two famous extensions of CAPM, the Fama-French Three Factor Model and the Fama-French Five Factor Model proposed by Eugene Fama and Kenneth French.

Past efforts have been devoted to statistical sleuthing on these models. Attempts have been made to prove the explanatory power of these models, as well as identify potential limitations and improvements by either adding more predictor variables or removing less significant ones.

A substantial amount of research has been done in U.S. stock market to validate the Fama-French model, most of which give significant results, because of the mature and well-regulated nature of U.S. equity market. The success of using Fama-French model to explain portfolio return in U.S. market makes us wonder if we could apply the same methodology to Chinese market. In the first part of the project, we try to examine these models using data in Chinese stock market and identify significant predictor variables by applying simple and multiple linear regressions. Based on the outcome, we then used T-tests to show whether the return of the portfolios constructed by our strategies is significantly higher than zero and further higher than the market return.

In order to quantitatively examine the Fama-French models, we firstly collected, cleaned and combined data (data source: CSMar, Wind, Yahoo API) from Chinese A-share stock market, excluding ST and PT (Special Treatment and Particular Transfer). Because of the inconsistency and incompleteness of raw data, a huge effort has been dedicated in the early stage of the project to purge and streamline the raw data into proper dependent variables and explanatory variables that we can work with. For instance, in order to calculate the difference between market return and risk free return, we utilized the monthly return of CSI 300 Index and quarterly deposit interest rate.

Chapter 2

Background

2.1 Asset Pricing Models

2.1.1 One-factor Model

Capital Asset Pricing Model (CAPM) (Lintner, 1965) (Sharpe, 1964) is based on the idea that investors should be compensated in two ways: time value of money and risk. The time value of money is represented by the risk-free (R_f) rate and compensates the investors for placing money in any investment over a period of time. The other half represents risk and calculates the amount of compensation the investor needs for taking on additional risk. This is calculated by taking a risk measure (beta) that compares the returns of the asset to the market over a period of time and to the market premium ($R_m - R_f$):

$$R_{it} - R_{Ft} = \alpha_i + \beta_i(R_{Mt} - R_{Ft})$$

2.1.2 Fama-French Three Factor Model

The Fama-French three-factor model is an empirical asset pricing model. Standard asset pricing models work forward, from assumptions about investor tastes and portfolio opportunities to predictions about how risk should be measured and the relation between risk and expected return. Empirical asset pricing models work backward. They take as given the patterns in average returns, and propose models to capture them. The Fama-French three-factor model is designed to capture the relation between average return and Size (market capitalization, price times shares outstanding) and the relation between average return and price ratios like the book-to-market ratio, which were the two well-known patterns in average returns at the time of Fama-French 1993 paper (Fama, 1993).

$$R_{it} - R_{Ft} = \alpha_i + \beta_i(R_{Mt} - R_{Ft}) + s_i \cdot SMB_t + h_i \cdot HML_t + e_{it}$$

Here R_{it} is the portfolio's expected rate of return, R_{Ff} is the risk-free return rate, and R_{Mt} is the return of the market portfolio. The "three factor" β is analogous to the classical β but not equal to it, since there are now two additional factors to do some of the work. SMB stands for "Small [market capitalization] Minus Big" and HML for "High [book-to-market ratio] Minus Low"; they measure the historic excess returns of small caps over big caps and of value stocks over growth stocks.

2.1.3 Fama French Five Factor Model

In Fama-French 2013 paper, two more variables are added into it: RMWt is the difference between the returns on diversified portfolios of stocks with robust and weak profitability, and CMAt is the difference between the returns on diversified portfolios of low and high investment stocks, which we call conservative and aggressive. If the sensitivities to the five factors, β_i, s_i, h_i, r_i, and c_i, capture all variation in expected returns, the intercept ai is zero for all securities and portfolios i.

$$R_{it} - R_{Ft} = \alpha_i + \beta_i(R_{Mt} - R_{Ft}) + s_i \cdot SMB_t + h_i \cdot HML_t + r_i \cdot RMW_t + c_i \cdot CMA_t + e_{it}$$

2.2 Data Processing

2.2.1 Data source

Returns of all the stocks in Chinese A-share Market since 2009

1. Why monthly return: to avoid trading noise
2. Why Chinese stocks: since Fama-French three-factor model works well in American markets, we'd like to examine it in Chinese stock markets.
3. Why after 2009: after the financial crisis, there would not be a lot of noise in markets
4. Why A-share: plenty of historical data

2.2.2 Data purging

Before running linear regression in R, a critical task is to compile the explanatory variables and dependent variable from raw data. In the full factor Fama-French model, five explanatory variables have been taken into consideration:

1. $R_m - R_f$: the excess return of market over interest rate
2. SMB: the excess average return of small-cap stocks over big-cap stocks
3. HML: the excess average return of stocks with high book-to-market ratio over stocks with low book-to-market ratio
4. RMW: the excess average return of stocks with robust profitability over stocks with weak profitability
5. CMA: the excess average return of stocks with conservative investment over stocks with aggressive investment

In order to calculate the last four variables, we first ranked all the stocks according to market capitalization, book-to-market ratio (book value divided by market value), earnings per share, and fixed asset growth ratio respectively, and then take the difference between average return of the n largest stocks and n smallest stocks in each category.

For example, in order to calculate SMB (Small minus Big cap), we need to take the difference between average return of the 20 stocks with smallest capitals and the 20 stocks with the largest capitals. The other three variables can be calculated in a similar approach.Eventually, we randomly chose 20 stocks to formulate the portfolio of interest to construct dependent variables by subtracting risk-free rate from the average return. As for t-tests, we examined whether our strategies lead to a higher return. One strategy is to long stocks with the smallest capital and short stocks with the largest capital in the last month, the other is to long stocks with the highest return and short stocks with the lowest return in the last month. To achieve this, we firstly got the monthly information for all A-share stocks, ranked each stock according to its monthly market capitalization and monthly return respectively. At the end, we long 20 stocks and short 20 stocks according to the strategy and calculate the average return.

Chapter 3

Multiple Linear Regression

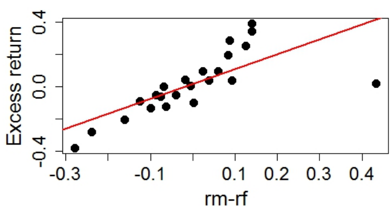
3.1 Simple Linear Regression using CAPM

```
> capm.model <- lm(Excess_ret ~ Market_excess_ret,data=data)
> summary(capm.model)

Call:
lm(formula = Excess_ret ~ Market_excess_ret, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-0.39545 -0.06722 -0.00359  0.05168  0.24573

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.01643    0.02810   0.584 0.565132
Market_excess_ret   0.92228    0.19404   4.753 0.000108 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1348 on 21 degrees of freedom
Multiple R-squared:  0.5183,Adjusted R-squared:  0.4953
F-statistic: 22.59 on 1 and 21 DF,  p-value: 0.0001076
```

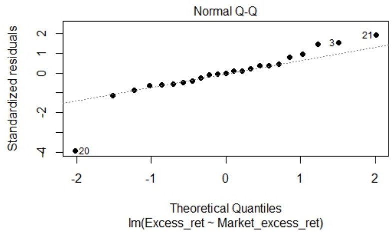
The simple linear regression gives the following result:

$$R_{it} - R_{Ft} = 0.01643 + 0.92228(R_{Mt} - R_{Ft})$$

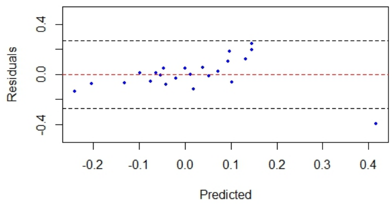
In finance, abnormal return is defined as the difference between actual return and expected return. In our model, the intercept can be interpreted as the abnormal return, while the slope can be interpreted as the expected reutnr. We can see from the regression result that the intercept is not significant but the slope is significant, which means the porfolio doesn't offer significant abnormal return (the intercept) in addition to normal return (the slope).
We noticed the model has a multiple R^2 of 0.5183, which means approximate 50% of the excessive return for the portfolio is explained by CAPM model. However, the QQ plot for residuals shows slightly thin tails, and the residual plot and regression plot both show there is an outlier. By taking a closer look at this outlier, we noticed it is associated with an extremely high market return and a relatively low portfolio return in December 2014. Although it is mentioned in class that usually a preferred way to handle this situation would be via transformation, in this case, we feel dropping the outlier directly makes sense due to the unique nature of Chinese stock market. Predictably, there is a



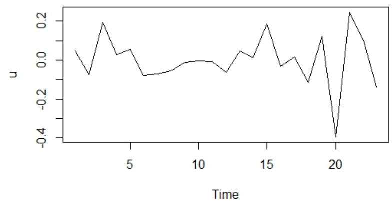
(A) Simple Linear Regression



(B) Residue QQ Plot



(C) Residue Plot



(D) Residue TS Plot

FIGURE 3.1: CAPM

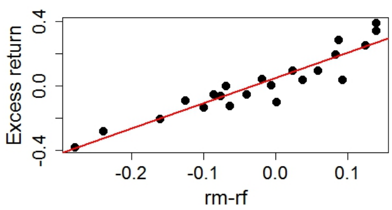
non-trivial improvement of explanation power after we dropped the outlier from the dataset.
Another potential cause for the outlier could be due to the data purging process. During the data cleaning process, we dropped all the stocks with missing data, while the majority of them turns to be highly volatile. As a consequence the remaining stocks used for the regression are much more stable and thusly the portfolio performance does not reflect the extreme changes of the market.

```
> modified=data[data$Date!='2014-12',]
> capm.modified <- lm(Excess_ret~Market_excess_ret, data=modified)
> summary(capm.modified)

Call:
lm(formula = Excess_ret ~ Market_excess_ret, data = modified)

Residuals:
     Min        1Q    Median        3Q       Max
-0.155022 -0.035799  0.009943  0.042098  0.125355

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.04811    0.01560   3.084  0.00586 **
Market_excess_ret  1.55777    0.13499  11.540  2.7e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07189 on 20 degrees of freedom
Multiple R-squared:  0.8694,Adjusted R-squared:  0.8629
F-statistic: 133.2 on 1 and 20 DF,  p-value: 2.703e-10
```

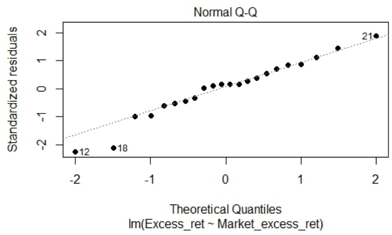
The simple linear regression gives the following result:

$$R_{it} - R_{Ft} = 0.04811 + 1.55777(R_{Mt} - R_{Ft})$$

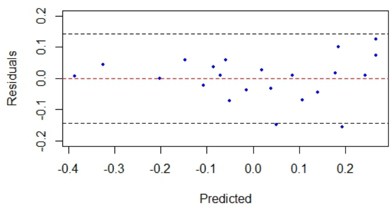
In this modified regression, both slope and intercept are significant. In financial scope, we say this portfolio offers abnormal return (the intercept) and normal return (the slope).



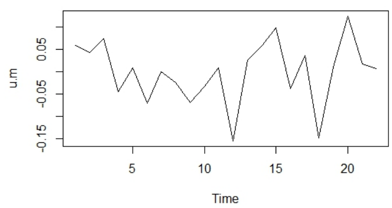
(A) Linear Regression



(B) Residue QQ Plot



(C) Residue Plot



(D) Residue TS Plot

FIGURE 3.2: CAPM Modified

From QQ-plot, although it has thin tails, normality assumption still holds. Also, we observe the residuals appear to be evenly distributed, therefore constant variance assumption is not violated. No obvious correlation can be identified from the time series plot, which means a fair assumption of independence.

3.2 Multiple Linear Regression using Fama-French Model

3.2.1 Fama-French Three Factors Model

```
> three.fac.model <- lm(Excess_ret~Market_excess_ret+SMB+HML, data=data)
> summary(three.fac.model)

Call:
lm(formula = Excess_ret ~ Market_excess_ret + SMB + HML, data = data)

Residuals:
     Min        1Q    Median       3Q      Max
-0.135798 -0.048239  0.000987  0.041602  0.117026

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.002259   0.026241   0.086 0.932286
Market_excess_ret  1.200102   0.107668  11.146 8.92e-10 ***
```

```
SMB                 -0.578047   0.136366  -4.239 0.000444 ***
HML                 -0.255188   0.138859  -1.838 0.081789 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.0692 on 19 degrees of freedom
Multiple R-squared:  0.8851,Adjusted R-squared:  0.8669
F-statistic: 48.78 on 3 and 19 DF,  p-value: 4.053e-09
```

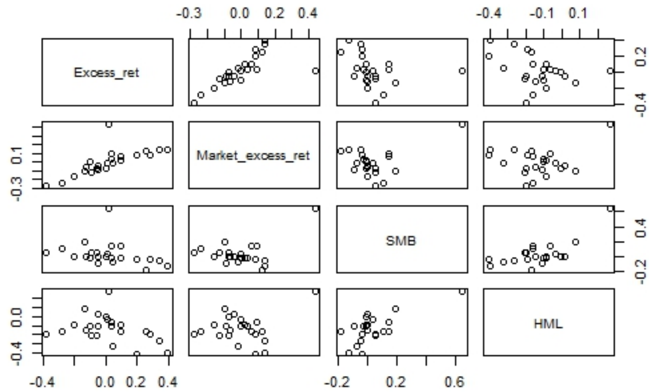


FIGURE 3.3: Fama-French Three Factors Scatter Plot

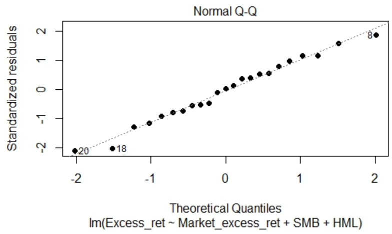
Multiple linear regression using Fama-French three-factor model gives the following coeffients:

$$R_{it} - R_{Ft} = 0.002259 + 1.200102(R_{Mt} - R_{Ft}) - 0.578047 \cdot SMB - 0.255188 \cdot HML$$

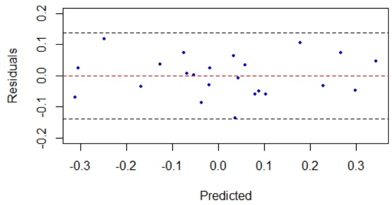
The estimate for market excess return and SMB is significant. It is interesting to find that the sign or SMB estimate is negative, which means that if the return difference between small companies and big companies increases 1 unit, the average return of this portfolio will decrease 0.58 units.
According to Adjusted R-squared, the model explains over 86.69% of the diversified portfolio returns. It is a quite an improvement from the simple linear regression using CAPM model including outlier. We can conclude that FF3 model is better than CAPM since we needn't do special treatment on outliers in FF3 model.

3.2.2 Fama-French Five Factors Model

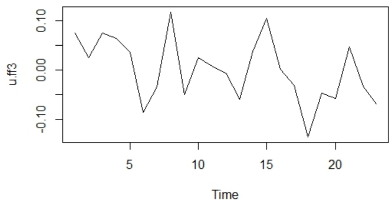
```
> five.fac.model <- lm(Excess_ret~Market_excess_ret+SMB+HML+RMW+CMA,
data=data)
> summary(five.fac.model)

Call:
lm(formula = Excess_ret ~ Market_excess_ret + SMB + HML + RMW + CMA,
data = five.factor)

Residuals:
     Min       1Q   Median       3Q      Max
```



(A) Residue QQ Plot



(B) Residue Plot



(C) Residue Time Series Plot

FIGURE 3.4: Fama-French Three Factors Model

```
-0.13504 -0.05054  0.01152  0.05416  0.10532

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)
(Intercept)        0.007215   0.029195   0.247  0.80777
Market_excess_ret  1.189206   0.117128  10.153 1.24e-08 ***
SMB               -0.540389   0.164080  -3.293  0.00429 **
HML               -0.235648   0.149130  -1.580  0.13250
RMW               -0.091264   0.202102  -0.452  0.65729
CMA                0.200104   0.407655   0.491  0.62980
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07238 on 17 degrees of freedom
Multiple R-squared:  0.8875,Adjusted R-squared:  0.8544
F-statistic: 26.82 on 5 and 17 DF,  p-value: 1.679e-07
```

As an extension of Fama-French three-factor model, the Fama-French five-factor model gives an Adjusted R^2 value of 0.8544, which is slightly less than the three-factor model:

$$\begin{aligned} R_{it} - R_{Ft} = 0.007215 + 1.189206(R_{Mt} - R_{Ft}) - 0.540389 \cdot SMB \\ a - 0.235648 \cdot HML - 0.091264 \cdot RMW - 0.200104 \cdot CMA \end{aligned}$$

Only the estimate for market excess return and SMB is significant. In terms of the Adjusted R^2, FF5 model has weaker explanation power than the FF3 model due to the possibility of overfitting. We further checked all the assumptions.

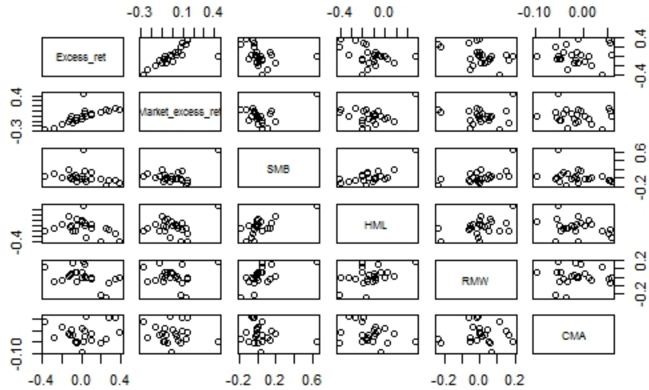
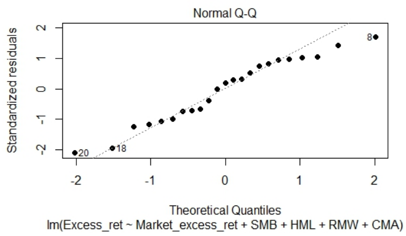
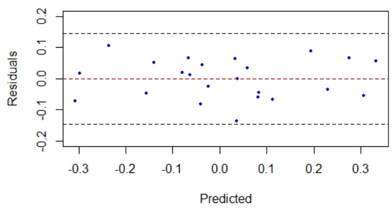


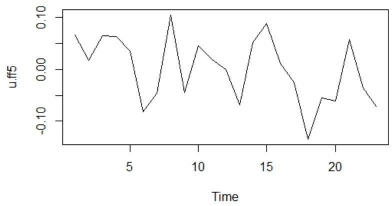
FIGURE 3.5: Fama-French Five Factors Scatter Plot



(A) Residue QQ Plot



(B) Residue Plot



(C) Residue Time Series Plot

FIGURE 3.6: Fama-French Five Factors Model

3.3 Finding the best model using backward elimination

Starting from the full model, we want to find a best model using these five factors by performing a sequential variable selection.

```
> full.model <- lm(Excess_ret~(Market_excess_ret+SMB+HML+RMW+CMA)^2
, data=data)
> summary(full.model)

Call:
lm(formula = Excess_ret ~ (Market_excess_ret + SMB + HML + RMW + CMA)^2
, data = data)

Residuals:
     Min         1Q    Median         3Q        Max
-0.094824 -0.016283  0.001876  0.020867  0.070361

Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           -0.01504    0.04362  -0.345    0.740
Market_excess_ret      0.83279    0.55532   1.500    0.177
SMB                    0.04310    0.64468   0.067    0.949
HML                   -0.29932    0.29437  -1.017    0.343
RMW                   -0.58815    0.57121  -1.030    0.337
CMA                    0.24344    0.86620   0.281    0.787
Market_excess_ret:SMB  1.18314    2.07228   0.571    0.586
Market_excess_ret:HML -2.82140    2.65095  -1.064    0.323
Market_excess_ret:RMW -2.53773    2.06168  -1.231    0.258
Market_excess_ret:CMA  4.88625    5.22205   0.936    0.381
SMB:HML                0.36933    1.41867   0.260    0.802
SMB:RMW                0.74916    2.93797   0.255    0.806
SMB:CMA                0.19525   18.61522   0.010    0.992
HML:RMW               -1.30417    1.99386  -0.654    0.534
HML:CMA                0.42770    5.02710   0.085    0.935
RMW:CMA               -9.11569   13.51054  -0.675    0.522

Residual standard error: 0.06471 on 7 degrees of freedom
Multiple R-squared:  0.963,Adjusted R-squared:  0.8837
F-statistic: 12.14 on 15 and 7 DF,  p-value: 0.001347

> model.back <- step(full.model,direction='backward')
> summary(model.back)
Call:
lm(formula = Excess_ret ~ Market_excess_ret + HML + RMW + CMA +
Market_excess_ret:HML + Market_excess_ret:RMW + RMW:CMA,
data = data)

Residuals:
     Min         1Q    Median         3Q        Max
-0.102020 -0.018525  0.000531  0.024135  0.070784
```

```
Coefficients:
                       Estimate Std. Error t value Pr(>|t|)
(Intercept)           -0.006586   0.019492  -0.338  0.74014
Market_excess_ret      1.016719   0.100195  10.147 4.12e-08 ***
HML                   -0.285203   0.107204  -2.660  0.01782 *
RMW                   -0.455672   0.179472  -2.539  0.02269 *
CMA                    0.253402   0.275467   0.920  0.37218
Market_excess_ret:HML  -1.936728   0.505799  -3.829  0.00164 **
Market_excess_ret:RMW  -1.719259   0.992338  -1.733  0.10368
RMW:CMA              -12.834854   4.396705  -2.919  0.01057 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04904 on 15 degrees of freedom
Multiple R-squared:  0.9544,Adjusted R-squared:  0.9332
F-statistic: 44.89 on 7 and 15 DF,  p-value: 6.112e-09
```

The best regression model we find by using backward elimination process provides an impressive adjusted R^2 value of 93.32%. However, this result is relatively artificial as it's hard to exlain the amazing explanation power using the theoretical foundation of Fama-French model. Nevertheless, FF3 model has excellent empirical performance as a financial theoretical model.

Chapter 4

Average Return of Stock Selecting Strategies

4.1 Average Return of Portfolios Small Cap v.s. Big Cap

In this section, we firstly test the difference of this month's return between portfolios constrcuted with 20 stocks with the largets market capitalization in last month and 20 stocks with the smallest capitalization in last month.

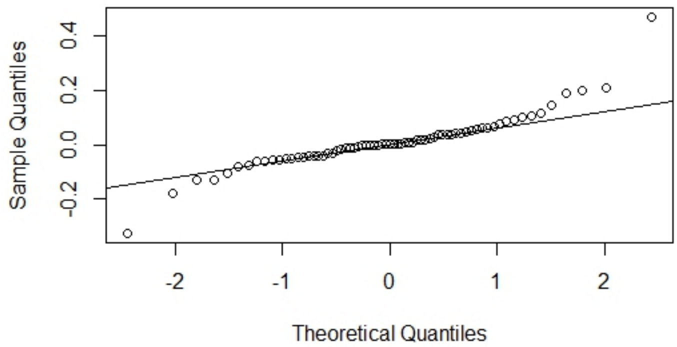


FIGURE 4.1: Small minus Big Stocks return QQ plot

The QQ plot shows that the difference relatively follows a normal distribution.

```
> t.test(strategy$small,strategy$big,paired = TRUE)

Paired t-test

data:  strategy$small and strategy$big
t = 0.91095, df = 68, p-value = 0.3655
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.01300289  0.03484658
sample estimates:
mean of the differences
             0.01092184
```

Since the p-value is bigger than 0.05, we can't reject the null hypothesis that there is no difference of this month's return between portfolios constrcuted with 20 stocks with the

largest market capitalization in last month and 20 stocks with the smallest capitalization in last month.
Then we test whether longing small stocks and shorting big stocks will perform better than the market.

```
> t.test(strategy$small-strategy$big,strategy$market,paired=TRUE)

Paired t-test

data:  strategy$small - strategy$big and strategy$market
t = 0.43413, df = 68, p-value = 0.6656
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.02403652  0.03740336
sample estimates:
mean of the differences
0.006683416
```

Since the p-value is bigger than 0.05, we can't reject the null hypothesis. Our strategy of longing small cap stocks and shorting big cap stocks doesn's seem feasible.

4.2 Average Return of Portfolios High v.s. Low

In this section, we will test the difference of this month's return between portfolios constructed with 20 stocks with the highest monthly return in last month and 20 stocks with lowest monthly return in last month.

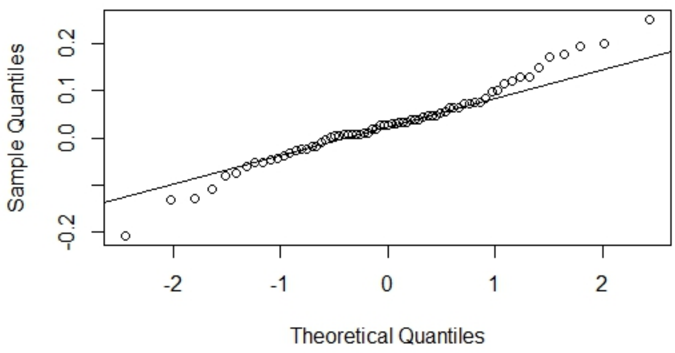


FIGURE 4.2: Low minus High Return Stocks QQ plot

The QQ plot shows that the difference relatively follows a normal distribution.

```
> t.test(strategy$low,strategy$high,paired = TRUE)

Paired t-test

data:  strategy$low and strategy$high
t = 3.0185, df = 68, p-value = 0.003574
alternative hypothesis: true difference in means is not equal to 0
```

```
95 percent confidence interval:
0.009852539 0.048287515
sample estimates:
mean of the differences
0.02907003
```

Then we test whether longing low return stocks and shorting high return stocks will perform better than the market. Since the p-value is smaller than 0.05, we reject the null hypothesis and conclude that there is a significant difference of this month's return between portfolios constrcuted by 20 stocks with highest return in last month and 20 stocks with lowest return in last month. It is interesting to see that the stocks which performed badly last month tend to perform better next month.

```
> t.test(strategy$low-strategy$high,strategy$market,paired=TRUE)

Paired t-test

data:  strategy$low - strategy$high and strategy$market
t = 1.7181, df = 68, p-value = 0.09032
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-0.004008304  0.053671501
sample estimates:
mean of the differences
0.0248316
```

Since the p-value is bigger than 0.05, we can't reject the null hypothesis that there's no differnce between the average return of low monthly stocks and high monthly return stocks. Our strategy of longing low return stocks and shorting high return stocks hence doesn't seem feasible.

Chapter 5

Conclusion

In the first part of the project, we used linear regression to test CAPM, Fama-French Three-Factor Model and Fama-French Five-Factor Model on Chinese capital market from 2010 to 2015. We noticed CAPM is quite sensitive to outliers, which makes it less appropriate in Chinese market, where outliers are quite common due to the high volatility. Instead of using a transformation to handle the outlier as mentioned in class, we dropped the outlier directlyl and observed an improvement of CAPM results. In terms of the adjusted R-squared value of the three models, FF3 Model shows the best performance. At last, we used sequential variable selection to find the model with minimum AIC. Although the model provides an impressive explanation power of 95%, it's more of an artificial outcome than something we could explain using the theoretical foundation of Fama-French models.

In the second part, we used t-tests to detect the differnce of average return of customized portoflios. First portfolio is created based on one of the underlying assumptions of FF3 that the average return of large cap stocks differs from small cap stocks. The second portofolio is inspired by the observation that momentum appears to be a dominant factor in Chinese market. For the first portfolio, we tested whether there is a difference between the average return by longing the stocks with the larger capital and by longing stocks with the lower capital. Although the average return of small cap stocks is slightly higher than big cap stocks, the result is not significant. For the second portfolio, we tested the difference between the average return by longing the stocks with the highest return in last month and by longing stocks with the lowest return in last month. The result showed the average return of low return stocks is higher than high return stocks, with a significant difference.

Appendix A

Data and Code files

1. “5_factor_plot_example.csv”: Data file which contains all data needed for examining Fama-French 5 factors model.
2. “strategy.csv”: Data file contains monthly return of portfolios constrcuted by various strategy.
3. “final_data_cleaning.py”: Python program used to perform initial data purging on raw data files.
4. “select_and_export.py”: Python program used to either manually or automatically create the portfolio and generate “5_factor_plot_example.csv”
5. “t_test_data.py”: Python program used to create “strategy.csv” data file.
6. “regression_test.R”: R program used to run various t-test for the data.
7. “stock_index.R”: R program used to download monthly and quarterly history return.

Bibliography

Fama E. F.; French, K. R. (1993). “Common risk factors in the returns on stocks and bonds”. In: *Journal of Financial Economics* 33.3.

Lintner, John (1965). “The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets.” In: *Review of Economics and Statistics* 47.1, pp. 13–37.

Sharpe, William F. (1964). “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk”. In: *Journal of Finance* 19.3, pp. 426–442.