

The Application of BARRA Model with different weighted methods in Chinese Stock Market

¹Guang Bian, ²Xingye Li

¹ Business School, University of Shanghai for Science and Technology, Shanghai, China

^{*2} Business School, University of Shanghai for Science and Technology, Shanghai, China

Email: bgalan@126.com

Abstract –Since the factors can represent the components of returns, as seen by the financial analyst, the multiple-factor model is a natural representation of the real environment. Barra multiple-factor model has been widely used in this mature capital market. Many investors use this model to get the estimates of the market risk and pursuit excess returns. As an active and effective model. Barra multiple-factor model can provide us robust and precise risk forecast in stock market. In this paper, Barra multiple-factor model is constructed with some effective factors in Chinese stock market. Then, we analysis the effectiveness of this model by using different weighted methods. After descriptor selection and testing, we choose nine significant factors to establish the Barra multiple-factors model. Through the research of different weighted methods of estimating factor returns, exponentially declining weighted method is better than average weighted method.

Keywords –Barra Model; risk ;factors ; active portfolio; preparation;

1. Introduction

As Chinese economic has played an important role in this world, the investment opportunities have been blooming through this world. In recent years the investment management industry has adjusted to continuing changes-theoretical advances, technological developments, and volatility. To address these changes, investment managers and financial institutions require the most advanced and powerful analytical tools available. Investors construct effective investment models to control security risk and pursuit excess returns, and expect to be able to obtain excess returns in this stock market. Establishing risk management models with financial theory and then applied in the securities portfolio management of assets is one of the problems in financial markets.

Markoeitz the beginning of the modern financial theory in investment field, had proposed a method of constructing an effective portfolio in 1952 .This theory tell us how to maximize returns in a given risk and how to minimize risk in a given returns.[1]But Markowitz portfolio models has some defects, the variance-covariance matrix which forms from the solution of that optimized problem. The Markowitz Efficient Frontier states that an investor can make decision along efficient frontier curve and exchange the given risk for expected return. From this theory since investors can reconstruct the portfolio to achieve a greater expected return for the same risk, this optimization motives the investors for diversification.

After the Markowitz model, financial theory of portfolio management is still developing. Sharpe (1964), Lintner (1965) and Mossin (1966) show how to use capital asset pricing model (CAPM) to determine a theoretically appropriate required rate of return of an asset, which is viewed as significant stage in modern investment theory. Under the CAPM, the expected residual return on any stock or portfolio is zero. Expected excess returns are proportional to the stock's (or portfolio's) beta. Sharp given the single index exponential model and multi-index exponential model based on this idea. In a 1969 conference in Buffalo, New York of a paper by Fischer Black, Michael Jensen, and Myron Scholes, it found that higher returns will be reward to low beta stocks, although the theory tells us that low beta should deserve the low return. This result offers wise investors to arbitrage from volatility strategy to beat the market.

The paper of Ross (1976) gives us a new way, APT, to do asset pricing. The theory argues that the linear function of various factors or theoretical market factors can be modeled to represent the expected return of the financial asset. If the price fluctuates from the theoretical expected price, arbitrage action should workout. Therefore the asset price should equal the expected end of period price implied by the model. The APT and CAPM theory are both important theories on asset pricing. These models are estimated by linear regression method. Fama and French (1992) studied three factors which can explain the returns of stock market, then constructed Fama-French three-factor model. This model is based on the

results of empirical research in the history of the rate of returns of the U.S. stock market, which wants to find the factors that could explain the average rate of returns of the stock market. Barra (1976) follows the APT methods, which uses multifactor method to build model. It takes some initial steps on this field. The fundamental factors related with risks are important with expected returns in BARRA model.

Barra Multiple-factor model identify common factors, which are categories defined by common characteristics of different securities, and determine the return sensitivity to these factors. Multiple-factor models of security market returns can be divided into three types: macroeconomic, fundamental, and statistical factor models. Macroeconomic factor models use observable economic variables, such as changes in inflation and interest rates, as measures of the pervasive shocks to security returns. Fundamental factor models use the returns to portfolios associated with observed security attributes such as dividend yield, book-to-market ratio, and industry membership. Statistical factor models derive their factors from factor analysis of the covariance matrix of security returns. Barra equity models are fundamental factor models, which outperform the macroeconomic and statistical models in terms of explanatory power.

2. Barra Multiple-factor Model

In Barra Rosenberg and Vinay Marathe (1976), the paper comes up development from APT theory with a new approach to construct portfolio. Generally speaking, BARRA model is one of multifactor models. The expected return is composed of returns due to different risks.

The general multifactor model is as follow:

$$R_i = \alpha_i + x_{i1}F_1 + \dots + x_{ik}F_k + \varepsilon_i \quad i=1,2,\dots,n \quad (1)$$

In equation (1), R_i is the excess return to each asset. α_i is the constant term of the regression equation, which is the common factor of risk factor. x_{ik} is risk exposure of security i to factor k . F_1, F_2, \dots, F_k is factor returns. ε_i is specific returns of security i .

Common factors of return are the influential factors for numerous stocks, such as liquidity, financial leverage and variability etc. Exposure coefficients on the factors are different within individual company, which is a measure of how risky the company is exposed to these factors. Specific return specifically influences the company alone, which is theoretically orthogonal to common factor return. It also implies that the specific returns should be independent among companies.

BARRA model will normalize the descriptors to form risk indices which describe the firm characteristics, regress the risk indices with excess return to get factor returns matrix and forecast the expected return. Combined with regression method, we can precisely capture the risks that company owns and do further risk analysis and portfolio construction.

$R_i, F_1, F_2, \dots, F_k$, and ε_i are random variables. The variance of ε_i is denoted by σ_i^2 . The covariance matrix of F_1, F_2, \dots, F_k is denoted by Φ . In addition, the validity

of this model depends on the following assumptions. Firstly, specific returns $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ is independent of each other. This means that the correlation between the two different securities rate of returns is only determined by the common factor F_1, F_2, \dots, F_k . Therefore, this assumption makes the estimation of the covariance matrix Σ which is formed by securities rate of returns simple. Secondly, $E(\varepsilon_i)=0$, the part which could not be explained by common factors F_1, F_2, \dots, F_k are involved in α_i . Thirdly, special returns and common factors are independent of each other.

The equation can be rewritten into a matrix form. $R=(R_1, R_2, \dots, R_n)$ is a random vector. Factor exposure matrix is denoted by \tilde{X} .

$$\tilde{X} = \begin{bmatrix} \alpha_1 & x_{11} & \dots & x_{1k} \\ \alpha_2 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \dots & \vdots \\ \alpha_n & x_{n1} & \dots & x_{nk} \end{bmatrix} \quad (2)$$

\tilde{F} denotes the k-dimensional vector $(1, F_1, F_2, \dots, F_k)$. Special returns vector is denoted by ε . Because the special returns is independent of each other, the covariance matrix of ε , which is denoted by Δ , is $diag(\sigma_{\varepsilon_1}^2, \sigma_{\varepsilon_2}^2, \dots, \sigma_{\varepsilon_n}^2)$. So the upper equation can be rewritten in matrix form,

$$R = \tilde{X}\tilde{F} + \varepsilon \quad (3)$$

that is,

$$\begin{pmatrix} R_1 \\ R_1 \\ \vdots \\ R_1 \end{pmatrix} = \begin{bmatrix} \alpha_1 & x_{11} & \dots & x_{1k} \\ \alpha_2 & x_{21} & \dots & x_{2k} \\ \vdots & \vdots & \dots & \vdots \\ \alpha_n & x_{n1} & \dots & x_{nk} \end{bmatrix} \begin{pmatrix} F_1 \\ F_2 \\ \vdots \\ F_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_k \end{pmatrix} \quad (4)$$

Then the covariance of R can be written as,

$$\Sigma = Cov(\tilde{X}\tilde{F} + \varepsilon) = XCov(F)X^T + Cov(\varepsilon) = X\Phi X^T + \Delta$$

Among the formula, $F=(F_1, F_2, \dots, F_k)$. Matrix X is a new matrix which is formed by matrix \tilde{X} without its first column. So, we can get $Cov(\tilde{X}\tilde{F} + \varepsilon) = XCov(F)X^T$. Therefore the covariance matrix of the multiple-factor model is the formula above. Covariance matrix used to construct the weights of the portfolio between each asset and factor exposure and to calculate the risk of the portfolio.

Assumed that there is a portfolio P , which is composed with N assets. Its weighted vector is denoted by h_p . So the vector of factor exposure x_p can be calculated as $x_p = X^T h_p$. The covariance matrix is used in conjunction with a portfolio's weight in each asset and the factor exposures of those assets to calculate portfolio risk. The following formula is the underlying form of Barra risk calculations:

$$\sigma_p = \sqrt{h_p^T (X\Phi X^T + \Delta) h_p} \quad (5)$$

Where σ_p is the volatility of portfolio returns, and h_p is the vector of portfolio weights for N assets.

After the section-cross regress, we can get the value of factor returns. In order to get the estimations of the rate of returns in practice, we need to estimate the value of factor returns. There are two methods. The one is average weighted; the other is exponential declining weighted method.

Average weighted is calculating the average of the factor returns of T periods, which are obtained by section-cross regression. Denote T by the current period, and by t any period in the past $t=1,2,\dots,T-1,T$. Then the weight of the factor returns in any periods are $\frac{1}{T}$.

In the method of exponential declining weighted, suppose that we think that observations that occurred 20 months ago should receive half the weight of the current observation. Let $\lambda=0.5^{\frac{1}{10}}$. If we assign a weight of λ^{T-t} to observation t , then an observation that occurred 20 months ago would get half the weight of the current observation, and one that occurred 40 months ago would get one-quarter the weight of the current observations. More generally, let $\lambda=(0.5)^{\frac{1}{HALF-LIFE}}$, and assign the weight $w(t)=\lambda^{T-t}$. Therefore, we can calculate the weight of every past period, then the estimation of the current factor returns is easy.

3. Empirical Analysis in Chinese stock market

Based on the characteristics of the Chinese securities market, this paper's data selects the CSI300 Index constituent stocks, and the time is from January 2009 to August 2012, a total of 44 months. There are 244 stocks with the removal of the listed and suspension of the stock during the period. We take the logarithmic rate as the stock rate of returns:

$$r_{it} = \ln\left(\frac{p_{it}}{p_{t-1,i}}\right) \tag{6}$$

Where r_{it} is the rate of returns of time t , and p_{it} is the closing price of stock i time t .

3.1 Descriptor Selection and Testing

Descriptors can be chosen from several sources. For some descriptors, market and fundamental information is combined. In this paper, we choose price to book value, price-earning/ratio, rate of return on total assets and some other descriptors. As showed in Table 1.

Table 1. Descriptors Selection and classification

Index	Descriptor
Profitability	Net Profit Margin on Sales
	Rate of Return on Common Stockholders' Equity(ROE)
	Return On Assets(ROA)
Growth	Total Assets Growth Rate
	Return On Net Assets
	EPS growth rate
	Increase rate of main business revenue
Leverage	Circulation market value/Total market value
	Debt asset ratio
Liquidity	Average trading volume of three-month
	Turnover Rate
Momentum	Three-month price reversal
	Six-month price reversal
Quality	Operating activities Net income / total profit
	Net cash flow generated from operating activities/ total profit
	Equity ratio
	Total Assets Turnover
Estimation	Price to book ratio
	Price to Earning Ratio
	Market value operating income ratio
Size	Circulation market value

In order to select the effective factors in the cross-section, we should remove the descriptive variables which are redundant. After the correlation analysis of

the factors above, that is, calculate the correlation between the stock rate of return in period t and descriptive variables in period $t-1$, then we obtain

correlation of each period. Next step is calculating their average value, and select 11 descriptors with strong correlation.

Now we should consider the multicollinearity between indicators that we get. Then we calculate the

correlations between each descriptor variables and average them, so that we get the correlation analysis table of the remaining 11 descriptive variables. As showed in table 2.

Table 2. Correlation Analysis after the significant test

Correlation Analysis	Net Profit Margin on Sales	ROE	ROA	EPS growth rate	Circulation market value/Total market value	Turnover Rate	Three-month price reversal	Six-month price reversal	Price to Earning Ratio	Price to book ratio	Circulation market value
Net Profit Margin on Sales	1.0000	0.3721	0.3974	0.0561	0.0402	0.0188	0.0190	0.0201	-0.0091	-0.0433	-0.0654
ROE	0.3721	1.0000	0.7680	0.2262	0.0447	0.0123	0.0349	0.0350	0.0008	-0.0058	-0.0803
ROA	0.3974	0.7680	1.0000	0.1451	0.0486	0.0461	0.0183	0.0186	0.0128	-0.0048	-0.0713
EPS growth rate	0.0561	0.2262	0.1451	1.0000	-0.0050	-0.0011	0.0222	0.0235	-0.0269	-0.0120	-0.0138
Circulation market value/Total market value	0.0402	0.0447	0.0486	-0.0050	1.0000	-0.1809	-0.1531	-0.1494	-0.0537	-0.2044	-0.1133
Turnover Rate	0.0188	0.0123	0.0461	-0.0011	-0.1809	1.0000	0.2241	0.2266	0.1232	0.2181	-0.2241
Three-month price reversal	0.0190	0.0349	0.0183	0.0222	-0.1531	0.2241	1.0000	0.9938	0.0494	0.1640	-0.0147
Six-month price reversal	0.0201	0.0350	0.0186	0.0235	-0.1494	0.2266	0.9938	1.0000	0.0502	0.1644	-0.0160
Price to Earning Ratio	-0.0091	0.0008	0.0128	-0.0269	-0.0537	0.1232	0.0494	0.0502	1.0000	0.2121	-0.0306
Price to book ratio	-0.0433	-0.0058	-0.0048	-0.0120	-0.2044	0.2181	0.1640	0.1644	0.2121	1.0000	-0.0772
Circulation market value	-0.0654	-0.0803	-0.0713	-0.0138	-0.1133	-0.2241	-0.0147	-0.0160	-0.0306	-0.0772	1.0000

Attention that the correlation coefficient between ROA and ROE is 0.77, and the correlation coefficient between three-month price reversal and six-month price reversal is 0.99. Hence, there is collinearity in these variables. Considerate the correlation coefficient obtained previously, we exclude ROA and six-month price reversal, so we get 9 significant descriptive variables at last.

3.2 extreme processing

Extremes can distort the analysis results seriously, so the processing of removing extremes is necessary. This paper takes the Skipped Huber Method. Mathematical formulas are described as follows:

$$D_{i,upper} = D_m + 5.2D_{MAD}, \text{ if } D_i \geq D_m + 5.2D_{MAD} \quad (8)$$

$$D_{i,lower} = D_m - 5.2D_{MAD}, \text{ if } D_i \leq D_m - 5.2D_{MAD}$$

Where D_i is denoted by the i -th observed value of the descriptors. D_m is denoted by the median of each descriptor. $D_{i,AD}$ is denoted by absolute deviation

between each observed value and the median, $D_{i,AD} = |D_i - D_m|$. D_{MAD} is the median of all the absolute deviation $D_{i,AD}$. $D_{i,upper}$ and $D_{i,lower}$ are denoted as upper and lower limits of the descriptive variables after the Skipped Huber Method.

3.3 Descriptor Standardization

Each descriptor can explain the rate of returns in the next period. Risk indices are composed of descriptors designed to capture all the relevant risk characteristics of a company. The descriptors are first normalized, that is, they are standardized with respect to the estimation universe. The normalization process involves setting random variables to a uniform scale. A constant (usually the mean) is subtracted from each number to shift all numbers uniformly. Then each number is divided by another constant (usually the standard deviation) to shift the variance.

The normalization process is summarized by the following relation:

$$normalized = \frac{[rawdescriptor] - [mean]}{[standard deviation]} \quad (7)$$

Then, the descriptors are combined into meaningful risk factors, known as risk indices.

3.4 Factor Return Calculation

The previous steps have defined the exposures of each asset to the factors at the beginning of every period in the estimation window. The factor returns over the period are then obtained via a cross-sectional regression of asset returns on their associated factor returns. But the variance of excess returns ε_{it} of every asset is not the same, and then the problem of heteroscedasticity is raised. In order to solve this problem, the method of estimating factor returns usually is generalized least squares. According to the studies of Grinold and Kahn (1995) and Barra, it's appropriate to replace the inverse of the excess returns matrix with a inverse matrix of diagonal matrix whose diagonal elements is the square root of the market value. Then, generalized least squares regression can be transformed into a weighted least squares regression. Factor returns can be calculated in formula (9).

$$F = (X^T w^{-1} X)^{-1} X^T w^{-1} R \quad (9)$$

Where R is excess return vector, w is the diagonal matrix which is composed with each stock's market value.

$$w = \begin{pmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{pmatrix} \quad (10)$$

3.5 Factor Return Estimation and Evaluation

In this paper, factor return estimation is using both average weighted method and exponential declining

method. We choose winning percentage (rate of stock returns bigger than CSI300 Index's), annualized rate of return and Information Ratio as the evaluation of the method.

An information ratio, denoted by IR, is a ratio of (annualized) residual return to (annualized) residual risk. Because the result we get before is monthly excess return, and denoted the standard deviation of monthly rate of return as residual risk. The monthly information ratio can be calculated. The relationship between annualized information ratio and monthly information ratio is the followed formula.

$$IR_y = \sqrt{12} IR_m \quad (11)$$

Where IR_y is annualized information ratio, and IR_m is monthly information ratio.

The larger the information ratio is, the higher the portfolio's excess return is.

3.6 Empirical Results

Based on the Barra multiple-factor model, we can calculate the value of the factor returns by cross-section of the weighted least squares regression. Then this paper estimate the factor returns using the average weighted method and exponential declining method. In order to estimate the rate of returns of the next period, factor returns multiply the factor exposure in last period. So we can get the estimation of the return ratio of the portfolio. Using the same method, we can select the top stocks whose rate of returns in the estimation is the top 20, then we can calculate these stocks' real rate of returns with average weighted. In order to compare the results of the two methods, we select the periods $T=2,4,8,16,32$. Detailed results are shown in Table 3. Figure 1 shows the Barra multiple-factor model results when $T=32$.

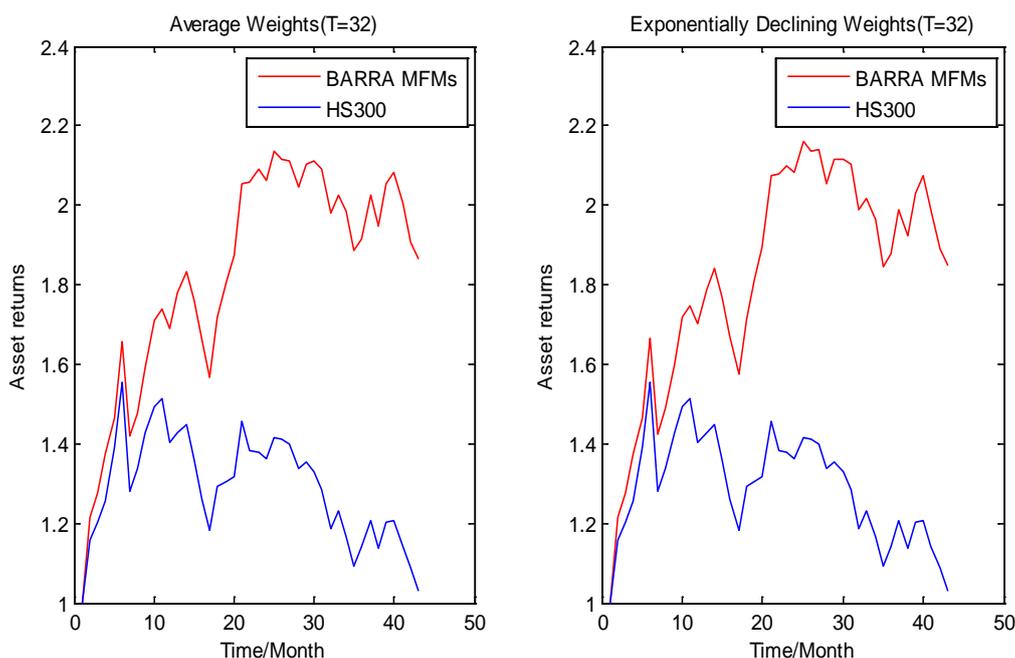


Figure 1. The comparison between Barra multiple-factor model and the rate of CSI 300 Index as T=32

Table 3. Comparison between different weighted method of BARRA Multiple-Factor Model

Weighting method	Comparing Index	Time Selecting				
		2	4	8	16	32
Average weighted	Winning Percentage	0.6190	0.5952	0.6429	0.6905	0.7381
	Annualized rate of return	0.1215	0.1533	0.1417	0.1589	0.1849
	Information Ratio	0.6618	0.8574	1.0548	1.1715	1.5964
Exponentially Declining Weighted	Winning Percentage	0.5476	0.5476	0.6667	0.6905	0.7619
	Annualized rate of return	0.1081	0.1823	0.1751	0.1681	0.1905
	Information Ratio	0.6833	1.0068	1.2784	1.1804	1.5899

As is shown in Table 3, with the selecting periods gradually an increase, the predictions of the model has a better improvement. But the exponentially declining weighted method has a better result than average weighted method. When the weighted time is two, the smallest time, the percentage of the rate of return of average weighted method outperformed CSI300 Index's is 61.90%, as the exponentially declining weighted method is 54.76%, which is much smaller. When the weighted time is thirty-two, the largest time, the percentage of the rate of return of average weighted method outperformed CSI300 Index's is 73.81%, as the exponentially declining weighted method is 76.19%, that is a little larger. In the aspect of annualized rate of return, exponentially declining weighted method is larger than average weighted method in any selecting times. Information ratio increases as the selecting times increasing in both methods. But the second method is a little better than the first method. As the time increases, each index becomes better. Generally speaking, exponentially declining weighted method is better than average weighted method.

4. Conclusions

This paper studies the different weighting method of the Barra Multiple-factor Model in predicting of Chinese stock market. We select the winning percentage, annualized rate of return and information ratio indexes as the evaluations to analysis the results. With different weighted method, the estimation of factor returns is different, so as the portfolio selection and the rate of return of stocks. Overall, with the application of Baara Multiple-Factor Model, the results from exponentially declining weighted method are better than average weighted method. Because exponentially declining weighted method gives the recent period a large weight, which means that the estimation of present factor return can be explained much more by recent factor returns. This conclusion can be a good reference in the empirical applications.

Chinese equity market is only 21 years since its inception, but the progress is enormous. In order to find a better way to estimate Chinese stock market, Barra Multiple-factor Model needs to constantly update and

improve. This study has a practical sense of constructing the portfolio with small risk and large rate of returns.

5. Acknowledgements

This research is supported by The Innovation Fund Project for Graduate Student of Shanghai (Project Number: JWCXSL1102).

References

- [1] Harry Markowitz, Portfolio Selection, The Journal of Finance, Vol.7, No.1.1952, pp. 77-91
- [2] Fischer Black and Robert Litterman, Global Portfolio Optimization, Financial Analysts Journal Vol. 48, No. 5 ,1992, pp. 28-43
- [3] William F. Sharpe, Mutual Fund Performance, The Journal of Business Vol. 39, No. 1,1966, pp. 119-138
- [4] Ross, Stephen A: The Arbitrage Theory of Capital Asset Pricing , Journal of Economic Theory 13, No.3. 1976, pp. 341-360.
- [5] Rosenberg, Barr; Extra- Market Components of Covariance in Security Returns , Journal of Financial and Quantitative Analysis 9:. 1974, pp. 263-274.
- [6] Fama, E. and K. French, Common Risk Factors in the Returns on Stocks and Bonds , Journal of Financial Economics, Vol. 33, No 1, 1993, pp. 3-54.
- [7] Fama,Eugene; French, Kenneth R, The cross-sections of the Expected Stock Returns , Journal of Finance 47, 1992. pp. 427-465.
- [8] BARRA, United States Equity Version3 (E3) ,BARRA Risk Model Handbook, 1998
- [9] Stephan, Thomas, Cz, Raimond Maurer and Margin Durr, Multiple Factor Model for European Stocks , working paper series: finance & accounting, No.57,Goethe University Frankfurt am Main, September. 2000
- [10] Richard C. Grinold, Ronald N Kahn, Active Portfolio Management: A Quantitative Approach for Providing Superior Returns and Controlling Risk, Second Edition, Professional Publishing, Columbus, OH,2000, pp. 41-147
- [11] Kalwaseth, Tarald O. Cautionary Note about R2, The American Statistician 39 No4. pp. 279-285.
- [12] Rosenberg, Barr; Marathe Vinay : Common Factors in Security Returns: Micro economic determinantsand macro correlates, Proceedings of the seminal on the analysis of security prices, University of Chicago, 1976, pp. 61-116.