Application of a Multi-factor Linear Regression Model for Stock Portfolio Optimization

Zhihao PENG, Xucheng LI
Department of Software Engineering, Dalian Neusoft University of Information, Dalian 116626, China
pengzhihao@neusoft.edu.cn

Abstract—Multi-factor models of asset pricing indicate a linear relationship between the expected return of assets while exposing to one or more risks. In this study, the sensitivity of portfolio returns to 4 selected macroeconomic factors (Market Performance, Real GDP, Inflation, Unemployment) were examined by means of multiple linear regression analyses with regard to multi-factors. Experiments show that the generalized approach of moment estimators of risk premiums lead to better results on individual assets over historical averages. Especially when the factors are weakly correlated with assets which indicates that the factors selected should be less correlated with each other.
.

Keywords- Factor Pricing Models, Optimization Model, Linear Regression

I. INTRODUCTION

Markowitz developed the portfolio construction theory in which investors should be compensated with higher returns for bearing higher risk. The Markowitz framework considers the risk as the portfolio's standard deviation, its measure of dispersion, or total risk. Sharpe (1964), Lintner (1965) and Mossin (1966) developed the Capital Asset Pricing Model (CAPM) which held that investors are compensated for bearing not only total risk, but also rather systematic risk, or market risk, as measured by the stock's beta. Investors are not compensated for taking stock specific risks, which can be diversified in the portfolio context. Stocks' beta are the slope of the stock's return regressed against the market's return. Now modern capital theory has been evolved from one beta, representing market risk, to multi-factor risk models (MFMs) with 4 or even more betas. Portfolio managers seeking the maximum return for a given level of risk using many different sets of models, based both on historical and expectation data.

The purpose of this present study is to use a multivariate statistical approach, i.e. factor analysis, to classify predictor variables according to their interrelationships and value to estimate the sensitivity of portfolio returns to 4 selected macroeconomic factors (Market Performance, Real GDP, Inflation, Unemployment).

II. PRELIMINARY DEFINITIONS

Some basic definitions of the factor Linear Regression Model are reviewed in this section[1].

In statistics, linear regression is a linear approach to modeling the relationship between a scalar response (or dependent variable) and one or more explanatory variables (or independent variables). The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression[1].

Given a data set $\{y_i, x_{i1}, \ldots, x_{ip}\}_{i=1}^{n}$, of n statistical units, a linear regression model assumes that the relationship between the dependent variable y and the p-vector of regressors x is linear. This relationship is modeled through a disturbance term or error variable ε, an unobserved random variable that adds "noise" to the linear relationship between the dependent variable and regressors. Thus the model takes the form

$$y_i = \beta_0 1 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i = \mathbf{x}_i^{\mathsf{T}} \beta + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where T denotes the transpose, so that $\mathbf{x}_i^{\mathsf{T}} \boldsymbol{\beta}$ is the inner product between vectors x_i and β.

Often these n equations are stacked together and written in matrix notation as

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2}$$

Where

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},$$

$$X = \begin{pmatrix} \mathbf{x}_1^{\mathsf{T}} \\ \mathbf{x}_2^{\mathsf{T}} \\ \vdots \\ \mathbf{x}_n^{\mathsf{T}} \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix},$$

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \quad \boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}. \tag{3}$$

III. THE MULTI-FACTOR LINEAR REGRESSION MODEL AND APPLICATION TO PORTFOLIO OPTIMIZATION

In this section a Multi-factor Linear Regression Model is proposed and the steps to implement the model is illustrated。

Generally speaking, there are 3 steps to implement the model:

(1)CAPM – Estimating beta of single stock portfolio using linear regression.

(2)Portfolio optimization – Minimizing variance and Finding weights.

(3)Estimating Portfolio Sensitivity – Multifactor Linear Regression Model.

978-1-5386-8031-5/18/$31.00 ©2018 IEEE
DOI 10.1109/ICVRIS.2018.00096

In step (1), first we need to construct the Capital Asset Pricing Model to Describes relationship between systematic risk and expected return for assets in which Beta is a measure of the volatility of a stock with respect to the market. Stock Return = (Risk Free Rate) + β (Market Risk Premium), is can be represented with the formula below:

$$R_p = R_f + \beta * (R_m - R_f)$$
$$\beta = \frac{Cov\ (Stock, Market)}{Var\ (Market)} \qquad (1)$$

The relationship between the expected return and Beta is shown in Fig.1.

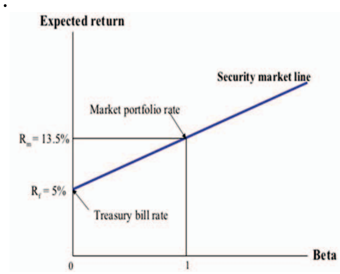

Fig. 1. Expected return vs Beta

In step(2), Markowitz mean variance portfolio is used to optimize the portfolio which combines assets to minimize the risk for a given return or maximize the return for a given risk. Another estimator Sharpe ratio is used which is a reward-to-risk ratio that focuses on total risk(as shown in Fig 2).So the Portfolio Optimization Process is demonstrated in fig 3.

$$\text{Sharpe ratio} = \frac{R_p - R_f}{\sigma_p} \qquad (2)$$

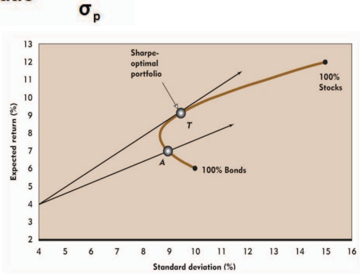

Fig. 2.Sharpe optimal portfolio

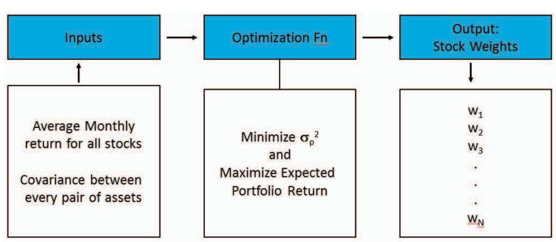

Fig. 3. Portfolio Optimization Process

To Calculating Stock Weights, first, we need to Vectorize the average returns and variances of the stocks:

$$E[\mathbf{R}] = E\left[\begin{pmatrix} R_A \\ R_B \\ R_C \end{pmatrix}\right] = \begin{pmatrix} E[R_A] \\ E[R_B] \\ E[R_C] \end{pmatrix} = \begin{pmatrix} \mu_A \\ \mu_B \\ \mu_C \end{pmatrix} = \boldsymbol{\mu}. \qquad (3)$$

$$\operatorname{var}(\mathbf{R}) = \begin{pmatrix} \operatorname{var}(R_A) & \operatorname{cov}(R_A, R_B) & \operatorname{cov}(R_A, R_C) \\ \operatorname{cov}(R_B, R_A) & \operatorname{var}(R_B) & \operatorname{cov}(R_B, R_C) \\ \operatorname{cov}(R_C, R_A) & \operatorname{cov}(R_C, R_B) & \operatorname{var}(R_C) \end{pmatrix} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} & \sigma_{AC} \\ \sigma_{AB} & \sigma_B^2 & \sigma_{BC} \\ \sigma_{AC} & \sigma_{BC} & \sigma_C^2 \end{pmatrix} = \boldsymbol{\Sigma}.$$

(4)

Where R is Vector of Average Returns of stocks, x is the Vector of Optimal Weights for stocks,E[R] is the Vector of Optimal Weights for stocks and Var(R) is the variance covariance matrix for stocks

So the equations for Expected Returns of portfolio is

$$\mu_{p,x} = E[\mathbf{x}'\mathbf{R}] = \mathbf{x}'E[\mathbf{R}] = \mathbf{x}'\boldsymbol{\mu} = (x_A, x_B, x_C) \cdot \begin{pmatrix} \mu_A \\ \mu_B \\ \mu_C \end{pmatrix} = x_A\mu_A + x_B\mu_B + x_C\mu_C. \qquad (5)$$

and the equations for risk of portfolio is

$$\sigma_{p,x}^2 = \operatorname{var}(\mathbf{x}'\mathbf{R}) = \mathbf{x}'\Sigma\mathbf{x} = (x_A, x_B, x_C) \cdot \begin{pmatrix} \sigma_A^2 & \sigma_{AB} & \sigma_{AC} \\ \sigma_{AB} & \sigma_B^2 & \sigma_{BC} \\ \sigma_{AC} & \sigma_{BC} & \sigma_C^2 \end{pmatrix} \begin{pmatrix} x_A \\ x_B \\ x_C \end{pmatrix} = x_A^2\sigma_A^2 + x_B^2\sigma_B^2 + x_C^2\sigma_C^2 + 2x_Ax_B\sigma_{AB} + 2x_Ax_C\sigma_{AC} + 2x_Bx_C\sigma_{BC}. \qquad (6)$$

Then, we can get the optimal weights of the portfolio

$$\begin{pmatrix} x_A \\ x_B \\ x_C \end{pmatrix} = \begin{pmatrix} \sigma_A^2 & \sigma_{AB} & \sigma_{AC} \\ \sigma_{AB} & \sigma_B^2 & \sigma_{BC} \\ \sigma_{AC} & \sigma_{BC} & \sigma_C^2 \end{pmatrix}^{-1} \begin{pmatrix} \mu_A \\ \mu_B \\ \mu_C \end{pmatrix}$$

$$W = \Sigma^{-1}E(R) \qquad (7)$$

In step(3),the workflow to estimate portfolio sensitivity is presented in Fig 4.

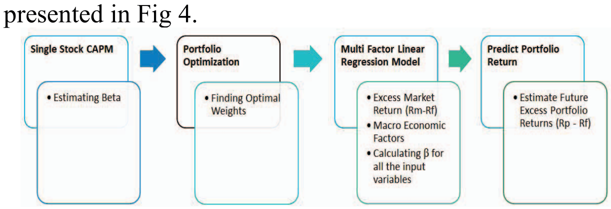

Fig. 4 Workflow for Estimating Portfolio Sensitivity

So, Portfolio Return (Rp) = CAPM + Macro Economic Factors, and

$$Rp = Rf + \beta * (Rm - Rf) + \beta 1 * Real\ GDP\ Rate + \beta 2 * Inflation\ Rate + \beta 3 * Unemployment\ Rate \qquad (8)$$

IV. RESULTS AND DISCUSSION

In this section, we tested the model with some sample data got from the real market. All the historical data for stocks and selected macroeconomic factors are fetched from Yahoo Finance. Only 8 stocks are selected, they are ABT, ACN, CME, AMZN, BBBY,CMG, GOOG, MMM. Returns (Ri) are calculated with the linear model based on the above stocks. S & P 500 index is selected as the Market Premium (Rm), the 13 week T-Bill as Risk Free Returns (Rf) while 4 Macro Economic Factors(Market Performance, Real GDP, Inflation, Unemployment) are picked as factors.

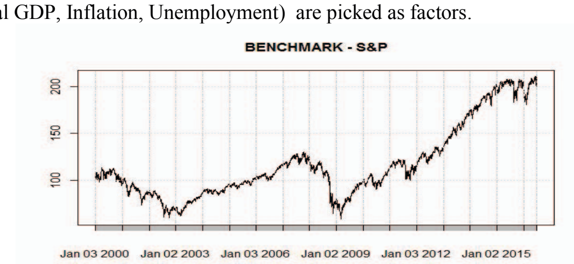

Fig. 5. S&P index from 2000 to 2015

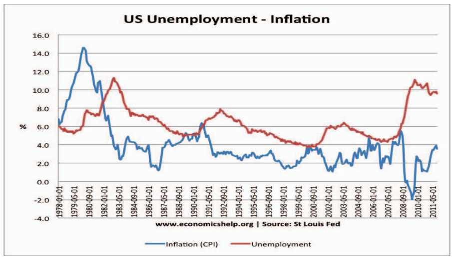

Fig. 6. US Unemployment-Inflation

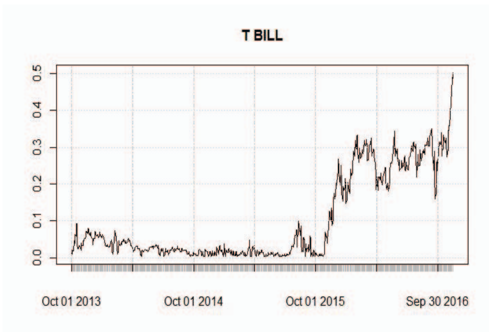

Fig. 7. 13 weeks T BILL

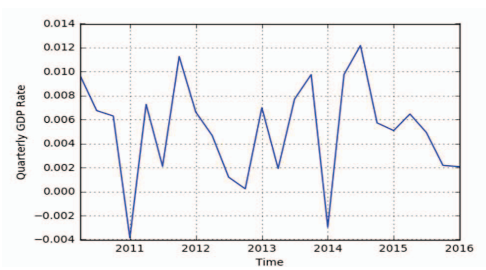

Fig. 8. Quarterly GDP

3 years historical data of date month returns are selected from September 2013 to September 2016(as shown in Fig.6-Fig.9).

Based on the model mentioned in section III, in step(1), first we need to calculate the Beta which is a measure of the volatility of a stock with respect to the market, results is shown in table 1.

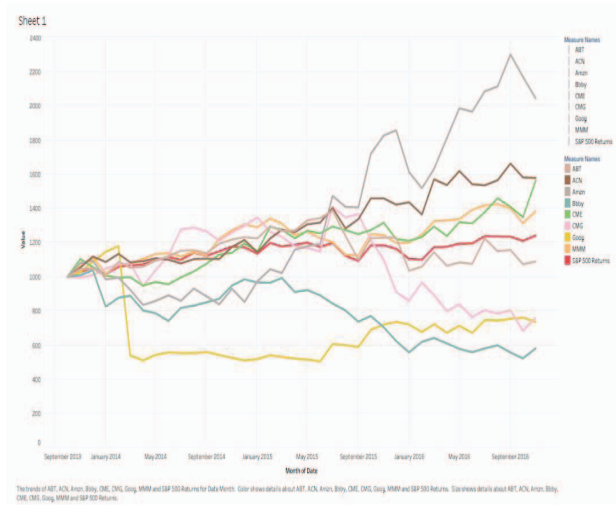

Fig. 9 The trends month return of the 8 stocks compared with S&P 500 index

Table 1. Variance – Covariance Matrix

TICKER	β FUNCTION	YAHOO (β)	GOOGLE (β)
ABT	1.52	1.55	1.33
ACN	1.03	1.08	1.20
AMZN	1.41	1.54	1.43
BBBY	1.03	0.97	0.88
CME	0.62	0.67	0.88
CMG	0.42	0.05	0.62
GOOG	0.87	0.88	0.89
MMM	0.96	1.03	1.05

* As of 12/3/16

Then is step (2), to optimize the portfolio, Covariance Matrix of the 8 selected stocks is calculated(As shown in table 2), then weighs of each stocks, average return(E[R]), stand deviation(SD) and Sharpe Ratio are formulated accordingly,finally, we can get the portfolio performance in terms of average return, stand deviation and Sharpe Ratio(Table 2-Table4).

Table 2. Variance – Covariance Matrix

	ABT	ACN	AMZN	BBBY	CME	CMG	GOOG	MMM
ABT	0.00352	0.00125	0.00226	0.00196	0.00117	0.00106	0.00216	0.00138
ACN	0.00125	0.0023	0.0018	0.00072	0.00083	0.00024	0.00212	0.00101
AMZN	0.00226	0.0018	0.00743	0.0003	0.0003	0.00082	0.00433	0.00067
BBBY	0.00196	0.00072	0.0003	0.00487	0.00198	0.00205	-0.0007	0.00178
CME	0.00117	0.00083	0.0003	0.00198	0.00259	0.00137	0.00061	0.00105
CMG	0.00106	0.00024	0.00082	0.00205	0.00137	0.00701	0.00107	0.00006
GOOG	0.00216	0.00212	0.00433	-0.0007	0.00061	0.00107	0.01172	0.00028
MMM	0.00138	0.00101	0.00067	0.00178	0.00105	0.00006	0.00028	0.00174

Table 3. Results for each stock

	Weights	E[R]	SD	Sharpe Ratio
ABT	-13.82%	0.41%	5.93%	0.07
ACN	19.25%	1.35%	4.80%	0.28
AMZN	36.32%	2.29%	8.62%	0.27
BBBY	-98.31%	-1.21%	6.98%	-0.17
CME	85.33%	1.34%	5.09%	0.26
CMG	7.16%	-0.40%	8.37%	-0.05
GOOG	-27.59%	-0.02%	10.83%	0.00
MMM	91.65%	0.97%	4.17%	0.23

Table 4. Portfolio performance

	E[R_p]	SD	Sharpe
Portfolio_Returns	4.23%	6.33%	0.67

In step (3), 4 Macro Economic Factors(Market Performance, Real GDP, Inflation, Unemployment) are picked as factors, the covariance and Beta value of each factor is shown in Fig 10 and Table 5.

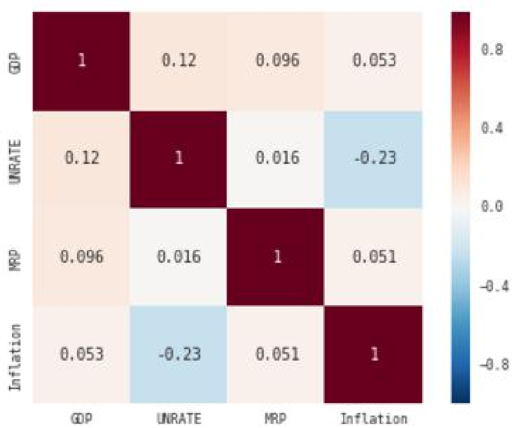

Fig. 10. Covariance of 4 macroeconomic factors

Table 5. Beta value of 4 factors

Factors	βi
GDP	5.655861
UNRATE	-0.62282
MRP	0.776697
Inflation	4.571911

We can see from the results that GDP and inflation are highly corelated with other factors, so the Bata value for them are 5.655861 and 4.571911, which influnce the portfolio return greatly. That means the more the factors corelated with each other, the less influnece will be imposed on the portfolio return.

V. CONCLUSIONS

This thesis introduces a multi-factor model and apply it to stocks portfolio optimization. First, capital asset pricing model is constructed, then the portfolio is optimized, finally multi-factors are used to estimate the portfolio sensitivity. Results show that better results could be achieved by selecting the factors that are least correlated to each other. Future work will be focused on the correlation of different factors and how to get desired better results based on them.

REFERENCES

[1] David A. Freedman. *Statistical Models: Theory and Practice.* Cambridge University Press. *p26*.2009.

[2] Bryzgalova, Svetlana. Spurious factors in linear asset pricing models. Working Paper, 2014.

[3] Burnside, Craig. Identification and inference in linear stochastic discount factor models with excess returns. Journal of Financial Econometrics, 2015.

[4] Chopra, Vijay K and Ziemba, William T. The effect of errors in means, variances, and covariances on optimal portfolio choice. Journal of Portfolio Management, 1993.

[5] Cochrane, John H. Asset pricing. New Jersey: Princeton University Press, 2001.

[6] DeMiguel, Victor, Garlappi, Lorenzo, Nogales, Francisco J and Uppal, Raman. A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms.Management Science,2009.

[7] DeMiguel, Victor, Garlappi, Lorenzo, and Uppal, Raman. Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy, The Review of Financial Studies, 2009.

[8] Fama, Eugene F. Determining the number of priced state variables in the ICAPM. Journal of Financial and Quantitative Analysis,1999.

[9] Fama, Eugene F and French, Kenneth R. The Cross-Section of Expected Stock Returns. The Journal of Finance, 1993.