Empirical tests of Fama-French three-factor model and Principle Component Analysis on the Chinese stock market

Jingjing Guo 890103-1321
Kaiwen Wang 900927-8350
Abstract

Date: 2014-06-03

Authors: Kaiwen Wang            Jingjing Guo
         fin13kwa@student.lu.se  fin13jgu@student.lu.se
         Mobile: 0762063660       0762187877

Title: Empirical tests of Fama-French three-factor model and Principle Component Analysis on the Chinese stock market

Tutor: Anders Vilhelmsson, Department of Business Administration, Lund University

Purpose: This paper aim to verify that the Fama-French three factor model (FF) captures more cross-sectional variation in returns for the Chinese stock market than the CAPM, over the period January 2004 to December 2013. Furthermore, we construct statistically optimal factors by using the principal component analysis (PCA) for the Fama-French portfolios and test whether the FF model leaves anything significant that can be explained by the PCA factors.

Method: Following the procedure in Fama and French (1993), first we construct FF factors and portfolios based on firm size and book-to-market equity, and then compare the performance between CAPM and FF models by applying time-series regressions. For deeper comparison, we continue to explain the return matrix (120*9) with principal component analysis, which produces several PCs for new time-series regressions and study the overall fitness and factor loadings of both FF and PCA models. To see which model captures the most variation, we run cross-sectional regressions with respect to all the three afore-mentioned models.

Conclusion: Our results show that the FF model tends to be more powerful than CAPM for explaining the variations in cross-sectional returns. Yet within the FF model, our data contains one divergence from the US market, we actually find a reversal of book-to-market equity effect. Finally, our results suggest that the PCA model performs better than the FF model.
Acknowledgements

We would like to express our sincere appreciation to our supervisor, Anders Vilhelmsson, who has dedicated a lot of time helping us with valuable advices. We also want to thank Mikael.o.k.karlsson for his support and useful comments.
# Table of contents

Abstract ................................................................................................................................................. 1
Acknowledgements ................................................................................................................................. 2
Table of contents .................................................................................................................................... 3
1. Introduction ........................................................................................................................................ 5
   1.1 Brief introduction ............................................................................................................................ 5
   1.2 Problem discussion .......................................................................................................................... 5
   1.3 Literature review ............................................................................................................................. 6
      1.3.1 Fama-French three factor model ............................................................................................. 6
      1.3.2 Principle component analysis ................................................................................................. 8
   1.4 Background of Chinese stock market ............................................................................................. 9
   1.5 Delimitation .................................................................................................................................... 10
2. Empirical Framework .......................................................................................................................... 12
   2.1 CAPM ........................................................................................................................................... 12
   2.2 Fama-French three factor model .................................................................................................... 12
   2.3 Cross-sectional regression ............................................................................................................. 13
   2.4 Principle component analysis ........................................................................................................ 13
3. Research methodology ......................................................................................................................... 16
   3.1 Data description ............................................................................................................................. 16
   3.2 Fama-French three factor model .................................................................................................... 18
   3.3 Principle component analysis ......................................................................................................... 19
   3.4 Regression .................................................................................................................................... 22
4. Results .................................................................................................................................................. 23
5. Analysis ................................................................................................................................................ 29
   5.1 The dependent and explanatory variables ....................................................................................... 29
   5.2 Model evaluation ............................................................................................................................... 32
   5.3 Diagnostics ..................................................................................................................................... 34
   5.4 PCA ............................................................................................................................................... 36
   5.5 Cross-sectional regressions ............................................................................................................ 38
   5.6 PCA based new FF model ............................................................................................................... 39
6. Conclusion ............................................................................................................................................ 42
7. Reference .............................................................................................................................................. 43
8. Appendix ............................................................................................................................................. 46
   8.1 Programming MatLab ...................................................................................................................... 46
   8.2 Regression results from synthetic FF model ..................................................................................... 47
List of tables

Table 1 Summary statistics for the monthly independent returns (in percent) in the regressions: January 2004 to December 2013 .......................................................23
Table 2 Descriptive statistics for 9 portfolios formed on size and book-to-market equity: 2004-2013, 10 years.................................................................23
Table 3 Regressions of stock excess returns on the market excess return: January 2004 to December 2013, 120 months .................................................................24
Table 4 Regressions of stock excess returns on the market excess return (Rm-Rf) and the mimicking returns for the size (SMB) and book-to-market equity (HML) factors: January 2004 to December 2013, 120 months ......................................................25
Table 5 Regressions of stock excess returns on the mimicking returns for the size (SMB) and book-to-market equity (HML) factors: January 2004 to December 2013, 120 months ......................................................................................26
Table 6 Regressions of stock excess returns on the PCA factors: January 2004 to December 2013, 120 months ......................................................................................26
Table 7 Cross-sectional regressions of stock excess returns on the estimated coefficients: January 2004 to December 2013, 120 months ......................................................................................27
Table 8 Correlation matrix ..................................................................................................................34
Table 9 Augmented Dickey-Fuller Unit Root Test .................................................................................34
Table 10 Breusch-Godfrey Serial Correlation LM Test .........................................................................35
Table 11 Heteroskedasticity Test: ARCH ..........................................................................................36
Table 12 R^2 for FF and PCA model ...................................................................................................38
Table 13 Correlations between original FF factors and synthetic FF factors ........................................40
Table 14 Regressions of stock excess returns on the synthetic market excess return (R_{m-R_f}) and the mimicking returns for the synthetic size (SMB) and book-to-market equity (HML) factors: January 2004 to December 2013, 120 months .................................................47

List of figures

Figure 1 SSE composite index from 2004 to 2013 .............................................................................. 17
Figure 2 Portfolios returns movement ............................................................................................... 35
Figure 3 Factor loadings for FF and PCA models .............................................................................. 37
1. Introduction

1.1 Brief introduction
Following the rapid development of the Chinese economy, the Chinese stock market has begun to show a competitive edge in the world’s financial and trading markets. Since the establishment of the Shenzhen Stock Exchange (SZSE) in 1991 and the Shanghai Stock Exchange (SHSE) in 1990, a growing number of individuals have invested in the Chinese stock market. But the problem is most of the A-share investors could not reasonably predict the risk and changes in earnings. Su (2003) proposed that domestic investors were unable to see the true nature of the equity market which leads to a confusing and aimless investment. Therefore, economists seek to find an efficient way to capture the variation in stock returns.

In early years, the most widely used model is the traditional Capital Asset Pricing model (CAPM) shaped by William Sharpe (1964) and John Lintner (1965). However, the CAPM is not accurate enough to predict the risk and explain variations in stock returns due to its unrealistic assumptions. An important reason of why it cannot describe the cross-section of expected returns is that CAPM only takes one factor, the market risk Beta, into consideration. Therefore, based on the CAPM, Fama and French (1993) proposed the well-known Fama–French three-factor model (FF model hereafter) by adding size and book-to-market factors into the regression analysis. They found that this model could significantly explain the cross-sectional variation of expected return on the US stock market.

1.2 Problem discussion
When comparing the Chinese stock market with the US stock market, it is not hard to notice that the former is an emerging market with relatively short history. Moreover, the regulatory environment in China lacks clearly defined property rights and inadequate legal protection under a transition economy. Additionally, the Chinese government plays an interfering role by retaining control of the country’s largest
enterprises. As a result, there is no clear evidence to show that the FF model can be properly applied to the Chinese stock market.

After referring to the old studies about the appropriate rules of pricing under the CAPM and FF model, we will evaluate both of them and make a comparison to see how much improvement the FF model can provide after two new risk factors (size and book-to-market ratio) are added. Besides, principal component analysis is a widely used statistical method, which reduces the dimensionality of a dataset that contains a large number of interrelated variables, while retaining as much of the variation as possible. We intend to construct the PCA factors and compare them with the FF factors. Finally, we will test which model that performs best in explaining the stock returns, by running cross-sectional regressions.

1.3 Literature review

1.3.1 Fama-French three factor model

Harry Markowitz (1952) proposed the “mean-variance” model and defined the concept of efficient frontier by explaining how a risk-averse investor constructs the optimal portfolio among a number of risky assets. Based on Markowitz’s work on diversification and modern portfolio theory, William Sharpe (1964), John Lintner (1965) and Jan Mossin (1966) independently introduced the traditional capital assets pricing model (CAPM). The traditional CAPM shows a positive linear relationship between average excess return and the risk-adjusted factor beta and, thereby providing a powerful explanation of the cross-section of expected returns. Stephen Rose (1976) proposed the arbitrage pricing theory (APT), which takes various macroeconomic factors into account to explain the expected returns of financial assets.

Banz (1981) found that market equity (ME) could be added to explain the cross-section of expected returns. He argued that the small size firms with low ME tend to have higher average returns, while big size firms, on the contrary, have lower average returns. Stattman (1980) and Rosenberg, Reid and Lanstein (1985) stated that
the firms’ book-to-market ratio should be positively related to the average returns on US stock market. Bhandari (1988) studied the effect of leverage and Basu (1983) tested if Earnings-to-price ratio could help to explain the cross-sectional variation. Based on the preliminary arguments, more recent studies by Fama and French (1992) examined the joint roles of market beta, size, book-to-market ratio, leverage and earnings-to-price ratio in average returns. They argued that beta alone cannot capture the cross-section of average returns in the U.S stock market. They found that the size and book-to-market equity were significant in explaining the cross-sectional variation in average return on the US stock market for the 1963-1990 period, while leverage and E/P ratio were not.

Fama and French (1993) extended their tests of common risk factors to both stock and bond markets. The established FF model well explained 95% variation of the excess return wherein they added two additional factors: SMB (Small minus Big: the difference between the return on a portfolio of small stocks and the return on a portfolio of large stocks) and HML (High minus Low: the difference between the return on a portfolio of high book-to-market ratio and the return on a portfolio of low book-to-market equity). They tested on 25 portfolios sorted on size and book-to-market equity from 1963 to 1990 and summarized that size and BE/ME ratio as the two mimic risk factors that play important roles in capturing variation in returns.

Fama and French (1995) strengthened the validity of their model and examined whether the fluctuation of stock prices, in relation to size and book-to-market equity (BE/ME), reflects the behavior of earnings. Concluding that the market beta and size factors in earnings could significantly help explain returns, but that there was no relationship between BE/ME factors in earnings and returns.

Positive evidence was also shown by Chui and Wei (1998) in five Pacific-Basic emerging markets (Hong Kong, Malaysia, Taiwan, Korea and Thailand). In their
paper, a weak relationship was shown between the market beta and average returns while the BE/ME gives powerful explanations of cross-sectional variation of expected returns in Hong Kong, Korea and Malaysia and the size effect is significant in all stock markets except Taiwan.

Due to the complexity of the FF model versus the simplicity of the CAPM, Phan (2007) created simple proxies for FF factors in the Japanese stock market across 33 industries by using four domestic stock indexes. In his paper, most of the market betas and size-factor betas are statistically significant, while only half of the BM-factor betas were found to be significant. The overall GMM statistics reveals that the FF model can be applied in the sample period. In the robustness check, the whole period was split into three non-overlapping sub-periods where the market risk premium and premium SMB were found to be negative in one sub-period (1990-1998) which verified a reversal of size effect by a number of researches in the Japanese market.

1.3.2 Principle component analysis
The key purpose of principal component analysis (PCA) is to reduce the dimensionality of the dataset that contains a large number of interrelated variables, meanwhile retaining as much of the variation as possible. This method is achieved by orthogonally transforming the original set of variables into new sets, called the principal components (PCs), in which the variables are linearly uncorrelated, and also PCs are ordered so that the first component retains most of the variation contained in all of the original variables.

It is generally accepted that the earliest descriptions of this technique were given by Pearson (1901) and later independently developed by Hotelling (1933). During the 25 years immediately following publication of Hotelling’s paper, it appears to have only a small amount of improvement regarding the applications of PCA. Since then, however, an explosion of new applications has occurred in respect that PCA requires a considerable computing power, and this expansion unsurprisingly coincided with the
widespread introduction of electronic computers.

However PCA does not require a priori specification of factors, which nowadays leads to its subsidiary position once being employed for asset pricing, otherwise there must be an additional step, as known as to create the mimicking portfolios, also like a realization of this statistical technique. Daskalakia, Kostakisb, and Skiadopoulou (2012) applied PCA after testing a number of traditional asset pricing models when trying to find the common factors in commodity futures returns. Kelley (2010) compared the conventional equity risk models with the PCA-based model which in the end delivers that PCA-based model, meets and sometimes exceeds the performance of standard supervised models. Xu (2007) studied the performance of a factor extracting method based on maximizing the explanatory power of the extracted factors which is mainly accomplished through PCA and compares its factors’ explanatory power with FF factors’. Moskowits (2003) also included PCA based approach to analyze covariance risk and pricing anomalies.

1.4 Background of Chinese stock market

Our research focuses on the price movement of those A-share stocks which are listed on the CSI300 index. Here, the A-shares do not refer to a “class” of common or preferred stocks as usual, in this paper, A-shares are specialized as shares of the RMB currency that are purchased and traded on the Shanghai and Shenzhen stock exchanges. As we mentioned earlier, the Chinese stock market is a young market with relative short history. Shanghai stock exchange (SSE) was established on November 26, 1990 and was in operation on December 19 of the same year. The other stock exchange, Shenzhen stock exchange (SZSE) was founded on December 1, 1990 and opened in July 3, 1991.

The CSI 300(2) stock index, abbreviated as CSI300, is a constitute stock which was published on April, 8th, 2005 by China security index co., Ltd, composed of the 300 largest A shares listed on the Shanghai stock exchange and Shenzhen stock exchange.
As the first equity index jointly launched by Shanghai and Shenzhen stock exchanges together, CSI300 aims to reflect the performance and fluctuation of China’s A-share market. The CSI300 index is a capitalization-weighted index, like the S&P500 index, which covers more than 60% of the market capitalization of Shanghai and Shenzhen stock markets. With characteristics of large-scale and high liquidity, it is designed as a performance benchmark and basis for derivatives' innovation and indexing. To determine the sample space of CSI300, ST shares (stocks with special treatment), the stocks with abnormal price fluctuations or manipulated by market, are deleted. So the CSI300 index is a representative index for analyzing the Chinese stock market.

1.5 Delimitation
Regarding the relation between BM-factor and average returns found in this paper that is opposite against Fama and French (1993), we may conjecture this phenomenon is mainly caused by the limited time horizon. Compared with Fama and French (1993), our sample period is shorter and only covers 10 years (2004-2013). The Chinese stock market also experiences a continuing “bear market” during most of this period. On the other hand, since the Chinese stock market is an emerging market, the number of stocks in our portfolios during some particular months will decrease to zero if we extend the time horizon from 10 years to 15 years. Meanwhile, we choose the top 100 stocks among 300 stocks as the proxy for the whole market, it is reasonable to believe that the results would be more significant and persuasive if more stocks were included in the sample.

Furthermore, risk-free rate is a typically proxy for the return on a one-month Treasury bill. But in China, the one-month Treasury bill has never been issued until 2007.02. To keep it consistent with our sample period, we replace it with the one-year Treasury bill. Not surprisingly, the return on the one-year Treasury bill is slightly higher than the one-month Treasury bill, which might explain the slightly negative average market risk premium over ten years.
1.6 Outline of this paper

The remainder of this paper is organized as follows. Section 2 explains the time-series and cross-sectional regressions used to test the asset-pricing model and the empirical framework of PCA factors. In section 3 we describe the dataset and explain how to construct the FF and PCA variables. The regressions and diagnostic test results are given in section 4. Section 5 presents our analysis with regard to the obtained results while section 6 concludes.
2. Empirical Framework

2.1 CAPM

Based on the “mean-variance” model which was proposed by Harry Markowitz (1959), Sharpe (1964) and Lintner (1965) introduced the traditional CAPM. The CAPM is used for pricing the individual stock or portfolio. Under the assumptions of CAPM, the expected return for any asset i at any time t is given by the following equation:

\[ E(R_i) = r_f + \beta_i [E(R_m) - r_f] \]  

(1)

where \( E(R_i) \) denotes the expected return on asset i, \( r_f \) is the risk-free interest rate, \( E(R_m) \) is the expected return on the market portfolio, and \( E(R_m) - r_f \) is the risk premium. This linear relationship is also called the security market line (SML). Since the market risk premium tends to be positive in the long run, SML indicates that there should be a positive relation between the market beta and average excess return.

\[ \beta_i = \frac{COV(r_i, r_m)}{\sigma_m^2} \]

In this equation, \( r \) is the returns in excess of the risk-free rate and \( m \) represents the market portfolio.

The time-series regression of CAPM at time t is defined as follow:

\[ r_t - r_f = \alpha + (r_m - r_f) \beta + \epsilon \]  

(2)

2.2 Fama-French three factor model

The FF model equation is given by the following equation:

\[ E(R_i) - r_f = b_1 [E(R_m) - r_f] + b_2 E(SMB) + b_3 E(HML) \]  

(3)

where \( E(R_i) \) is the expected return of any asset i, \( r_f \) is the risk-free interest rate. \( E(R_m) \) is the expected return of the market portfolio. \( E(SMB) \) and \( E(HML) \) are both expected excess return on proxy portfolios for the size factor (small minus big) and...
book-to-market factor (high minus low). The factor sensitivities $b_i$, $s_i$ and $h_i$ are the slopes in the time-series regression at time $t$ as follow:

$$R_{it} - r_f = \alpha_i + b_i(R_{mt} - r_f) + s_iSMB + h_iHML + \epsilon_{it}$$

(4)

### 2.3 Cross-sectional regression

After running time-series regressions on different models, we obtain the estimated betas for each portfolio. We then run month-by-month cross-sectional regressions with nine portfolio returns and the estimates of portfolio betas to test which model performs the best when explaining the cross-section of average excess returns. According to the theory, the time-series average value of all estimated intercepts should be equal to zero and the means of all estimated slopes are used to test whether the average premium for market beta is positive.

With respect to the one-factor asset pricing model (CAPM), the model should be organized as follow:

$$R_{pt} - r_f = \epsilon_i + \lambda_i \hat{\beta}_p + \epsilon_{p,t}, p=1,2,...,9$$

(5)

where the $\hat{\beta}_p$ denotes different estimated market betas for every FF portfolio which we can obtain from the time-series regressions in the first step. And $\lambda_i$ is interpreted as the risk premium.

With respect to the three-factor model (FF and PCA), the model is given below:

$$R_{pt} - r_f = \epsilon_i + \lambda_1\hat{\beta}_{1,p} + \lambda_2\hat{\beta}_{2,p} + \lambda_3\hat{\beta}_{3,p} + \epsilon_{p,t}, p=1,2,...,9$$

(6)

This formula gives us the risk price of SMB and HML beyond market beta and the $\lambda_1, \lambda_2, \lambda_3$ should be equal to the time-series mean of $E(R_m-r_f)$, $E(SMB)$ and $E(HML)$ if the model is constructed in an accurate way.

### 2.4 Principle component analysis

Suppose that $x$ is a vector of $p$ random variables, and then the variances and the structure of the covariance’s or correlations between the $p$ variables are targeted. However it will often not be very helpful to simply look at the variances or correlations, unless $p$ is small, or the structure is very simple, which requires an
alternative approach that is to look for a set of fewer derived variables meanwhile
preserving most of the information given by original variances, covariance’s and
correlations.

In mathematical term the first step is to look for a linear function $\alpha'_1 x$ having the
maximum variance of all $p$ elements in vector $x$, where $\alpha_1$ is a vector of $p$ constants
$\alpha_{11}, \alpha_{12}, \ldots \alpha_p$, and $'$ denotes transpose, so that

$$\alpha'_1 x = \alpha_{11}x_1 + \alpha_{12}x_2 + \cdots + \alpha_{1p}x_p = \sum_{j=1}^{p} \alpha_{1j}x_j \quad (7)$$

Next, identical way to find a linear function $\alpha'_2 x$ having maximum variance, only
being uncorrelated with $\alpha'_1 x$, and so on, till the $k$th stage a linear function $\alpha'_k x$ is
found, which has the maximum variance, also subject to being uncorrelated with $\alpha'_1 x$,
$\alpha'_2 x, \ldots, \alpha'_{k-1} x$. The $k$th derived variable $\alpha'_k x$ is the $k$th PC.

After we define PCs, coming alone is the question how to calculate them. Top of all,
the concept of covariance matrix $\Sigma$ of vector $x$ need to be introduced. This is the
matrix whose $(i, j)$th element is the covariance between the $i$th and $j$th elements of $x$
when $i \neq j$, and the variance of the $j$th element of $x$ when $i = j$. To derive the PCs, say
the first $\alpha'_1 x$ in which vector $\alpha_1$ maximizes the variance, $\text{var}[\alpha'_1 x] = \alpha'_1 \Sigma \alpha_1 $. 
Apparently for a finite $\alpha_1$, the maximum will not be achieved so a constraint must be
imposed. The constraint used in this derivation is $\alpha'_1 \alpha_1 = 1$ which indicates the sum
of squares of elements of $\alpha_1$ equals 1.

To maximize $\alpha'_1 \Sigma \alpha_1$ subject to $\alpha'_1 \alpha_1 = 1$, the standard approach is to use Lagrange
multipliers technique, that is to maximize $\alpha'_1 \Sigma \alpha_1 - \lambda(\alpha'_1 \alpha_1 - 1)$, where $\lambda$ is a
Lagrange multiplier. Then taking differentiation with respect to $\alpha_1$ gives $\Sigma \alpha_1 -
\lambda \alpha_1 = 0$ or $(\Sigma - \lambda I_p) \alpha_1 = 0$, where $I_p$ is the $(p \times p)$ identity matrix. Thus, $\lambda$ is an
Eigen value of $\Sigma$ and $\alpha_1$ is the corresponding Eigen vector. To decide which of the $p$
Eigen vectors gives the maximum variance of $\alpha'_1 x$, first note that the quantity which
is to be maximized is given by:
\[ \alpha'_1 \Sigma \alpha_1 = \lambda \alpha'_1 \alpha_1 = \lambda \alpha'_1 \alpha_1 = \lambda \quad (8) \]

That gives the sign of \( \lambda \) must be as large as possible. Thus, \( \alpha_1 \) is the Eigen vector corresponding to the largest Eigen value of \( \Sigma \), and \( \text{var}(\alpha'_1 x) = \alpha'_1 \Sigma \alpha_1 = \lambda_1 \), the largest Eigenvalue.

The second PC, \( \alpha'_2 x \), is to be produced by maximizing \( \alpha'_2 \Sigma \alpha_2 \) subject to being uncorrelated with \( \alpha'_1 x \) or equivalently \( \text{cov}[\alpha'_1 x, \alpha'_2 x] = 0 \), where \( \text{cov}(x, y) \) denotes the covariance of the random variables \( x \) and \( y \). Based on the given formula \( \text{cov}[\alpha'_1 x, \alpha'_2 x] = \alpha'_1 \Sigma \alpha_2 = \alpha'_2 \Sigma \alpha_1 = \alpha'_2 \lambda_1 \alpha'_1 = \lambda_1 \alpha'_2 \alpha_1 = \lambda_1 \alpha'_1 \alpha_2 \), any of those following equations \( \alpha'_1 \Sigma \alpha_2 = 0 \), \( \alpha'_2 \Sigma \alpha_1 = 0 \), \( \alpha'_2 \alpha_1 = 0 \), \( \alpha'_1 \alpha_2 = 0 \) could be used to specify zero correlation between \( \alpha'_1 x \) and \( \alpha'_2 x \). Then to maximize \( \alpha'_2 \Sigma \alpha_2 - \lambda (\alpha'_2 \alpha_2 - 1) - \varphi \alpha'_2 \alpha_1 \), where \( \lambda \), \( \varphi \) are the Lagrange multiplier, again differentiation with respect to \( \alpha_2 \) need to be taken, \( \Sigma \alpha_2 - \lambda \alpha_2 - \varphi \alpha_1 = 0 \).

If we multiply both sides with \( \alpha'_1 \), this gives us \( \alpha'_1 \Sigma \alpha_2 - \lambda \alpha'_1 \alpha_2 - \varphi \alpha'_1 \alpha_1 = 0 \), since the first two terms are zero and \( \alpha'_1 \alpha_1 = 1 \), it turns out \( \varphi = 0 \). Therefore, \( \Sigma \alpha_2 - \lambda \alpha_2 = 0 \), or equivalently \( (\Sigma - \lambda I_p) \alpha_2 = 0 \), which once again concludes that \( \lambda \) is an Eigenvalue of \( \Sigma \), and \( \alpha_2 \) the corresponding Eigen vector. Then \( \lambda = \alpha'_2 \Sigma \alpha_2 \), and \( \lambda \) is to be as large as possible.

As stated in previous sections, for the third, fourth, ... , \( p \)th PCs, the vectors of coefficients \( \alpha_3 \), \( \alpha_4 \) ... \( \alpha_p \), are always the Eigen vectors of \( \Sigma \) corresponding to \( \lambda_3, \lambda_4 \), ..., \( \lambda_p \), meanwhile are the third and fourth largest, ..., and the smallest Eigen value, respectively.
3. Research methodology

3.1 Data description

In our research, we collect data from CCER (China Center for Economic Research) database and DataStream. We select the top 100 largest stocks among the 300 stocks in our sample. According to the latest index weight, the top 100 stocks account for 69.04% of the whole index, and it convinces us that our sample is large enough to represent the market.

When selecting the sample stocks we apply the following criteria. First, the stock must have active trading during the twelve-month period preceding July of year \( t \), which means the stock without a trading record for more than three months will be excluded. Second, the stock should have been listed longer than two years. Third, the stock should have positive book value at the fiscal year-end in year \( t-1 \) in order to result in a positive BE/ME ratio (discussed below). Those stocks that do not satisfy those conditions above are excluded and replaced by the stock following behind them.

Our research covers the time period from January, 2004 to December, 2013, a period of 120 months. But due to the short history of the Chinese stock market, not all firms in our sample have the complete data covering our time period. For instance, many big size financial firms are only listed from 2007. Fama and French (1992) excluded financial firms in their research because they tested the effect of leverage in the regression, normally the financial firms have high leverage, and the high leverage has different meaning for nonfinancial firms. Since we do not test the leverage effect in this paper, financial firms will be included in our sample. Especially, during our sample period, after the first three years the market slumped around 2000 points and met the remarkable bull market in 2007. After reaching an all-time high of 6,124 points on October 16, the benchmark Shanghai Composite Index ended 2008 down a record 65%. Since then, the bear market started and has continued up until today.
In order to ensure that the risk premium factor is consistent with the excess return of each portfolio, when estimating the market returns for the CSI index, we compute the average expected return for the 100 stocks in the sample for each month as the market return instead of calculating the return for the whole market.

When determining the risk-free interest rate, to keep it consistent with our sample period, we use the monthly data of one-year Treasury bill instead of the one-month Treasury bill since the latter was not issued until 2007.02.

In FF three-factor model, our first step is to construct factor-mimicking portfolios in view of the size and book-to-market ratio. We use a firm’s market equity (ME) as measurement of its size, where the market equity is defined as market price times the number of shares outstanding for June of year t. With respect to the book-to-market ratio, we try to measure it in two different ways. In the first method, we take the reciprocal of price-to-book value (PTBV) at the fiscal-year end of year t-1. In the second method, we strictly follow Fama and French (1993) where the book-to-market equity is computed as the ratio between the book equity of a firm at the fiscal year-end in year t-1 and the firm’s market equity at the end of December in the preceding year. Since during our studied period, all the firms in the sample do not issue preferred stock, we just treat the book value of stockholders’ equity as the book value of common equity. In order to guarantee that the accounting information is known before the returns they are used to explain, we match the returns for July of the
year t to June of year t+1 with the accounting data for all the firms at the end of fiscal year t-1.

3.2 Fama-French three factor model

In order to examine the performance of FF factor model in the Chinese stock market, as implied in Fama and French (1993), we mimic the underlying risk factors related to size and book-to-market ratio as the explanatory variables. Firstly, when constructing the portfolios, we assume that evidence for the book-to-market ratio plays a stronger role than size in capturing the variation in stock returns based on the US market is suitable for the Chinese stock market as well. We sort firms into two groups on market equity and three groups on BE/ME exactly as in Fama and French (1993). As we mentioned above, our empirical tests focus on the primary portfolios, which consist of the top 100 firms that are chosen from the CSI300 index. We divide all the 100 firms into two groups based on the rank of their market value in the proportion 50% and 50%. It means that the Big group (B) contains 50 companies in big size and the other 50 companies are located in the Small group (S). Similarly, we also break our portfolio into three book-to-market equity groups according to the breakpoints for the bottom 30% (Low), middle 40% (Medium) and top 30% (High) of the ranked result of BE/ME.

At the end we manage to organize all the stocks into 6 specific portfolios according to the intersection of securities groups, that is, we find the intersection of each group and the six portfolios are: S/H, S/M, S/L, B/H, B/M, and B/L. (The letter in the numerator denotes the portfolios in different size groups; the letter in the denominator denotes the portfolios in different BE/ME ratio groups.) For example, the group B/L portfolio contains the stock in the big size group with a low BE/ME ratio, and S/H portfolio contains the stocks in small size group and also in high BE/ME group. In this way, we divide our original portfolio into six smaller portfolios.
Now the risk factors of SMB, HML and risk premium can be obtained. The size factor is defined as the difference between the average monthly return on small group and big group, likewise, the BE/ME factor is the difference between the returns on high and low group. Then we calculate the average monthly return for each portfolio and SMB and HML can be acquired by using the following formulas:

\[
\text{SMB} = \text{average monthly return of small size minus big size} = \frac{(S/H + S/L) - (B/H + B/L)}{2}
\]

\[
\text{HML} = \text{average monthly return of high BE/ME ratio minus low BE/ME ratio} = \frac{(S/H + M/H + B/H) - (S/L + M/L + B/L)}{3}
\]

The third risk factor that relates to beta is similar to what we have introduced in the CAPM model. It should be calculated as follows:

\[
\text{Excess return} = R_m - R_f
\]

For the dependent variables, we use excess returns on 9 portfolios formed on size and BE/ME instead of strictly following Fama and French (1993), which means we split the sample into three sub-groups based on the rank of size, following the same rule when we create the book-to-market groups. On each dimension there are three sub-groups and the intersection of them results in nine FF portfolios. The motivation of reducing the number of portfolios is because Fama and French (1993) covered three market indexes (NYSE, Amex and NASDAQ) which include 4797 stocks, on the contrary our sample only contains 100 stocks and is thereby much smaller.

3.3 Principle component analysis
Principal component analysis has been extensively used as part of factor analysis, like in our case the FF model, while PCA and factor analysis, as usually defined, are really
quite distinct techniques.

In previous sections of this paper, the FF model is estimated using monthly returns of 100 representative stocks, which then are divided into 9 portfolios according to firm size and book-to-market ratios. Given those monthly returns of nine portfolios, we would also like to use linear algebra technique such as PCA to construct the “optimal” factors.

Where the first step is to construct the portfolio returns matrix, where each column represents the return series of one portfolio (nine in total) and each row contains the monthly returns of all nine portfolios but at one time point (so 120 in total). Then all data is formed in a 120 by 9 matrix where the first row is the earliest return for each portfolio.

Step two is to calculate the nine by nine covariance matrix, denoted as $V$, with the built-in function in Excel, giving us the covariance matrix. The next step is to find the Eigen value decomposition, here and after recognized as a vector $x$ where initial values are set at random. Using Excel-solver, it is then possible to maximize the target term under certain constraints.

$$\max x'Vx$$
subject to $x'x = 1$

Solver produces a new vector $x_1$ that optimizes the equation above, and the first principal component can be computed as follow.

$$PC_{1,t} = x_1'R_t$$
where $R_t$ is a vector of portfolio returns at time $t$.

For second PC, same pattern is applied with the additional constraint.

$$\max x'Vx$$
subject to $x'x = 1$, $x_1'x = 0$ (orthogonal to $x_1$)

which works out a new vector $x_2$ that gives us $PC_{2,t} = x_2'R_t$. 
Repetition emerges when we continue to the third PC, only with the extra condition that $x'_2x = 0$ to ensure new optimal vector $x_3$ is orthogonal to $x_2$. The number of PCs that is possible to work out is as same as the dimension of covariance matrix, which in this paper is nine, yet usually not all of them will be employed for factor analysis since the fundamental idea of this method is to reduce the dimensionality of data.

There indeed have been many researches discussing the ‘stopping rule’, to determine the optimal number of components that adequately account for the total variation in an objective matrix meanwhile achieving its primary goal of reducing the dimensionality. Just like explained in many applications of PCA, we aim to replace the nine elements of returns matrix by a much smaller number of PCs, which nevertheless discard very little information. Even though there is no strict rule to determine the ‘right’ number, there does exist some widely spread and often followed guidelines to help extract the PCA factors that captures the majority of variation.

Most commonly mentioned is the cumulative percentage of variation approach. Under normal circumstance the required number of PCs ought to contribute to 70% to 90% of the total variation, and in this paper the number of three satisfies that condition at the same time being the smallest value of PC amount. So in following sections we always choose the first three principal components for analysis.

A more convenient alternative to get PCs applying professional computer software such as EViews, in which all cumulative proportions and PCs can be delivered in one spreadsheet.

After the factors are obtained, we run both time series and cross-sectional regression in the same way as with FF factors, and compare these two models’ performance specifically on the basis of the R squared coefficient.
3.4 Regression

So far, we have constructed excess return of nine portfolios sorted on the firm size and book-to-market equity as dependent variables in all time-series regressions. Meanwhile, with regard to different models (CAPM, FF and PCA), the excess market return, excess return on SMB, HML and three PCA factors are built respectively. Generally, the slopes and $R^2$ are direct evidence to judge how a model explains variation in stock returns, so we examine the explanatory power of market beta and FF factors separately, which means we test the performance of CAPM as well as a two-factor model. We then combine three factors together and examine regressions using $R_{m-r_t}$, SMB and HML factors jointly. As implied in the previous part, for comparison, we replace the FF factors by PCA factors and run time-series regressions on the same portfolios. Given the beta estimates for three models, we eventually examine cross-sectional regressions on these models and observe which model captures most of the cross-sectional variation in stock returns. In this paper we choose monthly return from 2004 to 2013 that gives us 120 time points, which implies that this procedure requires 120 repetitions of OLS regressions. Surely Excel is able to accomplish that but with loads of manual operations, and we choose to write a loop script in MatLab consequently to finish the task. The program script for FF model is attached in appendix, where the dataset includes betas for three variables as the first three columns and returns at each time point as the other 120 columns. For CAPM and PCA, all needed is to re-import the dataset and adjust the length of $lhs$ (left hand side) and first loop.
4. Results

The summary statistics for those independent variables used in the time-series regression are given below. In Panel A, we take the average value of each explanatory variable as expected risk premium. Results suggest only two out of six factors (HML and PCA2) are significant from the simple hypothesis test. Especially, the slightly negative expected return of market risk premium violates the common rule in asset-pricing model, an aspect that will be interesting to discuss. Panel B shows the correlation between each factor. We will not only focus on the correlation between factors within each model, but also analyze the relation between FF factors and PCA factors.

Table 1 Summary statistics for the monthly independent returns (in percent) in the regressions: January 2004 to December 2013

<table>
<thead>
<tr>
<th>Name</th>
<th>Mean</th>
<th>Std.</th>
<th>t-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rm-Rf</td>
<td>-0.00309</td>
<td>0.09237</td>
<td>-0.36649</td>
</tr>
<tr>
<td>SMB</td>
<td>0.003955</td>
<td>0.0457</td>
<td>0.947995</td>
</tr>
<tr>
<td>HML</td>
<td>-0.01906</td>
<td>0.063865</td>
<td>-3.26889</td>
</tr>
<tr>
<td>PCA1</td>
<td>-0.01021</td>
<td>0.276579</td>
<td>-0.40435</td>
</tr>
<tr>
<td>PCA2</td>
<td>0.02165</td>
<td>0.082701</td>
<td>2.867775</td>
</tr>
<tr>
<td>PCA3</td>
<td>0.005836</td>
<td>0.089137</td>
<td>0.717169</td>
</tr>
</tbody>
</table>

Panel B Correlations

<table>
<thead>
<tr>
<th></th>
<th>Rm-Rf</th>
<th>SMB</th>
<th>HML</th>
<th>PC1</th>
<th>PC2</th>
<th>PC3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rm-Rf</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMB</td>
<td>-0.01637</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td>-0.07538</td>
<td>-0.00874</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC1</td>
<td>0.996724</td>
<td>-0.04922</td>
<td>-0.06516</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PC2</td>
<td>0.052219</td>
<td>0.643197</td>
<td>-0.72093</td>
<td>0.022184</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>PC3</td>
<td>0.113909</td>
<td>-0.43486</td>
<td>-0.48649</td>
<td>0.164103</td>
<td>0.041298</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2 presents the annual mean of firm sizes and book-to-market equity for nine FF portfolios. Meanwhile, the number of firms and the percentage of market weights in each group are also shown to analyze the property of each portfolio. Most importantly, the average monthly return panel directly examines the size and book-to-market equity effect.
Table 2: Descriptive statistics for 9 portfolios formed on size and book-to-market equity: 2004-2013, 10 years.

<table>
<thead>
<tr>
<th>Size quintile</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average of annual averages of firm size</td>
<td>12276.45</td>
<td>12927.73</td>
<td>16780.65</td>
<td>0.208365</td>
<td>0.417837</td>
<td>0.57529</td>
</tr>
<tr>
<td>Average of annual B/E ratios for portfolio weight in portfolio</td>
<td>Small 0.055909</td>
<td>0.080461</td>
<td>0.029581</td>
<td>Small 11</td>
<td>14</td>
<td>5</td>
</tr>
<tr>
<td>Average of annual percent of market</td>
<td>Middle 0.01</td>
<td>0.01</td>
<td>-0.01</td>
<td>Middle 0.017881</td>
<td>0.010689</td>
<td>-0.01</td>
</tr>
<tr>
<td>Average of annual number of firms in portfolio</td>
<td>Big 0.014144</td>
<td>0.226002</td>
<td>0.252182</td>
<td>Big 0.207749</td>
<td>0.414882</td>
<td>0.698239</td>
</tr>
<tr>
<td>Mean</td>
<td>Small 0.031739</td>
<td>0.028243</td>
<td>0.017881</td>
<td>Small 0.33058</td>
<td>0.372558</td>
<td>0.36751</td>
</tr>
<tr>
<td>Yearly standard deviation</td>
<td>Middle 0.032451</td>
<td>0.024714</td>
<td>0.01037</td>
<td>Middle 0.340482</td>
<td>0.364748</td>
<td>0.332708</td>
</tr>
<tr>
<td>Big 0.026943</td>
<td>0.013524</td>
<td>0.01037</td>
<td>Big 0.417384</td>
<td>0.355227</td>
<td>0.330813</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Regressions of stock excess returns on the market excess return:
January 2004 to December 2013, 120 months.

\[ R(t) - R_f(t) = a + b[R_m(t) - R_f(t)] + \varepsilon(t) \]

Dependent variable: excess returns on 9 stock portfolios formed on size and book-to-market equity
Book-to-market equity(BE/ME) quintiles

<table>
<thead>
<tr>
<th>Size quintile</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00</td>
<td>Small 1.79</td>
<td>1.61</td>
<td>-0.84</td>
</tr>
<tr>
<td>b</td>
<td>0.01</td>
<td>0.00</td>
<td>-0.01</td>
<td>Middle 2.96</td>
<td>0.66</td>
<td>-2.87</td>
</tr>
<tr>
<td>t(a)</td>
<td>0.00</td>
<td>-0.01</td>
<td>-0.01</td>
<td>Big 0.60</td>
<td>-1.78</td>
<td>-2.76</td>
</tr>
<tr>
<td>t(b)</td>
<td>0.87</td>
<td>1.09</td>
<td>1.04</td>
<td>Small 15.41</td>
<td>24.82</td>
<td>20.35</td>
</tr>
<tr>
<td>b</td>
<td>1.02</td>
<td>1.05</td>
<td>0.96</td>
<td>Middle 26.37</td>
<td>22.13</td>
<td>22.27</td>
</tr>
<tr>
<td>Big 0.96</td>
<td>1.00</td>
<td>0.95</td>
<td>Big 11.04</td>
<td>19.27</td>
<td>20.43</td>
<td></td>
</tr>
</tbody>
</table>

\[ R^2 \]  
\[ S(e) \]

<table>
<thead>
<tr>
<th>Size quintile</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small 0.67</td>
<td>0.84</td>
<td>0.78</td>
<td>Small 0.06</td>
<td>0.04</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td>Middle 0.85</td>
<td>0.81</td>
<td>0.81</td>
<td>Middle 0.04</td>
<td>0.05</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td>Big 0.51</td>
<td>0.76</td>
<td>0.78</td>
<td>Big 0.09</td>
<td>0.05</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>
Table 3 tests the performance of CAPM on the nine constructed portfolios in which we examine the regressions that only use market risk premium ($R_m-R_f$) as explanatory variable. As we mentioned in section 3, the direct evidence for evaluating a model mainly includes slopes and $R^2$. From the results showed in the panel above, we notice the market betas distribute around one which then strongly supports the underlying theory of CAPM. Nevertheless, the $R^2$ locates on an unsatisfied level that motivates us to include more factors to help explain the variation in stock returns.

In table 4 we run a two-factor regression in the absence of market risk premium. It is easy to observe that $R^2$ largely decreases to a much lower level when SMB and HML factors are used. But the result still shows that the size and book-to-market equity factors can capture the variation in stock returns to a certain extent.

**Table 4 Regressions of stock excess returns on the market excess return ($R_m-R_f$) and the mimicking returns for the size (SMB) and book-to-market equity (HML) factors: January 2004 to December 2013, 120 months.**

$$R(t)-R_f(t)=a+Ssmb(t)+Hhml(t)+e(t)$$

<table>
<thead>
<tr>
<th>Size quintile</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td>t(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>0.00</td>
<td>-0.01</td>
<td>0.00</td>
<td>Small</td>
<td>-0.57</td>
<td>-0.57</td>
</tr>
<tr>
<td>Middle</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.01</td>
<td>Middle</td>
<td>-0.37</td>
<td>0.10</td>
</tr>
<tr>
<td>Big</td>
<td>-0.01</td>
<td>-0.01</td>
<td>-0.01</td>
<td>Big</td>
<td>-0.83</td>
<td>-0.75</td>
</tr>
</tbody>
</table>

| s            |      | t(s)   |      |      |        |      |
| Small        | 0.77 | 0.65   | 0.06 | Small| 4.37   | 3.10 | 0.28 |
| Middle       | -0.19| -0.28  | -0.09| Middle| -0.99  | -1.27| -0.45|
| Big          | -0.60| -0.63  | -0.61| Big  | -2.61  | -3.08| -3.28|

| h            |      | t(h)   |      |      |        |      |
| Small        | -0.44| -0.33  | 0.33 | Small| -3.46  | -2.19 | 2.16 |
| Middle       | -0.60| 0.02   | 0.27 | Middle| -4.42  | 0.11 | 1.89 |
| Big          | -0.70| 0.09   | 0.38 | Big  | -4.26  | 0.59 | 2.85 |

| $R^2$        |      | S(e)   |      |      |        |      |
| Small        | 0.21 | 0.11   | 0.04 | Small| 0.09   | 0.10 | 0.11 |
| Middle       | 0.15 | 0.01   | 0.03 | Middle| 0.09   | 0.11 | 0.10 |
| Big          | 0.18 | 0.08   | 0.14 | Big  | 0.11   | 0.10 | 0.09 |
Table 5 Regressions of stock excess returns on the mimicking returns for the size (SMB) and book-to-market equity (HML) factors: January 2004 to December 2013, 120 months.

\[
R(t) - R_f(t) = a + b[R_m(t) - R_f(t)] + S_{smb}(t) + H_{hml}(t) + e(t)
\]

Dependent variable: excess returns on 9 stock portfolios formed on size and book-to-market equity

<table>
<thead>
<tr>
<th>Book-to-market equity (BE/ME) quintiles</th>
<th>Size quintile</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Small</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Small</td>
<td>-0.12</td>
<td>-0.08</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>Middle</td>
<td>0.79</td>
<td>1.44</td>
</tr>
<tr>
<td></td>
<td>Big</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Big</td>
<td>-0.61</td>
<td>-0.57</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Small</td>
<td>0.86</td>
<td>1.09</td>
<td>1.07</td>
<td>Small</td>
<td>22.99</td>
<td>38.85</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>0.99</td>
<td>1.06</td>
<td>0.98</td>
<td>Middle</td>
<td>45.15</td>
<td>22.97</td>
</tr>
<tr>
<td></td>
<td>Big</td>
<td>0.92</td>
<td>1.00</td>
<td>0.97</td>
<td>Big</td>
<td>12.33</td>
<td>23.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Small</td>
<td>0.80</td>
<td>0.69</td>
<td>0.10</td>
<td>Small</td>
<td>10.64</td>
<td>12.22</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>-0.15</td>
<td>-0.24</td>
<td>-0.05</td>
<td>Middle</td>
<td>-3.49</td>
<td>-2.59</td>
</tr>
<tr>
<td></td>
<td>Big</td>
<td>-0.56</td>
<td>-0.60</td>
<td>-0.58</td>
<td>Big</td>
<td>-3.75</td>
<td>-6.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Small</td>
<td>-0.35</td>
<td>-0.21</td>
<td>0.45</td>
<td>Small</td>
<td>-6.37</td>
<td>-5.21</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>-0.49</td>
<td>0.13</td>
<td>0.37</td>
<td>Middle</td>
<td>-15.49</td>
<td>1.98</td>
</tr>
<tr>
<td></td>
<td>Big</td>
<td>-0.59</td>
<td>0.20</td>
<td>0.49</td>
<td>Big</td>
<td>-5.50</td>
<td>3.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Small</td>
<td>0.86</td>
<td>0.94</td>
<td>0.85</td>
<td>Small</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Middle</td>
<td>0.95</td>
<td>0.82</td>
<td>0.87</td>
<td>Middle</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>Big</td>
<td>0.64</td>
<td>0.84</td>
<td>0.95</td>
<td>Big</td>
<td>0.08</td>
<td>0.04</td>
</tr>
</tbody>
</table>

In table 5, by adding market factor to the regression equation that is shown in table 4, all the intercepts decrease to zero and the market betas do not deviate too much from the results in CAPM. In addition, R2 increases significantly for every portfolio in the sample, especially in SM, ML and BH portfolio, where it reaches 95% level and
accounts for the best fitting. Even the BL portfolio produces the lowest R2 of 0.64, there is still a significant augment compared with the CAPM. The standard deviations of the regressions also decline to a lower level.

Table 6 Regressions of stock excess returns on the PCA factors: January 2004 to December 2013, 120 months.

$R(t) - R_f(t) = a + b_1 PC_1(t) + b_2 PC_2(t) + b_3 PC_3(t) + e(t)$

<table>
<thead>
<tr>
<th>Size quintile</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th>b1</th>
<th>t(b1)</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th>b2</th>
<th>t(b2)</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th>b3</th>
<th>t(b3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.29</td>
<td>0.36</td>
<td>0.37</td>
<td>Small</td>
<td>32.52</td>
<td>35.38</td>
<td>25.51</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.01</td>
<td>0.33</td>
<td>0.36</td>
<td>0.33</td>
<td>Middle</td>
<td>27.44</td>
<td>23.83</td>
<td>27.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.29</td>
<td>0.33</td>
<td>0.33</td>
<td>Big</td>
<td>46.27</td>
<td>23.54</td>
<td>34.68</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>-0.22</td>
<td>-0.19</td>
<td>-0.20</td>
<td>0.61</td>
<td>0.43</td>
<td>-0.20</td>
<td>Small</td>
<td>20.37</td>
<td>12.68</td>
<td>-4.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle</td>
<td>0.09</td>
<td>-0.02</td>
<td>-0.23</td>
<td>0.20</td>
<td>-0.14</td>
<td>-0.18</td>
<td>Middle</td>
<td>5.11</td>
<td>-2.88</td>
<td>-4.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Big</td>
<td>0.90</td>
<td>0.08</td>
<td>-0.07</td>
<td>0.13</td>
<td>-0.35</td>
<td>-0.42</td>
<td>Big</td>
<td>6.24</td>
<td>-7.59</td>
<td>-13.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In practice, Principal component analysis can be easily applied using EViews. After importing all returns series of nine portfolios constructed above, the built-in principal
component analysis then provides Eigen values and Eigen vectors, and each of the
eigenvector contains nine numbers respectively for nine portfolios. To calculate the
principal component, by definition we need to multiply return vector at each time with
the Eigen vector, and get one time series now recognized as one principal component.
As mentioned earlier in this paper we have generated three uncorrelated factors to
explain the excess return. By running OLS regression with those three factors for each
portfolio returns, we generate exactly the same process as what we did in the FF
model. The performance of PCA model is best characterized by $R^2$, and slope on each
factor actually does not make economic sense, but we still present it for analyzing the
trend of variation in factor loadings among portfolios

Table 7 Cross-sectional regressions of stock excess returns on the
estimated coefficients: January 2004 to December 2013, 120 months

<table>
<thead>
<tr>
<th>Model</th>
<th>$c$</th>
<th>$\lambda_{mt}$</th>
<th>$\lambda_{smb}$</th>
<th>$\lambda_{hml}$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAPM</td>
<td>-0.0125</td>
<td>0.0093</td>
<td></td>
<td></td>
<td>0.12</td>
</tr>
<tr>
<td>FF3</td>
<td>-0.0249</td>
<td>0.0221</td>
<td>0.0052</td>
<td>-0.0181</td>
<td>0.58</td>
</tr>
<tr>
<td>PCA</td>
<td>-0.0328</td>
<td>0.0878</td>
<td>0.0234</td>
<td>0.0090</td>
<td>0.62</td>
</tr>
<tr>
<td>TSR coefficients</td>
<td>-0.0031</td>
<td>0.0036</td>
<td>-0.0191</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The results in table 7 provide the comparison of the performance of different models
in cross-sectional dimension. The MatLab program (see Appendix 8.1) produces a
1*120 vector for each variable that involved in CAPM, FF model and PCA model,
and then we take the means of all vector elements and sort them in one table. $R^2$ as
the most important criterion is surely included. For comparison, the average values of
the excess returns obtained in the time-series regressions are also presented in the
bottom line.
5. Analysis

5.1 The dependent and explanatory variables

The average market risk premium used in the time-series regression is only -0.309%, which indicates the overall expected return level is less than the return level of risk-free assets. Theoretically, investors would not invest in such a poor market with negative excess returns which are far from an investment perspective (normally 5%). This unexpected result may mainly be caused by the choice of risk-free interest rate. As we mentioned in the delimitation part, to ensure the data of risk-free interest rate is available during the study period, we use the expected return on a 12-month treasury bill, which is slightly greater than 1-month Treasury bill. But we note that, even if we replace the 12-month T-bill returns by 1-month T-bill returns for the post February 2007 period, this problem could be mitigated but not solved. The excess return is still around -0.1%. On the other hand, in the recent 5 years, most individual investors have to bear continuing losses due to the weak performance of the whole market. As a result, this problem may partly be explained by the low return of the overall market. The t-statistic for the market risk premium is also not significant.

The size factor SMB provides an average premium of 0.36%, but from the insignificant t-statistic (0.86) we have to infer that the size effect is not obvious in our sample. For the book-to-market factor HML, the negative risk premium is significant and in accordance with our finding that firms with high BE/ME ratio tend to have lower returns and vice versa. Among three PCA factors, both the second and third factors generate positive risk premiums and similar to the FF model, only the second factor is significant in statistical term.

In the correlation matrix, we observe that the correlation between SMB and HML is only -0.874%. The uncorrelated results indicate that the SMB proxy is largely independent from book-to-market equity effect and same for HML proxy. It confirms that these two factors capture the variation in stock returns along different dimensions.
On the other hand, the correlations between the market risk premium and SMB and HML are -1.6% and -7.5%, and when comparing with previous papers, the correlation of all three factors is much lower. But we would expect a stronger correlation between FF factors and the market premium (around 20%-30%). In Fama and French (1993), they found that market betas tend to collapse to one when adding FF factors into the model. Unfortunately this effect is not obvious in our case.

With respect to the PCA factors, since they are orthogonal to each other, the correlation between them should be extremely low, however as displayed in Table 1 the correlation between the first factor and the third factor is not low enough. By comparing the PCA factors with FF factors, we would say the first PCA factor is perfectly correlated with the market risk premium; the second PCA factor is positively correlated with SMB factor and negatively correlated with HML factor, and for the last PCA factor there ought to be a mixed effect of size and book-to-market ratio due to the nearly same negative correlation with SMB and HML.

Table 2 includes the summary statistics of 9 portfolios. As mentioned many times there are 100 stocks in our sample and after sorting the BL and SH portfolios contain the least number of stocks, 4 and 5 respectively. The numbers of stocks in the remaining portfolios are similarly around 12. The average book-to-market equity ranges from 0.20 to 0.69 since none of the firms have a BE/ME greater than one. It indicates that the price-to-book values of all the firms in the sample are larger than one, and the real value of the firms tend to be over-estimated during this period. Together, the largest three portfolios formed on size covers 82% of the total market value and 52% of the weight of the whole portfolio in the sample.

The 9 portfolios provide a virtual wide spread of average monthly return from 1.03% to 3.24%. In every BE/ME group, average returns have a tendency to decrease from the small size portfolios to big size portfolios. It then confirms the conclusion in Fama and French (1993) that there is a strong negative relation between size and average
return. In contrast, our results in book-to-market equity dimension are totally opposite to FF’s evidence. According to the same paper, there should be a stronger positive relation between book-to-market equity and average return which indicates the portfolios with a higher BM gain higher returns. But in our sample, for every group of the same size, the average returns always tend to monotonically increase from high BE/ME portfolios to low BE/ME portfolios. Since most of the firms in the high BE/ME groups are state-owned financial or industrial enterprises, we suggest this phenomenon is directly caused by those firms’ poor performance compared to the low BE/ME firms.

Fama and French (1995) found that high BE/ME firms are typically those who have distressed during the bad time whereas low BE/ME firms are typically growth stocks with high average returns on capital. Here, the academic from behavioral finance area (DeBondt and Thaler (1987), Lakonishok, Shleifer and Vishny (1994)) provide another perspective. They argued that investors tend to overreact to the past performance of those firms that are sorted on book-to-market-equity, which means the stock price is underestimated for the distressed firms (high BE/ME) and overestimated for the growth firms (low BE/ME). If the overreaction is corrected, then the return should be higher for value firms and lower for growth firms.

This odd result shed light on the singularity of this market where we reckon the participants are not as rational as postulated in the original FF model. However the firms’ sizes tend to influence the return the same way as in the US market. That phenomenon is totally reasonable, since during that period we have chosen, which is the very early stage of this industry, most of the participants in China’s stock market are new to this activity and relatively lack professional information. Also given the strong government intervention and ambiguous regulations, it is plausible that this market contains particular characters other than US market. By taking the t-test between the 9 portfolios, we cannot reject the null hypothesis that they have the same returns due to the large p-values, which indicates this result is found by chance.
5.2 Model evaluation

In Table 3, we examine the traditional CAPM model and test whether market beta explains the excess returns in the time-series regression. The results are virtually in accordance with the theory. An effective factor model should have intercept terms are close to zero. In this case, the absolute values of all portfolios’ intercept terms are less than 0.01 and the average value of market beta is close to 1. Except for the BL portfolio with a relative low $R^2$ of 0.5, the $R^2$ mainly range from 0.65 to 0.85, combining all the significant t-statistics, we could conclude that the CAPM captures most common variation in stock returns. Nevertheless, we cannot ignore the fact that no portfolios produce a $R^2$ greater than 0.9 and some intercept terms are still slightly higher. Thus, it indicates the excess return on market portfolios might leave some variation in the expected return which could be attributed to SMB and HML factors.

In Table 4, we simply test how much variation in stock returns can be captured by the size factor and BE/ME factor. In the absence of market risk premium, as expected, the $R^2$ reduce to a much lower level, only one portfolio exceeds 0.2; there are six out of nine $R^2$ are above 0.1 and especially for the portfolios in the low BE/ME group. Even the two-factor model shows poor performance, we still infer that those two factors capture some variation that risk premium factor missed and adding SMB and HML factors to the CAPM will substantially help explain the excess returns.

Table 5 shows that after adding excess market return to the regression, the three factors model (FF model) captures most of the common variation in stock excess returns. All the intercept terms are close to zero and are significant in statistical terms. At the same time, due to the slight correlation between SMB, HML and market factor, betas are more close to 1 in five portfolios. All of the market betas are significant and more than 12 standard deviations from zero.

With respect to the SMB factor, the factor loadings on SMB are related to the size. In each BE/ME group, the coefficients for SMB factors monotonically decrease from small size portfolio to big size portfolio which verifies the statistical results that
average returns tend to decrease from the small size portfolios to big size portfolios. However, in the SH and MH portfolios, the t-statistics are insignificant, less than 2 standard deviations from zero.

Similarly, the slopes on HML factors are also associated with BE/ME ratio and all the slopes are significant. In each size group, the slopes for HML factors tend to increase from the low ratio portfolio to high ratio portfolio. The contrasting results of the relation between book-to-market equity and average return still result in the same tendency with FF’s conclusion. To address this concern, we recognize that in FF’s portfolios, the firms with high BE/ME ratio tend to have higher returns, so the HML factor should be identified as a positive risk premium. Not surprisingly, portfolios belong to the high BE/ME group should be assigned a higher positive factor loading and vice versa.

In our case, contrarily the HML factor is a negative risk premium, which leads to the results that those firms, which have low BE/ME ratios and higher positive returns, should be given lower negative slopes. After constructing the three factors model, the standard deviation of the equation decreases significantly. Given the significant slopes on SMB and HML factors, we expect the new model will better explain the excess return and result in higher $R^2$.

In CAPM, no portfolios could produce the $R^2$ higher than 0.9; in the FF model, there are three portfolios that produce $R^2$ greater than 0.9 and eight out of those nine $R^2$ are higher than 0.8. The BL portfolio still results in the lowest $R^2$, but it also increases with a wide margin from 0.50 in the CAPM to 0.64 in the FF model. In the small size group, $R^2$ increases from the value between 0.50 and 0.77 to 0.64 and 0.94, similarly, for the three portfolios in the big size group, $R^2$ increases from the value between 0.66 and 0.83 to 0.84 and 0.93.
5.3 Diagnostics

To ensure that the regression results are convincible, we make some diagnostic tests. Firstly, if near multicollinearity exists, a high $R^2$ of the model accompanies with high standard error of the coefficients. As a result, the individual variables are insignificant. Similarly, the regressions become very sensitive to the explanatory variables. By checking the correlation matrix, since the correlation between three independent variables are less than 0.2 so that the possibility of multicollinearity can be excluded according to the rule of thumb (0.8).

Table 8 Correlation matrix

<table>
<thead>
<tr>
<th></th>
<th>SMB</th>
<th>RM_RF</th>
<th>HML</th>
</tr>
</thead>
<tbody>
<tr>
<td>SMB</td>
<td>1.0000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RM_RF</td>
<td>-0.01637</td>
<td>1.0000</td>
<td></td>
</tr>
<tr>
<td>HML</td>
<td>-0.00874</td>
<td>-0.07538</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Since even two uncorrelated non-stationary series will also result in high $R^2$ in the regression, to prevent spurious regressions, next, we test the stationarity of all the series: excess return for each portfolio i, excess return on the risk premium, SMB and HML. Based on the monthly data we have, we use the augmented DF test, and choose $p=12$ as the frequency to check if there are unit roots in the series. Given the results in the table, we can reject the null hypothesis at any levels and conclude that all the series in our sample is stationary. Moreover, it is more understandable to observe from the graph of returns. It is visible to note that all the series have no trending behavior and the curves frequently cross the mean of zero. Not surprisingly, stock returns are always white noise process.

Table 9 Augmented Dickey-Fuller Unit Root Test

<table>
<thead>
<tr>
<th></th>
<th>SL</th>
<th>SM</th>
<th>SH</th>
<th>SMB</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-value</td>
<td>-5.74961</td>
<td>-9.81737</td>
<td>-7.96013</td>
<td>-10.8636</td>
</tr>
<tr>
<td>p-value</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>ML</th>
<th>MM</th>
<th>MH</th>
<th>HML</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-value</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>BL</th>
<th>BM</th>
<th>BH</th>
<th>Rm-Rf</th>
</tr>
</thead>
<tbody>
<tr>
<td>t-value</td>
<td>-9.19847</td>
<td>-9.37308</td>
<td>-5.35538</td>
<td>-3.60796</td>
</tr>
<tr>
<td>p-value</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>(-0.007)</td>
</tr>
</tbody>
</table>
After running the time-series regressions with respect to the FF three factor model, we are interested in testing whether there is heteroscedasticity and serial autocorrelation exists in the residual series. To test heteroscedasticity, based on the time-series regressions we have, the ARCH test for testing whether the residual variance is constant over time is used. From the results in the table, eight out of the nine portfolios pass the test except for the MH portfolio. The p-value of it is only 0.01, which suggests us to reject the null hypothesis that states they have constant variance over time. To deal with this problem in the regression, we use robust standard errors to get rid of the heteroscedasticity as usual. Breusch-Godfrey Serial Correlation LM Test is applied to test for autocorrelation. Since all p-values are greater than 0.05, we can conclude that autocorrelation does not exist at 95% confidence level. The results also verify the Durbin-Watson test results, which we obtain from the regression output view in EViews.

**Table 10 Breusch-Godfrey Serial Correlation LM Test**

<table>
<thead>
<tr>
<th>df=5</th>
<th>L</th>
<th>M</th>
<th>H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Obs*R S</td>
<td>0.310079</td>
<td>0.97606</td>
<td>1.061793</td>
</tr>
<tr>
<td>p-value</td>
<td>-0.8629</td>
<td>-0.6278</td>
<td>-0.6025</td>
</tr>
<tr>
<td>Obs*R M</td>
<td>0.824057</td>
<td>0.110059</td>
<td>5.515462</td>
</tr>
<tr>
<td>p-value</td>
<td>-0.6752</td>
<td>-0.949</td>
<td>-0.0684</td>
</tr>
<tr>
<td>Obs*R B</td>
<td>0.128221</td>
<td>0.902074</td>
<td>3.311051</td>
</tr>
<tr>
<td>p-value</td>
<td>-0.9409</td>
<td>-0.6504</td>
<td>-0.2029</td>
</tr>
</tbody>
</table>
5.4 PCA

The so-called factor loadings of PCA are obtained and shown with clear slope. Still since they contain no explicit and direct indication of real life factor such as firm size or book-to-market equity, we choose to connect the PCA factors with the variables formed in FF model as mentioned earlier in this section.

When observing the correlation matrix in table 1, a rather notable number implies PC1 is strongly related to \((R_m-t_i)_t\), as clearly displayed in the figures below, the trend are significantly similar, while PC2 and PC3 both appear to have equal correlation with SMB and HML. More specifically speaking, there is a positive relation between PC2 and SMB, as 0.64, and negative between PC2 and HML as -0.72. The surfaces of them deliver the same information that in SMB it inclines from SL corner diagonally yet in HML it presents the opposite, and both of them work together to produce a less steep slope in PC2 chart that leaning just like it in SMB. As to PC3 which is simultaneously affected by both SMB and HML with respective correlation -0.43 and -0.48. Even though the absolute numbers are not outstanding, their influences are obvious to catch. The SH corner stands higher than others in both SMB and HML and double negative force makes SH the deep down hollow in PC3. Also BL corner in SMB and HML is low-lying while in PC3 it turns into the peak that again emphasizes the negative sign.
Furthermore, the basic idea of principal component analysis suggests that the first component usually captures the majority of variation. Then to conclude we would say in this particular market, even though firm size and book-to-marker ratio at some level affect the excess return of stocks, market return plays a rather more important role in explaining the risk of stock or portfolio excess returns. We consider it possible to explain this correlation since the market return we use is actually the average of those 100 stocks that chosen, its prior position in the correlation matrix appears to be rather understandable.

Recall the time series regression tables, besides the intercepts and the coefficients for
factors, summary statistics table also includes coefficient of determination, usually denoted as \( R^2 \) that indicates how well data points fit the statistical model. For further comparison, we extract \( R^2 \) coefficients of FF model and PCA model and put them in one table.

**Table 12 R\(^2\) for FF and PCA model**

<table>
<thead>
<tr>
<th></th>
<th>FF</th>
<th>PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average ( R^2 = 0.86 )</td>
<td>Average ( R^2 = 0.89 )</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Middle</td>
</tr>
<tr>
<td>Small</td>
<td>0.86</td>
<td>0.94</td>
</tr>
<tr>
<td>Middle</td>
<td>0.95</td>
<td>0.82</td>
</tr>
<tr>
<td>Big</td>
<td>0.64</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>Low</td>
<td>Middle</td>
</tr>
<tr>
<td>Small</td>
<td>0.93</td>
<td>0.92</td>
</tr>
<tr>
<td>Middle</td>
<td>0.88</td>
<td>0.83</td>
</tr>
<tr>
<td>Big</td>
<td>0.98</td>
<td>0.85</td>
</tr>
</tbody>
</table>

Evidently in above table, \( R^2 \) for each portfolio applied with FF and PCA are all over 0.8 besides the only one in FF model that is BL portfolio. However PCA fits almost perfectly with the highest \( R^2 \) out of all. This particular portfolio actually has already shown its unique quality earlier when constructed. In initial FF model the conclusion is that firms with bigger size and low book-to-marker ratio both tend to bring less return. Nevertheless the very few stocks from China’s market that sorted into this portfolio happen to be the leading high-yield stocks of all 100. That makes this supposed-to-stay-down-the-pyramid portfolio level up to the top of all, which then disturbs the normal function of FF model in this market. Furthermore in respect that PCA factors are generated algebraically from PCA model then performs equally well for all portfolios. All in all considering that the average \( R^2 \) of FF and PCA models are respectively 0.857 and 0.893, then we would say from the perspective of data fitness, FF model performs just as well as PCA does, which once again provide sufficient validity of FF model.

**5.5 Cross-sectional regressions**

Clearly when analyzing the cross-sectional regression outcomes, as shown in table 7,
R² of CAPM is only 0.12 which means this model delivers nearly no significance when applied for this dataset. This result is consistent with the early studies that reject the CAPM. Even there is a positive relation between excess return and beta, but it is too “flat” to describe different portfolios. Although the market risk premium is slightly different in the sample, the model brings a positive risk premium contrarily. As implied by afore-mentioned empirical framework, the dependent variable in the cross-sectional regression is the excess return for each portfolio, so we would expect an intercept of zero. But in this case, the intercept is -0.012 indicating that the zero-beta-portfolio gives lower return than r_f.

FF model, with R² of 0.58, continuously shows its advantage compared with the former, although its performance is slightly weaker than PCA’s, whose R² achieves 0.62. That is not quite a surprise because the latter is derived purely mathematically. Also, Daskalaki, Kostakis, and Skiadopoulos (2012) tested various asset-pricing models, the results clearly state that PCA works better than all other models with its R² top to 0.6 while others float around 0.3. In Xu’s (2007) paper the several R² of FF model are near 0.3 yet PCA produces R² that half of them are above 0.4.

With respect to FF cross-sectional regression, if the model is constructed properly, then λ₁, λ₂, λ₃ should be equal to the times-series mean of market risk premium, SMB and HML. By comparing those between cross-sectional regression results and time-series mean we find that they are almost equivalent for SMB and HML factors but the spread in market risk premium seems a little wide. Similarly, the intercepts for FF and PCA models are also negative, and all models predict positive market risk premiums. To conclude we still would say FF model fits Chinese market nicely and deserves a leading place in asset pricing field.

5.6 PCA based new FF model
Since previous analysis tells PCA generally functions better than FF model in data fitting, with the intention of improving FF model, we implant PCA in FF factors
construction part. More precisely the idea\(^1\) is to create new \((R_m - r_f)_t\), SMB and HML using afore-mentioned three principal components so that those FF factors not only represent their essences but also contain most of the variation of portfolio returns.

To achieve that we firstly run another regression for old \((R_m - R_f)_t\) with the three PCs extracted from original portfolio returns. That is to say treat old \((R_m - R_f)\) as the dependent variable on the left hand side and PC1, PC2 and PC3 as the independent variables on the right hand side then to regress three new \(\beta_1, \beta_2, \beta_3\).

Second is to compute new \((R_m - R_f)'_t\) by applying the new betas. Surely for SMB and HML it is the identical application. Putting this transformation in term of formula will be as follow.

\[
(R_m - R_f)'_t = \beta_{R,1}PC1_t + \beta_{R,2}PC2_t + \beta_{R,3}PC3_t
\]

\[
SMB'_t = \beta_{S,1}PC1_t + \beta_{S,2}PC2_t + \beta_{S,3}PC3_t
\]

\[
HML'_t = \beta_{H,1}PC1_t + \beta_{H,2}PC2_t + \beta_{H,3}PC3_t
\]

Afterwards new factors should be proved to contain enough credibility before being put to use. One direct approach is once again to introduce the correlation table.

**Table 13 Correlations between original FF factors and synthetic FF factors**

<table>
<thead>
<tr>
<th></th>
<th>RM-RF</th>
<th>SMB</th>
<th>HML</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.998509</td>
<td>0.791914</td>
<td>0.854025</td>
<td></td>
</tr>
</tbody>
</table>

Obviously we can say the new factors are relatively strongly connected with the old ones. Thus it is worth trying to substitute new factors into FF model.

As noted in Table 14 in appendix, the \(R^2\) of generating FF model with new factors is 0.893 which is exactly the same as previous \(R^2\) of PCA model. Continuing to cross-sectional regression the result is surprisingly notable as well. Although the coefficients of intercepts and factors here are not the same with that in any other models, the \(R^2\) (0.62) is once again identical with PCA, slightly higher than initial FF models. At this point we might be able to state this so called new FF model contains

\(^1\) [http://www.calculatinginvestor.com/2013/03/18/pca-factors-vs-fama-french-factors/](http://www.calculatinginvestor.com/2013/03/18/pca-factors-vs-fama-french-factors/)
some improvement over the old one, however the correlation between old and new factors are not always strong enough. That is to say even though new factors include more information of return matrix, but the correlation between them and the old FF factors is lessened.
6. Conclusion

Through all procedures, it is not difficult to notice that \( R^2 \) in FF model is much greater than CAPM, which leads us to conclude that FF model tends to be much more powerful than CAPM for explaining the cross-sectional variation in average stock returns. Thus PCA in the same way slightly dominates the FF model, which is also plausible following the extent literature where we also detect the statistical edge of PCA in asset pricing. Although the new FF factors that we synthesize construct a new model with the exact same \( R^2 \) as PCA factors do, which seems to have a promising improvement over the old FF model, however the correlation between new and old FF factors are not as high as expected. Therefore its validity in representing the initial FF factors still needs to be studied further. In addition, the FF model constructed based on our dataset produces a reversal of BM effect, and we attribute it to the oddness of this particular market, which is relatively more complex and affected by multiple forces other than the US market. In a general view, CAPM, FF model and PCA model applications in China’s stock market function nicely and once again show their feasibility for asset pricing.
7. Reference


8. Appendix

8.1 Programming MatLab

%%% Import dataset, and name the matrix as “data”. This dataset also include factors column, then independent variables are extracted from “data” as well.

lhs=[ones(1,9);data(:,1:3)];%%% add a ones vector to get the intercept coefficients.

a=[]; b1=[]; b2=[]; b3=[]; Rsq=[];

%%% loop for each time point
for i=4:123

coef=data(:,i)\lhs;%%% simple OLS regression

a=[a,coef(1)];%%% collecting coefficients from each regression.

b1=[b1,coef(2)];
b2=[b2,coef(3)];
b3=[b3,coef(4)];

resid2=[];
for j=1:9 %%% loop within each vector, to calculate R squared by definition.

y=data(:,i);
resid=y(j)-coef(1)-coef(2:4)*lhs(j,2:4);%%% compute individual residual.
resid2=[,resid,resid^2];
end

SSresid=sum(resid2);
SStotal=(9-1)*var(y);
Rsq=[Rsq,1-SSresid/Sstotal];

%%% after loop Rsq ends up a vector containing 120 elements.

End
8.2 Regression results from synthetic FF model

Table 14: Regressions of stock excess returns on the synthetic market excess return ($R_m - r_t$) and the mimicking returns for the synthetic size (SMB) and book-to-market equity (HML) factors: January 2004 to December 2013, 120 months.

\[
R(t) - R_f(t) = a + b(R_m(t) - R_f(t)) + S_{smb}(t) + H_{hml}(t) + e(t)
\]

Dependent variable: excess returns on 9 stock portfolios formed on size and book-to-market equity (BE/ME) quintiles

<table>
<thead>
<tr>
<th>Size quintile</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
<th>Low</th>
<th>Middle</th>
<th>High</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Small</td>
<td>-0.85</td>
<td>-0.45</td>
</tr>
<tr>
<td>Middle</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.01</td>
<td>Middle</td>
<td>1.70</td>
<td>1.54</td>
</tr>
<tr>
<td>Big</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>Big</td>
<td>-1.83</td>
<td>-0.24</td>
</tr>
<tr>
<td>t(a)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>0.87</td>
<td>1.08</td>
<td>1.08</td>
<td>Small</td>
<td>32.59</td>
<td>35.48</td>
</tr>
<tr>
<td>Middle</td>
<td>1.00</td>
<td>1.07</td>
<td>0.99</td>
<td>Middle</td>
<td>27.76</td>
<td>23.92</td>
</tr>
<tr>
<td>Big</td>
<td>0.90</td>
<td>0.99</td>
<td>0.97</td>
<td>Big</td>
<td>48.53</td>
<td>23.68</td>
</tr>
<tr>
<td>t(b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>1.14</td>
<td>0.81</td>
<td>-0.01</td>
<td>Small</td>
<td>15.58</td>
<td>9.68</td>
</tr>
<tr>
<td>Middle</td>
<td>-0.10</td>
<td>-0.31</td>
<td>0.10</td>
<td>Middle</td>
<td>-0.99</td>
<td>-2.55</td>
</tr>
<tr>
<td>Big</td>
<td>-1.93</td>
<td>-0.81</td>
<td>-0.56</td>
<td>Big</td>
<td>-38.06</td>
<td>-7.06</td>
</tr>
<tr>
<td>t(s)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>h</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>-0.29</td>
<td>-0.18</td>
<td>0.42</td>
<td>Small</td>
<td>-5.94</td>
<td>-3.16</td>
</tr>
<tr>
<td>Middle</td>
<td>-0.38</td>
<td>0.12</td>
<td>0.47</td>
<td>Middle</td>
<td>-5.74</td>
<td>1.52</td>
</tr>
<tr>
<td>Big</td>
<td>-1.48</td>
<td>0.17</td>
<td>0.46</td>
<td>Big</td>
<td>-43.84</td>
<td>2.24</td>
</tr>
<tr>
<td>t(h)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Small</td>
<td>0.93</td>
<td>0.92</td>
<td>0.85</td>
<td>Small</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Middle</td>
<td>0.88</td>
<td>0.83</td>
<td>0.87</td>
<td>Middle</td>
<td>0.04</td>
<td>0.04</td>
</tr>
<tr>
<td>Big</td>
<td>0.98</td>
<td>0.85</td>
<td>0.92</td>
<td>Big</td>
<td>0.02</td>
<td>0.04</td>
</tr>
</tbody>
</table>

S(e)