

Empirical Test of Arbitrage Pricing Model for the SSE 50 Index Stocks

Mingyue Ding¹, Yumei Pan² and Yanqiang Ding^{3,*}

¹Ji'nan Center for Marketing via Phones, China Ping'an Life Insurance Corp., Ji'nan, China

²Accounting Department, Shandong University, Ji'nan, China

³School of Economics, Shandong University, Ji'nan, China

*Corresponding author

Abstract—The objective of this paper is to test the applicability of arbitrage pricing theory specifically for the performance of the SSE 50 Index composite stocks listed on China's Shanghai Stock Market. Two groups of the SSE 50 Index stocks have been selected and tested. Principal component analysis is used to extract the common factors that influence the daily mean rate of return without dividends of the selected stocks for the period from January 2, 2014 to July 21, 2017. Two-pass regression method is employed to test the significance of the common factors in arbitrage pricing model. The main findings in this analysis are that for the first group of 15 persistently included stocks in the SSE 50 index there could not be found out significant common factors, and for the second of 11 in-then-out stocks of the SSE 50 index there could be found out two significant common factors. The results of this study suggest that we should investigate separately into specific blocks of capital markets when we are checking the applicability of arbitrage pricing theory and designing optimal quantitative investment strategies based on arbitrage pricing model.

Keywords—arbitrage pricing model; principal components analysis; SSE 50 index composite stocks

I INTRODUCTION

SSE 50 index was launched on January 2, 2004 by China's Shanghai Stock Exchange (SSE). The index is computed using the 50 most representative stocks listed on the Shanghai Stock Exchange. These composite stocks are of large scale and actively traded with sound liquidity. Arguably, these blue chip stocks are the most important potential targets of investment. However, specific analysis on these stocks is rare and outdated.

The purpose of this paper is to make a specific and in-depth analysis of the performance of the stocks in the SSE 50 index and the common factors influencing them based on the arbitrage pricing theory, using the most recent data available. This paper will examine whether the arbitrage pricing model and quantitative methods can apply to China's market of pillar blue chips. This analysis will be helpful for us to understand the core part of China's capital markets and the nature of the ballast stone stocks, and let us know whether optimal quantitative investment strategies can be designed based on arbitrage pricing model.

The research scope of this paper will be limited to two groups of the stocks that have been included in the SSE 50 index as its composite stocks. The first group includes the stocks that are persistently included in the index not only during the period

from as late as July 1, 2013 to July 21, 2017, most of the selected stocks were also in the index before July 1, 2013, and even since the inception of the index on January 1, 2004. The second group is composed of the stocks that entered and then left the SSE 50 index during the period from December 16, 2013 to July 21, 2017. The data to be analyzed will be limited to the daily *mean rate of return* (MRR) without dividends of the chosen stocks during the recent three and a half years from January 1, 2014 to July 21, 2017, when the data was downloaded from CSMAR® database.

There will be two stages in this research. Principal component analysis will be first used to discover whether there are and to extract common factors that influence the rate of return to investment in the stocks, and then the two-pass regression procedure will be employed to test the significance of the common factors and decide whether arbitrage pricing models will apply to the pricing of these stocks. Microsoft Excel will be used to clean and organize the daily trading data of the selected stocks, IBM SPSS for principal component analysis, and MathWorks' MATLAB for the two-pass regression procedure to test the significance of the common factors.

The paper is structured as follows. The following section will provide a brief review of the theoretical and empirical studies on asset pricing and China's stock markets. The third section will introduce the arbitrage pricing theory and models, and relevant analytical methodology. Data description will be supplied in the fourth section, and principal component analysis and testing for arbitrage pricing model are done in section five and six respectively. The thesis concludes in section seven with main findings of the research and suggestion for future investigation.

The major contribution of this paper to the arbitrage pricing theory literature and practical investment and risk management is its ad hoc examination on the applicability of arbitrage pricing model to the SSE 50 index stocks which are the blue chips on China's capital markets.

II REVIEW OF THEORETICAL AND EMPIRICAL LITERATURES

Ross [1] developed the arbitrage pricing theory that derives the gains of stocks from the characteristics of the gain producing process of stocks, providing a broader point of view on the market equilibrium process and the risk-return relation.

Chamberlin and Rothschild [2] first proved that extracting the eigenvectors, i.e. principal component analysis, is a good way to find the approximate factor structures. Connor and Korajczyk [3, 4] extended the asymptotic principal components analysis to identify statistically the approximate common factors and compute the risk premium to overcome the shortcomings of the factors analysis methodology to empirically specify and test arbitrage pricing models.

Ravi Shukla [5] pointed out that the assumptions of the arbitrage pricing theory may not be satisfied when the random process of rate of return of stocks is not linear, or the random variance is not totally diversified, or there might exist arbitrage opportunities.

Zhu [6] introduced arbitrage pricing theory to Chinese readers. Later on Chinese scholars have used Chinese market data to test empirically the general applicability of arbitrage pricing theory in China's stock markets and relationship between macroeconomic forces and the market performance (rate of return and its risk) of stocks. Most of the empirical analysis on China's stock markets are positive and confirm the fitness of arbitrage pricing theory, while some other disagreed. Some authors have also tried to make investment recommendations based on their research results.

Ji and Hu [7] analyzed the daily rates of return of the 50 composite stocks during the first half year after the launching of the index from January 2, 2004 to May 27, 2004. The authors drew out nine common factors that had impacts on the daily rates of return of the composite stocks included in the index. The first common factor contributed 38.56% of the variance in the daily rates of return, and the nine factors together contributed 68.4% of the variance in the daily rates of return.

Ji and Hu [7] recommended investors to choose the representative stock of each risk factor, and to create a portfolio by the correspondent weight contributed to the variance in the rate of return by each risk factor, thus to simulate the risk-return structure of the investment in SSE 50 index to get the market rate of return to investment.

Su Ping [8] used the trading data of 16 selected stocks from the 88 stocks which were in the SZSE 100 index to investigate a multi-factor arbitrage pricing model for China's Shenzhen Stock Market under the conditions of permitted and forbidden short sales respectively. The author showed that the multi-factor arbitrage pricing model can give investors reliable investment suggestions, and calculated the optimal quantitative strategies for investing in the selected stocks.

Yin and Chen [9] made empirical analysis on single factor, five-factor, and ten-factor arbitrage pricing model by means of asymptotic principal component analysis, using 463 prices of 35 stocks on China's Shanghai Stock Market and Shenzhen Stock Market. The authors found out that arbitrage pricing models with one, five, and ten factors are all better than capital assets pricing model (CAPM) in explaining the variance in the rates of return of the stocks.

III BRIEF REVIEW OF ARBITRAGE PRICING MODEL AND ANALYTICAL METHODS

A. Arbitrage Pricing Model

According to Reinganum [10], there are three assumptions in arbitrage pricing theory: capital market are perfectly competitive; investors always avoid the risk and prefer more wealth than less with certainty; a multi-factor model could represent the stochastic process of the generating asset return. Focusing on a strict factor structure, the arbitrage pricing theory holds that the expected return is linearly related to its covariance with the factors when there is no arbitrage. That is to say the factor loadings or betas are proportional to the returns' covariance with the factors. Huberman [11] expounded the mathematics of arbitrage pricing theory.

A multi-factor arbitrage pricing model can be specified as following:

$$P_i = E(P_i) + \beta_{i1}\Phi_1 + \beta_{i2}\Phi_2 + \dots + \beta_{ik}\Phi_k + \varepsilon_i \quad (1)$$

Where R_i is the return rate of company i 's stock; $E(R_i)$ is the expected return rate of company i 's stock; F_j is the deviation from the expected value of the j^{th} factor, and its expected value is zero; b_j is the sensitivity of the stock of company i to the j^{th} factor; e_i is the influence of company's own characteristic on return, and its expected value is zero and totally independent with the factors and other companies' characteristics.

In this way, the return behavior of different risk assets are linked with multi factors, and any part of change of asset return that could not be explained by the change of multi-factors just belongs to the asset.

In a large portfolio, zero-investment's return with zero-systematic-risk portfolio is zero only if the idiosyncratic effects vanish. United the theory of linear algebra and the rationales of economics, we can express the expected return on any asset i :

$$E_i = \lambda_0 + \lambda_1\beta_{i1} + \dots + \lambda_k\beta_{ik} \quad (2)$$

Where λ_0 could be regarded as the expected return rate on an asset with the risk which is zero-systematic (i.e., $b_{01} = b_{02} = \dots = b_{0k}$). $\lambda_1, \dots, \lambda_k$ are the weight of factor, and could be explicated as factor risk premium. And the relationship between the price or return of asset i and the risk premium is shown by b_i .

B. Principal Components Analysis

Principal component analysis (PCA) is a statistical process invented by Karl Pearson in 1901 that converts observation values of a set of possibly correlated variables into values of linearly irrelevant variables called principal components (or sometimes the main variation pattern) using orthogonal transformations. Principal components analysis is mostly used as a tool in exploratory data analysis and for making predictive models. Using the principal components analysis we could get the number of main components that is less than or equal to the number of original variables, the first principal component causes the greatest share of variance, and the variance caused by the following components decreases.

According to Chamberlain and Rothschild [2], when asset returns have a factor structure with a small number of factors, the covariance matrix of asset returns may be written as the sum of a diagonal matrix and a matrix of short rank. That is, in a strict k -factor asset market arbitrage pricing model, the return on the i^{th} asset is generated by

$$\rho_i = E(\rho_i) + \otimes_{i1}\Phi_1 + \dots + \otimes_{ik}\Phi_k + \varpi_i, \quad (3)$$

Where the factors F_k are uncorrelated with the idiosyncratic disturbances v_i , which are in turn uncorrelated with each other. This implies that the covariance matrix may be decomposed into a matrix of rank k and a diagonal matrix. That is, for any N , $\sum_N = B_N B_N' + D_N$, where B_N is the $N \times k$ matrix of factor loadings and D_N is a diagonal matrix.

Chamberlain and Rothschild [2] defined an approximate k -factor markets structure when there is a sequence $\{\beta_{i1}, \dots, \beta_{ik}\}$, $i \in (1, \infty)$, such that for any N , $\sum_N = B_N B_N' + R_N$, where the i, j element of the $N \times k$ matrix B_N is β_{ij} and R_N is a sequence matrices with uniformly bounded eigenvalues, and showed that extracting the eigenvectors of \sum_N is a good way of finding approximate factor structures. That is to say, principal component analysis is an appropriate technique for finding an approximate factor structure.

The procedures of asymptotic principal component analysis is composed of five steps according to Yin and Chen [9]. In general circumstances, the sample is sufficiently large, steps 4 and 5 can be skipped. The advantage of asymptotic principal component method is that it is not necessary to select factors ex ante, and the extracted common factors are linearly independent with each other and with the residual terms.

C. Two-pass Regression Method for the Test of Arbitrage Pricing Model

For the test of arbitrage pricing model, the two-pass regression method is often to use [8]. In accordance with principal components analysis, the first pass is to ensure the structure of arbitrage pricing model, and get the estimated values of the sensitivity to the factors through the multiple linear regression, and the second pass, using the estimated values of the sensitivity to the factors as explanatory variables, to check whether arbitrage pricing theory is established using cross section regression.

$$\rho_{i\tau} = \alpha_i + \otimes_{i1}\Phi_{1\tau} + \otimes_{i2}\Phi_{2\tau} + \dots + \otimes_{ik}\Phi_{k\tau} + \varepsilon_{i\tau} \quad (4)$$

$$\rho_i = \lfloor_0 + \lfloor_1 \otimes_{i1} + \lfloor_2 \otimes_{i2} + \dots + \lfloor_k \otimes_{ik} + \varepsilon_i \quad (5)$$

r_i is the mean return rate of the stock i , that is the sum of daily return rate divided by the number of days. $\beta_{i1}, \beta_{i2}, \dots, \beta_{ik}$ are the loading values obtained by principal components analysis, that is, the sensitivities. In generally, we do the regression with all the principal components which were extracted. However, there has another method to do the regression, which is stepwise regression. Stepwise regression is a method of fitting regression models in which predictive variables are selected by automated procedures. It is basically repeated several times, each time removing the weakest relevant variables, then the best variables to explain the distribution are left.

IV DESCRIPTION OF DATA AND ITS SOURCES

SSE 50 index takes December 31, 2003 as its base day, and its base point is 1000. Its 50 composite stocks are adjusted half a year according to the principles of sample stability and dynamic tracking, and every time of adjusting at most 10% of the composite stocks can be changed. Tentative adjustment can be made under some special cases such as when some stocks emerge to be among the top stocks or acquisition and merging occurred.

As of July 21, 2017, the composite stocks of SSE 50 index have been adjusted normally for twenty-seven times and tentatively for fourteen times. There are a total of 163 stocks that have been selected as composite stocks of the index for a total of 196 inclusions into the index and a total of 146 exclusions from the index. Thirty stocks have been remained in the index during this research period. These stocks are presented in Table 1. The data in the rightmost three columns can be found in Fact Book 2016 [12].

* 600887 was excluded from it on January 5, 2009; * 601766 was excluded on January 4, 2011; * 601688 was excluded on January 4, 2012; ** Included in the index 10 trading days after their initial public offer; † Used for principal components analysis later on because they had data on all the trading days. Sources: CSMAR® database provided by Shenzhen GTA (GuoTaiAn) Information Technology Corporation; Shanghai Stock Exchange [12].

TABLE I. PERSISTENT COMPOSITE STOCKS OF SSE 50 INDEX FROM 2014 TO 2017

Stock Code	Short Name of the stocks	Date of Inclusion	Date of IPO	Number of Tradable Shares [10,000 shares]	Closing Price [RMB yuan]	Weight in SSE 50 Index [%]
600000	Pudong Development Bank	2004/1/2	1999/11/10	1865347	18.27	4.20
600016†	Minsheng Bank	2004/1/2	2000/12/19	2955177	9.64	6.75
600028†	China Petroleum & Chemicals	2004/1/2	2001/8/8	9555777	4.96	1.27
600030†	CITIC Securities	2004/1/2	2003/01/06	981466	19.35	3.76
600036	China Merchants Bank	2004/1/2	2002/4/9	2062894	17.99	4.56
600050†	China Unicom	2004/1/2	2002/10/9	2119660	6.18	1.30
600104	Shanghai Automobile Industry Group	2004/1/2	1997/11/25	1102557	21.22	1.71
600887*	Yili Group	2004/1/2; 2012/7/2	1996/3/12	601565	16.43	2.43
600519†	Kweichow Moutai	2005/7/1	2001/8/27	125620	218.19	2.66
601006***†	Daqin Railway	2006/8/15	2006-08-01	1486679	8.62	1.26
601398**	Industrial and Commercial Bank	2006/11/10	2006/10/27	26961221	4.58	2.41
600048	Poly Real Estate	2007/1/4	2006/7/31	1075525	10.64	1.70
601628***†	China Life	2007/1/23	2007/1/9	2082353	28.31	1.15
601166**	Industrial Bank	2007/2/19	2007/2/5	1617962	17.07	5.57
601318**	China Ping An Insurance	2007/3/15	2007/3/1	1083266	36.00	9.39
601328***†	Bank of Communications	2007/5/29	2007/5/15	3925086	6.44	3.69
601088***†	China Shenhua	2007/10/23	2007/10/9	1649104	14.97	0.72
601857***†	Petro China	2007/11/19	2007/11/5	16192208	8.35	0.99
601601†	China Pacific Insurance	2008/7/1	2007/12/25	628670	28.86	2.20
600837**	Haitong Securities	2009/1/5	1994/2/24	809213	15.82	3.16
601169	Bank of Beijing	2009/1/5	2007/9/19	1267223	10.53	2.61
601766*	CRRC	2009/7/1; 2011/7/1	2008/8/18	2291769	12.85	2.88
601668†	China State Construction	2010/1/4	2009/7/29	2990056	6.34	2.34
601288***†	Agricultural Bank of China	2010/7/29	2010/7/15	29405529	3.23	2.99
601818**	Everbright Bank of China	2011/1/4	2010/8/18	3981036	4.24	1.64
601688*	Huatai Securities	2011/1/4; 2013/1/4	2010/2/26	544372	19.72	1.59
600111†	Northern Rare-Earth	2011/1/4	1997/9/24	221920	14.02	0.75
601989†	China Shipbuilding Industry	2011/7/1	2009/12/16	1795786	9.40	2.11
600518	Kangmei Pharmaceutical	2013/07/01	2001/3/19	439743	16.95	1.26
600999	China Merchants Securities	2013/07/01	2009/11/17	480310	21.70	1.54

During the period from December 16, 2013 to July 21, 2017, there are eight normal adjustments and one times tentative adjustment of the composite stocks. There are forty-one stocks being included into and forty-one stocks excluded from the SSE

50 index at the same time. Among the newly included stocks, twenty-one stocks were excluded from the index later on during the same period. These stocks are presented in Table 2.

TABLE II. IN-THEN-OUT COMPOSITE STOCKS OF SSE 50 INDEX FROM 2014 TO 2017

Stock Code	Short Name of the stocks	Date of Inclusion	Date of exclusion	Date of IPO	Number of Tradable Shares [10,000 shares]	Closing Price, 2005 [RMB yuan]
600018	Shanghai International Port Group	2013/12/16	2016/06/13	2000/07/19	2275518	6.48
600637	BesTV (Shanghai Oriental Pearl Media)	2013/12/16	2017/06/12	1993/03/16	215844	37.89
600406	NARI Technology Development	2013/12/16	2015/12/14	2003/10/16	220575	16.68
600332	Guangzhou Baiyunshan	2013/12/16	2015/06/15	2001/02/06	103634	30.27
601117†	China National Chemical	2013/12/16	2014/12/15	2010/01/07	493300	6.89
600089†	TBEA	2014/06/16	2015/12/14	1997/06/18	317812	11.77
600196†	Fosun Pharmaceutical	2014/06/16	2015/06/15	1998/08/07	190561	23.49
600703	San'an Optoelectronics	2014/06/16	2015/06/15	1996/05/28	237038	24.28
601118†	China Hainan Rubber Industry Group	2014/06/16	2015/06/15	2011/01/07	393117	7.56
600832	Shanghai Oriental Pearl Group/Old	2014/06/16	2015-05-20 delisted	1994/02/24	n.a.	n.a.
600690	Qingdao Haier	2014/12/15	2015/12/14	1993/11/19	550947	9.92
601998†	China CITIC Bank Corp	2014/12/15	2017/06/12	2007/04/27	3190516	7.22
600109†	Sinolink Securities	2014/12/15	2017/06/12	1997/08/07	283686	16.12
600150†	China State Shipbuilding	2014/12/15	2016/06/13	1998/05/20	137812	34.83
600372†	Aviation Electronics	2014/12/15	2015/06/15	2001/07/06	175916	24.63
600893	Aero-engine	2015/05/20	2017/06/12	1996/04/08	136514	45.03
600583†	Offshore Oil Engineering	2015/05/20	2015/12/14	2002/02/05	404901	8.95
600795†	GD Power Development	2015/12/14	2016/12/12	1997/03/18	1781554	3.93
601669	Power Construction Corporation of China	2015/12/14	2016/12/12	2011/10/18	330000	8.03
601377†	Industrial Securities	2016/06/13	2017/06/12	2010/10/13	520000	11.00
601919	COSCO SHIPPING Holdings	2016/06/13	2016/12/12	2007/06/26	763567	9.02
601727	Shanghai Electric Group	2016/06/13	2016/12/12	2008/12/05	985100	11.54

† Used for principal components analysis later on because they had data on all the trading days. Sources: CSMAR® database provided by Shenzhen GTA (GuoTaiAn) Information Technology Corporation; Shanghai Stock Exchange [12].

We can observe from Table 1 that among the 30 stocks which are remained inclusion in the SSE 50 index there are eight stocks which were originally included in the index when the index was initiated on January 2, 2004; there are nine stocks which were included tentatively in the index 10 trading days after their IPO in 2006, 2007, and 2010; there are seven stocks that were included in the index within half a year or around a year from 2006 to 2010; there are only two stocks which were lately included in July, 2013.

There are nine stocks in the banking sector, four in the security sector, three in the insurance sector, taking weights of 32.78%, 10.05%, and 12.74% in the SSE 50 index respectively; thus the financial sector takes a weight of 55.57% in the SSE 50 index. There are two stocks in the petroleum and chemicals sector, one in the coal and electricity sector (601088), one in the telecommunication sector (60050), two in the railway and locomotive sector (601006 and 601766), together the energy and infrastructure sectors take a weight of 8.42% in the index; the rest eight stocks are in the sectors of manufacturing, construction (601668), mineral (600111), and real estate (600048), taking a weight of 14.96% in the index. In total, the persistent composite stocks of the SSE 50 index take a weight of 78.95% in the index in 2015 [12].

The daily trading data of the thirty persistent stocks of the SSE 50 index and that of the twenty-one in-then-out stocks from January 1, 2014 to July 21, 2017 is retrieved from CSMAR® database provided by Shenzhen GTA (GuoTaiAn) Information Technology Corporation at the website of the library of China's Shandong University. Because of missing trading data for some of the chosen stocks, later on in this paper fifteen of the thirty persistent stocks and eleven of the in-then-out stocks with full trading data (as indicated in Table 1 and Table 2 by a † symbol) will be selected to do principal components analysis and test the applicability of arbitrage pricing theory.

V ANALYSIS OF THE DATA USING PRINCIPAL COMPONENT ANALYSIS

In this paper, IMB SPSS 23 was used to do the principal component analysis for both the fifteen persistent stocks (first data set) and eleven in-then-out stocks (second data set) with full trading data. The result is shown in Figure I and Figure II.

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	8.860	59.064	59.064	8.860	59.064	59.064
2	1.333	8.887	67.952	1.333	8.887	67.952
3	.802	5.348	73.299			
4	.690	4.601	77.901			
5	.632	4.214	82.115			
6	.497	3.312	85.427			
7	.397	2.644	88.071			
8	.368	2.455	90.526			
9	.308	2.054	92.580			
10	.274	1.830	94.410			
11	.262	1.748	96.158			
12	.200	1.333	97.491			
13	.151	1.008	98.499			
14	.122	.811	99.310			
15	.103	.690	100.000			

Extraction Method: Principal Component Analysis.

FIGURE I. TOTAL VARIANCE EXPLAINED FOR THE 15 PERSISTENT STOCKS

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.361	57.830	57.830	6.361	57.830	57.830
2	1.390	12.640	70.470	1.390	12.640	70.470
3	.640	5.820	76.290			
4	.543	4.939	81.229			
5	.475	4.314	85.543			
6	.368	3.350	88.893			
7	.327	2.975	91.868			
8	.316	2.871	94.739			
9	.243	2.213	96.952			
10	.203	1.842	98.795			
11	.133	1.205	100.000			

Extraction Method: Principal Component Analysis.

FIGURE II. TOTAL VARIANCE EXPLAINED FOR THE 11 IN-THEN-OUT STOCKS

In Figure I, there are 15 factors influencing the rate of return. According to the principle that the eigenvalues is greater than one to extract factors, percentage of variance and the cumulative percentage explained by the factors is as shown. Two principal components were extracted, they can explain 67.95% of the variance in the set of data. Figure II shows that there are 11 factors influencing the rate of return, and also two principal components were extracted, they can explain 70.47% of the variance in the set of data.

Figure III shows the loading value matrix of the two principal components for the first and second data set. There lists the loading values of the two components, which are sensitivity of the rate of return to the common factors. The loading values are the basis for regression analysis next.

Component Matrix^a

	Component	
	1	2
\$600016	.776	-.384
\$600028	.829	.015
\$600030	.749	.199
\$600050	.728	.341
\$600111	.604	.565
\$600519	.582	.142
\$601006	.741	.285
\$601088	.834	.171
\$601288	.831	-.395
\$601328	.840	-.312
\$601398	.794	-.445
\$601601	.836	-.071
\$601628	.804	-.079
\$601668	.732	.332
\$601857	.788	-.084

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Component Matrix^a

	Component	
	1	2
\$601998	.697	-.530
\$600089	.742	.096
\$600109	.832	-.266
\$600150	.823	.228
\$600196	.811	-.292
\$600372	.589	.610
\$600583	.712	.364
\$600795	.884	-.200
\$601117	.746	.238
\$601118	.647	.396
\$601377	.849	-.366

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

FIGURE III. LOADING VALUE MATRIX OF COMPONENTS FOR THE 15 PERSISTENT STOCKS (LEFT PART) AND THE 11 IN-THEN-OUT STOCKS (RIGHT PART)

There is no significant difference in the coefficients of the original variables in each factor, and the coefficients of the individual variables in the factor can be differentiated to the maximum and minimum poles by rotation. We use the mean orthogonal rotation, so that the number of variables with a high loading value on a factor and the number of factors in the variable are minimized. After the rotation, the loading value matrix is as Figure IV.

Rotated Component Matrix^a

	Component	
	1	2
\$600016	.830	.245
\$600028	.598	.574
\$600030	.415	.654
\$600050	.303	.745
\$600111	.060	.825
\$600519	.331	.499
\$601006	.351	.712
\$601088	.496	.692
\$601288	.879	.274
\$601328	.828	.341
\$601398	.885	.212
\$601601	.663	.515
\$601628	.644	.488
\$601668	.312	.741
\$601857	.636	.474

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

Rotated Component Matrix^a

	Component	
	1	2
\$601998	.873	.065
\$600089	.492	.564
\$600109	.799	.352
\$600150	.466	.716
\$600196	.801	.319
\$600372	.036	.847
\$600583	.292	.744
\$600795	.780	.423
\$601117	.401	.673
\$601118	.222	.726
\$601377	.878	.289

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

FIGURE IV. LOADING VALUE MATRIX OF COMPONENTS FOR THE 15 PERSISTENT STOCKS (LEFT PART) AND THE 11 IN-THEN-OUT STOCKS (RIGHT PART) AFTER ROTATED

VI TEST OF ARBITRAGE PRICING MODEL BY TWO-PASS REGRESSION METHOD

A. Regression for the First Data Set

For the first data set, regress the mean rate of return to the coefficients of two principal components (loading values) based on the Eq. 5, and the result is shown as Fig. 5. Beta₁ is β_{11} and Beta₂ is β_{12} . $R^2 = 0.12$ means the coefficients of the two principal components can explain 12% of the variance in the

mean rate of return. So, the mean rate of return might not have linear relationship with the coefficients of the two principal components. Besides, the coefficients (λ_0 's) of the β_0 's are not significant, and β_{11} is more significant than β_{12} . T-test shows that λ_1 could be zero with 32.9% possibility, and λ_2 could be zero with 98% possibility. So, the two principal components are not significant.

Considering that loading value of the two principal components may affect each other, we regress the mean rate of return to β_{11} and β_{12} respectively. The results are shown as Fig. 6 and Fig. 7. As we can see, the significant level of both two λ 's are increased. There is 79.4% possibility that λ_1 could not be zero, and 54.7% possibility that λ_2 could not be zero. However, they didn't achieve the 90% significant level. In addition, using the stepwise regression method to do regression, no variables entered into the equation.

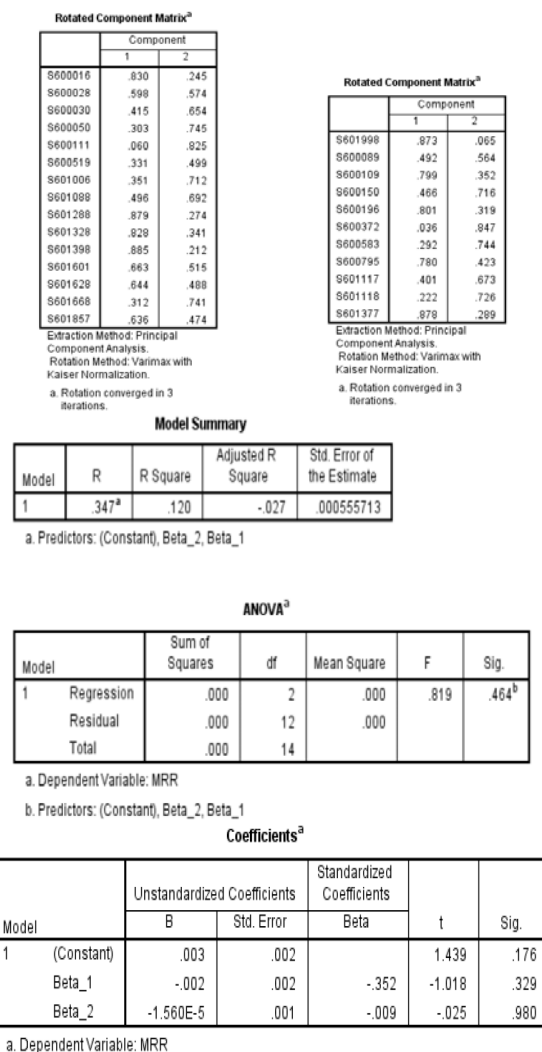


FIGURE V. REGRESSION TO THE COEFFICIENTS OF THE TWO PRINCIPAL COMPONENTS

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.346 ^a	.120	.052	.00533926

a. Predictors: (Constant), Beta_1

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	.000	1	.000	1.773	.206 ^b
Residual	.000	13	.000		
Total	.000	14			

a. Dependent Variable: MRR

b. Predictors: (Constant), Beta_1

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	.003	.001		1.894	.081
Beta_1	-.002	.002	-.346	-1.332	.206

a. Dependent Variable: MRR

FIGURE VI. REGRESSION TO THE COEFFICIENT OF THE FIRST PRINCIPAL COMPONENT

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.210 ^a	.044	-.030	.00556510

a. Predictors: (Constant), Beta_2

ANOVA^a

Model	Sum of Squares	df	Mean Square	F	Sig.
1 Regression	.000	1	.000	.599	.453 ^b
Residual	.000	13	.000		
Total	.000	14			

a. Dependent Variable: MRR

b. Predictors: (Constant), Beta_2

Coefficients^a

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
1 (Constant)	.001	.000		5.369	.000
Beta_2	.000	.000	.210	.774	.453

a. Dependent Variable: MRR

FIGURE VII. REGRESSION TO THE COEFFICIENT OF THE FIRST PRINCIPAL COMPONENT

B. Regression for the Second Data Set

For the second data set, there has 11 stocks, and two principal components. As well as the first data set, regress the mean rate of return to the coefficients of the two principal components (loading values) based on the Eq. 5. This time, use the stepwise regression method directly.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.945 ^a	.893	.881	.000130746
2	.966 ^b	.937	.921	.000106470

a. Predictors: (Constant), Beta_2

b. Predictors: (Constant), Beta_2, Beta_1

ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	.000	1	.000	75.130	.000 ^b
	Residual	.000	9	.000		
	Total	.000	10			
2	Regression	.000	2	.000	59.434	.000 ^c
	Residual	.000	8	.000		
	Total	.000	10			

a. Dependent Variable: MMR

b. Predictors: (Constant), Beta_2

c. Predictors: (Constant), Beta_2, Beta_1

FIGURE VIII. MODEL SUMMARY FOR THE SECOND DATA SET WITH STEPWISE REGRESSION

Figure VIII illustrates the order of the factors which are introduced. According to the results of the variance analysis, the factor which has the greatest contribution to the return rate variance is introduced first. The first model introduces the factor β_{i2} and the second model introduces the factor β_{i1} . R^2 of the first model is 0.893, meaning that the first model can explain 89.3% of the variance in the mean rate of return. When the factor β_{i1} is added, R^2 increases to 0.937. In F-test, $F = 59.434$ is high as well, that means factors β_{i1} and β_{i2} could explain the variance in the mean rate of return significantly, and the mean rate of the return has a linear relationship with the factors β_{i1} and β_{i2} .

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.001	.000		14.549	.000
	Beta_2	-.001	.000	-.945	-8.668	.000
2	(Constant)	.000	.000		-.810	.442
	Beta_2	-.001	.000	-.773	-6.725	.000
	Beta_1	.001	.000	.271	2.361	.046

a. Dependent Variable: MMR

FIGURE IX. COEFFICIENT OF THE STEPWISE REGRESSION

Figure IX lists the unstandardized coefficients, standardized coefficients, hypothesis test of regression coefficient (t-test) and the P-value. According to t-test, the P-value for factors β_{i1} and β_{i2} are 0.046 and 0.000, respectively. At 95% significance level, the coefficient of factors β_{i1} and β_{i2} are significant. So, this data set is suitable for analysis by arbitrage pricing model.

VII SUMMARY AND FURTHER WORK

The main findings in this paper is that the first data set of the 15 persistently included stocks in the SSE 50 index is not consistent with the arbitrage pricing theory, that is, there could not be found out significant common factors for this group of stocks, and the 11 in-then-out stocks of the SSE 50 index is

consistent with the APT, that is, there could be found out two significant common factors for this group of stocks. This is different from the much more general outcomes in the academic literatures about the applicability of arbitrage pricing theory to China's stock markets.

This study suggests that we should investigate separately into specific blocks of capital markets when we are checking the applicability of arbitrage pricing theory and designing optimal quantitative investment strategies accordingly. Future studies must further examine the reasons why the markets of stocks which are included in the SSE 50 index performed differently in terms of characteristics of the industries and governance, long term variance in their rates of return, and trading dynamics. In a separate paper we are going to calculate the optimal quantitative investment strategies for the 11 in-then-out stocks of the SSE 50 index. We should further check the applicability of arbitrage pricing model to other groups of the SSE 50 Index stocks in earlier and later periods than that in this paper.

ACKNOWLEDGEMENT

This paper is part of the first author's dissertation - Quantitative Portfolio Strategy: An Empirical Analysis of SSE 50 Index Stocks Based on APT Model, submitted in partial fulfilment of the degree of MSc in Financial Engineering (2016/2017), University of Birmingham, The United Kingdom. The author (Mingyue Ding, ID: 1727011) gratefully appreciates the pertinent and patient supervision by her tutor Tatiana Gaydaenko.

REFERENCES

- [1] S. Ross, The arbitrage Theory of Capital Asset Pricing, Journal of Economic Theory, 23 (1976) 341-360.
- [2] Chamberlain, Gary, Michael Rothschild, Arbitrage, Factor Structure, and Mean Variance Analysis on Large Asset Markets, Econometrica, 51 (1983) 1281-1304.
- [3] Connor, Gregory, Robert A. Korajczyk, Performance measurement with the arbitrage pricing theory: A new framework for analysis, Journal of Financial Economics, 15 (1986) 373-394.
- [4] Connor, Gregory, Robert A. Korajczyk, Risk and return in an equilibrium APT: application of a new test methodology, Journal of Financial Economics, 21 (1988) 255-290.
- [5] Ravi Shukla, An Empiricist's Guide to the Arbitrage Pricing Theory, Working Paper of the 10th Conference of America's Financial Management Association, Florida (1999).
- [6] Zhu Jing, the 9th Lecture on the Theories and Practices of Security Markets: Portfolio Choice Theory, World Economic Papers, 1993, pp.55-60, 79.
- [7] Ji Jun, Hu Qinglin, A Study on the Arbitrage Pricing of the SSE 50 Index, Statistics and Decision, 12 (2004) 32-33.
- [8] Su Ping, A Case Study of Multi-factor APT Model Used in Shenzhen Stock Market, Journal of Zhejiang Business Technology Institute, 1 (2007) 12-17.
- [9] Yin Xiangfei, Chen Liuqin, An Empirical Study of Arbitrage Pricing Theory on China's Stock Markets: Based on the Asymptotic Principal Components Methodology, Journal of the Henan Institute of Financial Management, 1 (2008) 90-95.
- [10] M. Reinganum, The Arbitrage Pricing Theory: Some Empirical Results, The Journal of Finance, 2 (1981) 313-321.
- [11] G. Huberman, A simple approach to arbitrage pricing, Journal of Economic Theory, 28 (1982) 183-91.
- [12] Shanghai Stock Exchange, Fact Book 2016, pp.35-42.