A safety-first portfolio selection framework: Estimating returns of exchange traded funds through regression analysis

M N Young
School of Industrial Engineering and Engineering Management, Mapu’a University, Manila, Philippines
E-mail: mnyoung@mapua.edu.ph

Abstract. This research provides a possible variant of the basic framework of portfolio selection. First, for the return estimation, considering Asian exchange-traded funds as the selection pool for the investment, regression analysis is applied to generate future return scenarios. Second, for the assignment of weights, the likelihood of these scenarios are assumed to be equally likely. Third, for the portfolio selection model, the safety-first model is utilized with consideration of 5 different diversification factors. The proposed procedure is modeled as a linear programming model into AIMMS optimization software where the resulting portfolios show that safety-first portfolios can outperform the benchmark. These observations can pave the way to a new generic portfolio selection framework that can possibly help investors and may lead to alternative investment instruments.

1. Introduction

The basic framework of portfolio selection is composed of 3 parts namely return estimation, assignment of weights, and portfolio selection model. In the first part, this is where investment pools are identified and future returns are estimated based on the individual perception of investors on future performances of the market. Typically, one can use the historical returns as rough estimates for future returns or they can use sophisticated models to generate return scenarios e.g. regression equations and other technical forecasting techniques. After having these return estimates, the second part of portfolio selection is to provide appropriate weights to these future returns or return scenarios. This is still based on the individual perception of the investor where he/she provides the appropriate weights to the corresponding return scenarios. Commonly, equal weights can be given to returning scenarios or the SP/A (security, potential, and aspiration) theory of [9] can be applied wherein hopeful (fearful) investors will give more weights to favorable (unfavorable) scenarios. There are also other weighting schemes like prospect theory [10], cumulative prospect theory [11], etc. Now that we have the return estimates and corresponding probability weights, the chosen portfolio selection model can now be applied. Normally, the 2 most frequent models utilized are the mean-variance (MV) model [12] and safety-first (SF) model [13]. The basic idea in the MV model is that investors are assumed to be rational such that they will always choose the investment with the highest return (lowest risk) among choices with similar risk (return). In the MV model portfolio return is maximized and risk is minimized considering a risk-return trade-off parameter of the investor. SF model on the other hand is based on the logic that individuals will have their own risk level for corresponding mental frames of the investments [14] as such for each investment investor will have a different loss threshold and the loss probability threshold depending on the source of the invested capital. Thus, in the SF model portfolio return is maximized while limiting the probability of losing a certain percentage of the investment to a specified probability threshold. Thus, this research provides a simple variation of 3 parts of portfolio selection.

In today’s investment world, an exchange-traded fund (ETF) is a relatively new investment instrument that aims to track a specific index. It is similar to a mutual fund (MF) in that regard but ETF can be traded multiple times within a trading day like stocks while MF can only be traded once a day. Moreover, the transaction costs for ETF is relatively cheaper than stocks and MFs.
From 2008-2018, there is a steady increase in cash invested in ETFs. Looking at the world economic outlook, Asia is projected to overtake both US and UK in the year 2022 in terms of GDP. The probable reason according to [15] is a combination of political, regional & global integration, management of renewable sources, and other economic factors. Thus, [8] considered Asian ETFs as the investment pool. They collected market indices of 9 Asian countries and identified the benchmark index and the investment pool. China was considered as the benchmark because it has the largest market share while Malaysia, Singapore, Hong Kong, Indonesia, Korea, Japan, India, and Taiwan are considered into the investment pool. They used equally weighted historical returns as return scenarios and applied a safety-first portfolio selection model to get the optimal portfolios. Then, they compare portfolio performance to the benchmark. They show that their framework can outperform the benchmark and can possibly be a profitable alternative investment option for investors. Thus, this study will expound on their work by focusing more on the return estimation part.

In the work of [8], they consider only the historical returns, but the better estimation of returns may lead to a superior portfolio, so in this research, the work of [8] will be replicated but with the addition of the estimating and generating future return scenarios of the investment pool utilizing regression analysis. The assignment of weights will remain the same by using equiprobable return scenarios. As for the portfolio selection models, the generic safety-first portfolio selection model will be applied. The resulting portfolios will then be compared to the benchmark and also to the corresponding portfolios in [8].

This study flow is as follows. Section 1 provides the basic background of the study. Section 2 discusses the methodologies considered in the proposed portfolio selection framework. Section 3 shows the results and analysis of portfolio performance. Section 4 states the concluding remarks.

2. Methodology
This section discusses the regression equations used to generate the return scenarios, the portfolio selection model, and how the portfolios are evaluated.

2.1. Regression Analysis
As mentioned, similar to [7] and [8], the investment pool considered are 8 Asian ETFs namely Malaysia, Singapore, Hong Kong, Indonesia, Korea, Japan, India, and Taiwan. The benchmark is China ETF. The market index of these countries is collected and assumed that there are corresponding ETFs that perfectly tracts their performance. Thus, in estimating the returns of these ETFs, similar to [5] for each trading day regression analysis is applied to generate return scenarios. Let \( R_{M,t} \) and \( R_{i,t} \) respectively be the returns of the benchmark and ETF \( i \) at time \( t \). Then through a single index model, we can calculate for the return of ETF \( i \) on the next period by the regression equation:

\[
R_{i,t+1} = \beta_0 + \beta_1 R_{M,t+1} + \sigma_{ei} \tag{1}
\]

Note that \( \sigma_{ei} \) is the error term for ETF \( i \). It is also ensured that residuals are normally distributed before using the corresponding regression equation to generate 500 return scenarios to be used in each trading day. These scenarios are assumed to have equal likelihood of occurrence. Most importantly, the scenarios are anchored based on 1 estimated return of the benchmark. This only mean that for every scenario there is a corresponding return estimate for the benchmark.

2.2. Portfolio Selection Model
To determine the optimal portfolios, the generic safety-first portfolio selection model is applied to the generated and equally weighted return scenarios of the 8 Asian ETFs. The resulting SF portfolios will then be compared to the benchmark (China Market Index) and the corresponding portfolios in [8]. SF model focuses on limiting the probability to lose a certain amount (depending on the investor) to a
certain probability (also depending on the investor). Thus, assuming that there are \( n \) stocks and \( m \) scenarios, let portfolio \( Y \) be denoted as \( Y = (y_1, y_2, \ldots, y_n) \), and let \( r_Y \) be the return of portfolio \( Y \) such that \( \sum_{i=1}^{m} y_i = 1 \). Let \( R_L \) and \( \gamma \) represent the loss threshold and loss probability threshold respectively, the generic SF model is represented as:

\[
\max E[r_Y] \quad (2)
\]

\[
s.t. \quad P(r_Y \leq R_L) \leq \gamma, \quad (3)
\]

Equation (2) maximizes the expected return of portfolio \( Y \). \( P(r_Y \leq R_L) \leq \gamma \) in Equation (3) is the portfolio’s downside risk which will be limited to at most the value of \( \gamma \).

Consider scenario \( j \) as a row vector of returns such that \( (r_{1j}, r_{2j}, \ldots, r_{nj}) \). \( R_d \) is the return of ETF \( i \) on scenario \( j \), \( P_j \) as the nominal probability weight on scenario \( j \). Then, for scenario \( j \), denote \( r_{Yj} \) as the return of portfolio \( Y \) on scenario \( j \) such that \( r_{Yj} = \sum_{i=1}^{m} y_i R_{ij} \). In addition let \( I \) denote the threshold portfolio weight that can be allocated to ETF \( i \), then the scenario-based SF model is formulated as:

\[
\max E[r_{Yj}] = \sum_{i=1}^{m} r_{ij}P_j \quad (4)
\]

\[
s.t. \quad r_{ij} = \sum_{i=1}^{m} y_i R_{ij}, j = 1,2,\ldots,m \quad (5)
\]

\[
R_L - r_{ij} \leq M \omega_j, j = 1,2,\ldots,m \quad (6)
\]

\[
\sum_{j=1}^{m} P_j \omega_j \leq \gamma \quad (7)
\]

\[
y_i \leq 1 \quad (8)
\]

\[
0 \leq y_i \leq 1, \quad \omega_j \text{ is binary} \quad (9)
\]

In summary, equation (4) reflects equation (2) as it maximizes the expected return of portfolio \( Y \). Equation (5) shows how portfolio return is calculated at scenario \( j \). Equations (6) & (6) reflect equation (3) as it ensures that the probability of losing \( R_L \) is limited to at most \( \gamma \). Equation (8) makes sure that the SF portfolio only \( I \) of the total budget where \( I = 0.2, 0.4, 0.6, 0.8, 1 \).

### 2.3. Performance Evaluation

Following the comparison format of [1, 2, 3, 4, 5, 6, 7, 8] on evaluating portfolio performance, this study examines the return statistics like mean return, standard deviation, and cumulative returns. For a more rigorous testing pair-return difference t-tests are also applied. The null and alternative hypothesis are:

\[
H_0 \text{ - paired portfolio returns} = 0 \quad (10)
\]

\[
H_1 \text{ - paired portfolio returns} > 0 \quad (11)
\]

### 3. Back-test Results and Analysis

This section contains the data description and the back-test results.

#### 3.1. Data Description

Data are from yahoo finance, where market indices of the 9 Asian countries are collected. These indices are then used to calculate for the corresponding ETFs return. The period of data collected is from 2015-September-1 to 2018-August-20 wherein the back-test period is from 2017-November-11 to 2018-August-20. In total 11 portfolios, 5 SF portfolios or \( Pi \) portfolios where \( Pi = \{P1, P2, P3, P4, P5\} \), 5 counterpart SF portfolios from [8] or SF - 1 portfolios where \( SF - I = \{SF - 0.2, SF - 0.4, SF - 0.6, SF - 0.8, SF - 1\} \), and 1 benchmark (China ETF) or \( M \). Note that the counterpart portfolios are \( P1 \& SF - 0.2, P2 \& SF - 0.4, P3 \& SF - 0.6, P4 \& SF - 0.6, \) and \( P5 \& SF - 1 \). As \( i \) or \( I \) increases the portfolio becomes less diversified. \( Pi \) portfolios used 500 generated return scenarios and \( SF - 1 \) portfolios used 400 historical returns as the return estimate for the next period estimates for each trading day. The safety-first parameters considered are \( RL = -5\% \) and \( \gamma = 5\% \).

#### 3.2. Back-test Results

Tables 1 and 2 show the respective performance of \( Pi \) and \( SF - 1 \) portfolio. Let’s first compare the
returns of $Pi$ with the benchmark M. Looking at the mean returns ($\bar{R}_Pi$), it is noticeable that $Pi$ portfolios have superior mean returns than M. Moreover, the corresponding risks ($\sigma_{RP_i}$) are also smaller than M. These also translate with the cumulative return wherein $Pi$ portfolios have higher $CR_{Pi}$ and $\bar{CR}_{P_i}$ than M. This is also clearly evident in Figure 1. Moreover, through pair-return difference with M, t-tests show that the returns of $P2$ and $P4$ are significantly larger than M with P-values of 0.1000 and 0.0900 respectively. In addition, $P1$, $P5$, and especially $P3$ are very close to being significant. With these observations, it can be said that $Pi$ portfolios can outperform the benchmark.

**Table 1. Return statistics of $Pi$ portfolios over 100 days test period.**

|       | P1   | P2   | P3   | P4   | P5   | M    |
|-------|------|------|------|------|------|------|
| $\bar{R}_{Pi}$ | -0.0003 | 0.0001 | 0.0002 | 0.0004 | 0.0005 | -0.0018 |
| $\sigma_{RP_i}$ | 0.0053 | 0.0069 | 0.0078 | 0.0089 | 0.0101 | 0.0149 |
| $CR_{Pi}$ | -0.0296 | 0.0052 | 0.0180 | 0.0380 | 0.0464 | -0.1766 |
| $\bar{CR}_{P_i}$ | -0.0051 | 0.0083 | 0.0271 | 0.0351 | 0.0357 | -0.0866 |
| P-value with M | 0.1480 | 0.1000* | 0.1010 | 0.0900* | 0.1500 |
| P-value with SF-I | 0.9480 | 0.5770 | 0.5980 | 0.6820 | 0.7370 |

$\bar{R}_{Pi}$ denotes the mean return of portfolio $Pi$
$\sigma_{RP_i}$ denotes the standard deviation of the mean return of portfolio $Pi$
$CR_{Pi}$ denotes the cumulative return of portfolio $Pi$
$\bar{CR}_{P_i}$ denotes the mean cumulative return of portfolio $Pi$
P-value denotes the P-value of the pair-return difference T-test with M
* denotes significance at 0.1 level

Now, let’s compare the $Pi$ portfolios with their $SF-I$ counterpart. From the return statistics of respective portfolios from Tables 1 and 2, insufficient evidence is available to claim that the $Pi$ portfolios are superior than the $SF-I$ portfolios. This is even supported by the resulting P-values of the pair-return t-test between $Pi$ and $SF-I$ portfolios.

**Table 2. Return statistics of SF-I portfolios over 100 days test period.**

|       | SF-0.2 | SF-0.4 | SF-0.6 | SF-0.8 | SF-I | M    |
|-------|--------|--------|--------|--------|------|------|
| $R_{SF-I}$ | 0.0002 | 0.0000 | 0.0000 | 0.0004 | 0.0008 | -0.0018 |
| $\sigma_{RSF-I}$ | 0.0052 | 0.0066 | 0.0077 | 0.0087 | 0.0107 | 0.0149 |
| $CR_{SF-I}$ | 0.0227 | -0.0055 | 0.0020 | 0.0389 | 0.0745 | -0.1766 |
| $\bar{CR}_{SF-I}$ | 0.0312 | 0.0250 | 0.0261 | 0.0515 | 0.0755 | -0.0866 |
| P-value with M | 0.0880* | 0.1310 | 0.1360 | 0.1000* | 0.0800* |

$R_{SF-I}$ denotes the mean return of portfolio $SF-I$
$\sigma_{RSF-I}$ denotes the standard deviation of the mean return of portfolio $SF-I$
$CR_{SF-I}$ denotes the cumulative return of portfolio $SF-I$
$\bar{CR}_{SF-I}$ denotes the mean cumulative return of portfolio $SF-I$
P-value denotes the P-value of the pair-return difference T-test with M
* denotes significance at 0.1 level
Figure 1. Cumulative returns of $P_i$ portfolios.
Nonetheless, since both $P_i$ and $S_F - I$ portfolios can both beat the benchmark $M$. Then $P_i$ portfolios are still a good investment option. This proposed portfolio selection framework can still be a profitable investment for investors.

4. Conclusion
This research provides a variant of the basic framework of portfolio selection. The proposed framework utilizes Asian ETFs as the investment pool and then applies regression analysis to estimate and generate future return scenarios. On each trading day, the return scenarios are assumed to be equally likely which are then used in the safety-first portfolio selection model to obtain optimal portfolios. The resulting $P_i$ portfolios are compared to the benchmark ($M$) and to corresponding $S_F - I$ portfolios $S_F - I$ that uses historical return as return scenarios. Back-tests reveal that both $P_i$ and $S_F - I$ portfolios can outperform the benchmark $M$, but there is insufficient evidence to claim that $P_i$ portfolios are better than $S_F - I$ portfolios. Although this is the case, since $P_i$ can outperform the benchmark, it can still be a profitable investment procedure for some investors.

This study focuses on finding superior methods in the return estimation part of the basic framework of portfolio selection. Although back-test results that the scenario generator technique employed does not show significant superiority over using historical data the research is very optimistic that more sophisticated methods can provide superior portfolios. Thus, this can be the future studies for this research. Other extensions can focus more on the assignment of probability weights to consider the fear and hope levels of the investors. Others can also develop unique models for the selection model. Lastly, finding the best possible combination of techniques for the 3 parts (return estimation, assignment of weights, selection model) of portfolio selection can be the ultimate goal of this research topic.

References
[1] Chang KH, Young MN 2019 Behavioral stock portfolio optimization considering holding periods of B-stocks with short-selling. Comput Oper Res. 1 p 112:104773.
[2] Chang KH, Young MN 2019 Portfolios Optimizations of Behavioral Stocks with Perception Probability Weightings. Ann Econ Financ. 20(2).
[3] Chang KH, Young MN, Liu CC, Chung HP 2018 Behavioral Stock Portfolio Optimization through Short-Selling. Int J Model Optim. 8(2) p 125-30.
[4] Chang KH, Young MN, Hildawa MI, Santos IJ, Pan CH 2015 Portfolio selection problem considering behavioral stocks. In Proceedings of the World Congress on Engineering 2015 (Vol. 2, pp. 1-3).
[5] Chang KH, Young MN, Diaz JF 2015 Portfolio Optimization Utilizing the Framework of Behavioral Portfolio Theory. Int J Model Optim. 15(1) p 1-3.
[6] Mercado MJ, Garcia MA, Ilagan JB, Reveche RR, Young MN 2019 An Empirical Analysis of Financial Ratio Trends of Several Companies Listed in the Philippine Stock Exchange. In Conference Proceeding 10th Annual International Conference on Industrial Engineering and Operations Management (IEOM) 2019.
[7] Young MN, Chuahay TJTN 2019 Mean-Variance Portfolio Selection Utilizing Exchange Traded Funds in Asia. 2019 IEEE 6th International Conference on Engineering Technologies and Applied Sciences (ICETAS). IEEE; doi.org/10.1109/icetas48360.2019.9117564
[8] Young MN, Chuahay TT, Diaz JF 2019 Portfolio Selection Utilizing Safety-First Optimization Model on Exchange-Traded Funds in Asia. In Conference Proceeding 10th Annual International Conference on Industrial Engineering and Operations Management (IEOM) 2019.
[9] Lopes LL 1987 Between Hope and Fear The Psychology of Risk. Adv Exp Soc Psychol. Elsevier; p 255–95. doi.org/10.1016/s0065-2601(08)60416-5
[10] Kahneman D, Tversky A. 1979 Prospect Theory: An Analysis of Decision under Risk. *Econometrica*. JSTOR; 47(2) p 263. doi.org/10.2307/1914185

[11] Tversky A, Kahneman D 1992 Advances in prospect theory: Cumulative representation of uncertainty. *J Risk Uncertain*. Springer Science and Business Media LLC; (4) p 297–323. doi.org/10.1007/bf00122574

[12] Markowitz H 1952 PORTFOLIO SELECTION*. *J Finance*. Wiley; 7(1) p 77–91. doi.org/10.1111/j.1540-6261.1952.tb01525.x

[13] Roy AD 1952 Safety First and the Holding of Assets. *Econometrica*. JSTOR; 20(3) p 431. doi.org/10.2307/1907413

[14] Shefrin H 2000 Statman M. Behavioral Portfolio Theory. *J. Financial Quant. Anal*. JSTOR; 35(2) p 127. doi.org/10.2307/2676187

[15] Kim Y 2018 The Southeast Asian Economic Miracle. Routledge.