Stochastic Dominance and Index Fund Investment Decision

Kai Shi*
School of Economics
Northeast Normal University
Changchun, China

Tianfeng Qi
School of Economics
Northeast Normal University
Changchun, China

Abstract—**The concept of stochastic dominance provides a more general evaluation criterion for expected utility under uncertainty. This paper discusses the investment choice of index funds in light of the concept of stochastic dominance. It is found that small and medium-sized board index funds provide greater expected utility for greedy investors. As for risk-averse investors, small and medium-sized board index funds are less risky and therefore appear to be better choice. In general, stochastic dominance has great application potential in the areas of evaluation, selection, R&D and marketing for funds.**

Keywords—index fund; stochastic dominance; investment tool selection

I. INTRODUCTION

Low-cost index funds provide investors the most profitable tool in the past 35 years. As a passive investment strategy, the index fund obtains the yield and trend consistent with the benchmark index by establishing a portfolio that tracks the performance of the benchmark index. Since the first index fund in China, the Huaan Shanghai 180 Index Enhanced Securities Investment Fund issued at the end of 2002, the index fund market had flourished. As of November 2011, more than a hundred and ten index funds have been issued, with a total market value of 311.094 billion yuan. The issuance of index funds is increasingly favored by the market. In the current environment of weakly efficient Chinese stock market and stable economic development, long-term investment in index funds, either for institutional investors or small and medium-sized retail investors, is a good choice to share the bonus of Chinese economy growth.

At present, there are a wide range of domestic index fund products. Although portfolio investment can spread non-systematic risks, due to different benchmark indexes, there are bound to be differences in relative risks and returns among index funds. Zhao and Ba (2009) [1] believe that Chinese index fund can closely follow the benchmark index, so the determination of the benchmark index is very important. With the increasingly diversified index fund products and increasingly fierce market competition, the realistic difficulty

for investors is how to identify the best fund products for investment between the homogeneous and differentiated products when the purchase and redemption rates are not much different. For fund companies, what is more important is how to choose the basic index and design fund products. These all involve issues of utility comparison and selection of investment tools under uncertain situations.

This paper attempts to answer the question: Which benchmark investment fund should investors choose to maximize their expected utility? More importantly, we provide a new analysis perspective (stochastic dominance) for fund companies' product design, fund rating, and fund management performance evaluation.

II. STOCHASTIC DOMINANCE

Any investor, regardless of which kind of attitude he has towards risks, will strive to maximize the expected utility of wealth, and thus we can use expected utility theory to analyze the behavior choice under uncertainty. If an individual can obtain more wealth from an asset (or portfolio) in any (ordered) natural state, the asset is stochastic dominance relative to other assets. Stochastic dominance theory is a non-parametric approach. The choice of risk assets could be done by analyzing various outcomes and the objective probabilities related under the general preference hypothesis without parameter setting.

First, let us see the first-order stochastic dominance. Assuming the cumulative probability of the asset (or portfolio) x be distributed as $F_x(W)$ and the cumulative probability distribution of the asset (or portfolio) y as $G_y(W)$, for all sets of non-decreasing utility functions, if for all wealth W, $F_x(W) \leq G_y(W)$; for certain wealth W_i, $F_x(W_i) < G_y(W_i)$, then x is first stochastic dominance to y.

In other words, when the x is first stochastic dominance to y, the cumulative probability distribution of asset y is always located on the left of the cumulative probability distribution of asset x. Let $U(W)$ be the utility function, W be the wealth level, $f(W)$ be the probability distribution of wealth, and the expected utility can be express as the probability-weighted sum of utility for all possible wealth levels, $E\big[U(W)\big] \equiv \int_{-\infty}^{\infty} U(W)f(W)dW$. It means that no matter for risk appetite, risk neutral or risk averse investors, x will be preferred to y. The expected utility provided by asset x is greater than that of y.

Furthermore, the second-order stochastic dominance could be introduced. Assume that the marginal utility of wealth in the utility function is positive, while the total utilities are increasing at a decreasing rate (*i.e.*, the marginal utility function is decreasing, and the utility function is strictly concave). At this time, the individual is regarded as a risk-averse person. If for all wealth W, $\int_{-\infty}^{W_i} \left[G_y(W) - F_x(W) \right] dW \geq 0$; for certain wealth W_i, $G_y(W_i) \neq F_x(W_i)$, x is second-order stochastic dominance to y.

For all risk-averse investors, x is better than y means that at any given level of wealth, the coverage of the cumulative probability distribution of y is greater than the coverage of x. It shows that although x and y provide the same expected wealth, the risk of y is even greater.

Stochastic dominance is an extremely important and powerful conclusion which is based on the maximization of expected utility. More importantly, it is suitable for all of probability distribution. It can even be affirmed that if an asset is second-order stochastic dominant, any risk-averse investor will favor the asset regardless of the shape of its utility function.

III. COMMONLY USED INDEXES AND ASSUMPTIONS

Except for the active management behavior of enhanced index funds and various theme investments, copying and tracking the market-represented basic index is the basic operating model of index-based funds based on passive management strategies. Therefore, the choice of the basic index has become the key to determining the risks and benefits of index funds. At present, the basic indexes used in China mainly include:

CSI 300 Index is a component stock index compiled from a sample of 300 A-shares in the Shanghai and Shenzhen stock markets. It covers a market value of around 60% of the Shanghai-Shenzhen Stock Market and a total net profit of over 80%. It has a good market representative. Also it is the first time to reflect the overall trend of the A-share market, which is issued by the Shanghai and Shenzhen Stock Exchanges. As of November 2011, the market value of funds for the purpose of tracking the CSI 300 Index is up to 108.38 billion yuan, accounting for 34.8% of the market value of all index funds.

The Shanghai Composite Index (abbreviated as the Shanghai 180 Index) is the Shanghai Stock Exchange's adjustment and rebranding of the original Shanghai 30 Index. Its sample stocks are large-scale, highly liquid, and the most market representative drawn from all Shanghai A-shares. The 180 kinds of sample stocks reflect the general and operational conditions of the Shanghai securities market. They are operable and investment-capable and can serve as a benchmark for investment evaluation and the basis for financial derivatives.

The Shenzhen 100 Index is the first index in the domestic securities market that is oriented to investment functions and represents a multi-tiered market system. It includes the Shenzhen A-shares market (including the small and medium board) with the largest market value and the most actively traded 100 constituent stocks. It also represents the core high-quality asset of the Shenzhen A-share market, who maintains strong growth, low valuation, and high investment value.

The CSI 100 Index consists of the largest 100 stocks selected from the sample stocks of the CSI 300 Index to comprehensively reflect the overall situation of a group of large cap companies with the most market influence in the Shanghai and Shenzhen stock markets.

The Small and Medium Board Index is a comprehensive index of stocks that are normally traded on the SME board. As the index of the first all-tradable sector, it more comprehensively and objectively reflects the overall price trend of the SME board under all-trading conditions. It is an important reference for investment operations, and becomes a wind vane for the market trend in the era of post-share distribution.

In addition, the Shanghai Composite Index and the Shenzhen Component Index are also often used as the basis of the ETF, which fully reflects the overall situation of the two markets.

Due to the availability of relevant data and statistical data, this paper mainly seeks the dominant index within the scope of the above discussion. However, the research method of this article is not limited to this, it can easily be extended to a wider range of asset selection issues.

The basic interesting assumptions in this paper: within the candidate set, the index i maximizes the expectation and utility of the investor, so a rational investor should choose an i-based index fund.

Based on this, build the following assumptions:

H_0^1 : All index returns have basically the same distribution.

H_0^2 : Index i stochastic dominance to all other indexes.

H_0^3 : At least one of the other indexes stochastic dominants the index i.

Obviously, H_0^1 shows that all indexes are homogenous to investors. Therefore, the rejection of it is the premise to find stochastic dominance. It is worth noting that not rejecting H_0^2 does not mean that i provides investors with the greatest expectation and utility. Rejecting H_0^3 indicates that index i at least provides investors with the same returns as other indexes.

In summary, the index, only satisfies the conditions of rejecting H_0^1, not rejecting H_0^2 and rejecting H_0^3, can maximize the expected utility of investors.

IV. METHODOLOGY

We use $X_1 \sim X_7$ to respectively represent the daily rate of return of CSI 300 Index, Shanghai 180 Index, Shenzhen 100 Index, CSI 100 Index, Small and Medium Board Index, Shanghai Composite Index and Shenzhen Component Index. We also use $F_1(x) \sim F_7(x)$ to represent the corresponding accumulation distribution function. Referring to Levy's (2006) study [2], $D_k^{(1)}(x) = F_k(x)$, k = 1,..., 7, for s ≥ 2, iteratively define

$$D_k^{(s)}\left(x\right) = \int_{-\infty}^{x} D_k^{(s-1)}\left(t\right)dt \tag{1}$$

If for all x, $D_i^{(s)}(x) \leq D_j^{(s)}(x)$; for some x, the inequality is established strictly, then the X_i s order is stochastic dominating X_j.

We use χ to denote the support set of X_k, s to represent the order of stochastic dominance. For any $k, l, i = 1, \ldots, 7$ and $x \in \chi$, we define:

$$\Delta_{k,l}^{(s)}(x) = D_k^{(s)}(x) - D_l^{(s)}(x) \tag{2}$$

$$d_{1s} = \max_{k \neq i} \sup_{x \in \chi} \left| \Delta_{i,k}^{(s)}(x) \right| \tag{3}$$

$$d_{2s} = \max_{k \neq i} \sup_{x \in \chi} \Delta_{i,k}^{(s)}(x) \tag{4}$$

$$d_{3s} = \min_{k \neq i} \sup_{x \in \chi} \Delta_{k,i}^{(s)}(x) \tag{5}$$

For $j = 1, 2, 3$, the null hypothesis and alternative hypothesis for $H_0^1 \sim H_0^3$ above can be expressed as

$H_0^j : d_{js} \leq 0$ vs. $H_1^j : d_{js} > 0$

In order to achieve the inspection on the basis of the data set $\{X_{kt}: t = 1, \ldots, N; k = 1, \ldots, 7\}$, taking H_0^2 as an example, refer to Cho *et al.* （2007）[3], Linton *et al.* (2005) [4] and Bai *et al.* （2011）[5] to construct test statistics:

$$D2_N^{(s)} = \max_{k \neq i} \sup_{x \in \chi} \sqrt{N} \left[\overline{D}_N^{(s)}(x; \overline{F}_i) - \overline{D}_N^{(s)}(x; \overline{F}_k) \right] \tag{6}$$

Among them

$$\overline{D}_N^{(s)}(x; \overline{F}_k) = \frac{1}{N(s-1)!} \sum_{i=1}^{N} 1(X_{ki} \leq x)(x - X_{ki})^{s-1} \tag{7}$$

The rest of the test statistics can be constructed in a similar way.

The rejection of each null hypothesis is based on the larger positive test statistic. Under the appropriate constraints of Linton *et al.* (2005) (Hypothesis 1 to Hypothesis 3) [4], the test statistic converges weakly to the function of the Gaussian process. Since the limit distribution depends on the unknown true distribution of X_k, the asymptotic critical value cannot be constructed once and for all. Only a bootstrap and subsampling method can be used to estimate the asymptotic p-value. Referring to Horowitz （2000）[6] and Lahiri （1999）[7], the basic process is as follows:

A. Subsampling method

We use W_N to represent any test statistic

Step 1: Use the full sample $W_N = \{Z_i = (X_{1i}, \ldots, X_{7i})^T : i = 1, \ldots, N\}$ to calculate the test statistic W_N;

Step 2: Generate a subsample $W_{N,b,i} = \{Z_i, \ldots, Z_{i+b-1}\}$, where $i = 1, \ldots, N-b+1$;

Step 3: Calculate test statistics $W_{N,b,I}$ by using subsamples $W_{N,b,i}$;

Step 4: Calculate asymptotic p –value

$$p_{s,b} = \frac{1}{N-b+1} \sum_{i=1}^{N-b+1} 1\left(W_{N,b,i} > W_N\right) \tag{8}$$

The choice of subsample size depends on the length of the data and should satisfy: when $N \to \infty$, $b \to \infty$ and $b/N \to 0$.

B. (re-centralized overlap) Bootstrap method

Step 1: Same as above;
Step 2: Same as above;

Step 3: the bootstrap sample $W_N^* = \{Z_i^* : i = 1, \ldots, N\}$ is generated by extracting N-b+1 overlapping sample blocks and arranging them end to end in sequence. This process is repeated M times, where M is the bootstrap sample times.

Step 4: Use the bootstrap sample W_N^* to calculate the re-centralized test statistic W_N^*.

For example, for $D2_N^{(s)}$, define:

$$W_N^* = D2_N^{(s)*} = \max_{k \neq i} \sup_{x \in \chi} \sqrt{N} \left[\overline{D}_N^{(s)*}(x; \overline{F}_i) - \overline{D}_N^{(s)*}(x; \overline{F}_k) \right]$$

Among them,

$$\overline{D}_N^{(s)*}(x; \overline{F}_k) = \frac{1}{N(s-1)!} \sum_{i=1}^{N} \left\{ 1(X_{ki}^* \leq x)(x - X_{ki}^*)^{s-1} - \omega(i,b,N) 1(X_{ki} \leq x)(x - X_{ki})^{s-1} \right\}$$

$$\omega(i,b,N) = \begin{cases} i/b & i \in [1, b-1] \\ 1 & i \in [b, N-b+1] \\ (N-i+1)/b & i \in [N-b+2, N] \end{cases}$$

Repeat M times of this process;

Step 5: By calculating the ratio of W_N^* over W_N in M repetitions, asymptotic p-values are approximated. Similar to the subsampling method, the subblock size b should satisfy: when $N \to \infty$, $b \to \infty$ and $b/N \to 0$.

V. EMPIRICAL ANALYSIS

A. Data

In the empirical analysis, daily return rates of CSI 300 Index, Shanghai 180 Index, Shenzhen 100 Index, CSI 100 Index, Small and Medium Board Index, Shanghai Composite Index and Shenzhen Component Index as of March 31, 2018 are used. All data come from CSMAR Financial (stock) database. In order to ensure the comparability of the samples, we choose the shortest sample length as the basis. The remaining index returns were truncated, eliminating the missing data and non-comparable period. Finally we obtain 7 sets of yields.

B. Results

When the sample size is small, whether calculating the p-value by using the bootstrap method or the subsampling method does not pose a problem. However, when the sample size is large, the former has a heavier computational burden and a longer computation time. Linton *et al.*(2005) [4] proved that under very weak conditions, the subsampling method provides a fairly consistent approximation of the critical value. Therefore, this article will mainly focus on the results of subsampling. Determining the size b of the subsample is very important but quite difficult. Linton *et al.* proposed various criteria for selecting b, including the thumb of rule used in this paper. Referring to the practice of Cho *et al.* (2007) [3], this paper calculates test statistics for 30 different subsample size b within the range of $[N^{0.3}, N^{0.7}]$ and reports the median of corresponding p values. The relevant hypothesis test results are shown in Table I.

Before analyzing the results of the stochastic dominance test, reviewing the interesting fundamental assumptions would be helpful: In order to achieve the expectation and utility maximization, greedy investors prefer to select an index fund that satisfies the conditions of rejecting H_0^1 and H_0^3 and not

rejecting H_0^2 under the first-order stochastic dominant criteria.; a risk-averse investor (regardless of the shape of its utility function) prefers to select an index fund that satisfies the conditions of rejecting H_0^1 and H_0^3 and not rejecting H_0^2 under the second-order stochastic criteria.

TABLE I.   STOCHASTIC DOMINANT TEST OF MEDIAN P VALUES

	CSI 300	SH 180	SZ 100	CSI 100	SME	SHCI	SZCI
H_0^1:							
FSD	0.0085***	0.0000***	0.0000***	0.0162**	0.0000***	0.0000***	0.0000***
SSD	0.0551**	0.0423**	0.0426**	0.1899	0.0435**	0.0653*	0.0363**
H_0^2:							
FSD	0.0082***	0.0000***	0.0000***	0.0364**	0.7973	0.0000***	0.0110**
SSD	0.0499**	0.0381**	0.0378**	0.1067	0.4257	0.0442**	0.0245**
H_0^3:							
FSD	0.1530	0.5229	0.2832	0.7684	0.0119**	0.0000***	0.3849
SSD	0.0200**	0.3665	0.0576*	0.6366	0.0291**	0.0000***	0.5202

Notes: FSD represents a first-order random dominance, SSD represents a second-order random dominance, *** represents a significant 1% level, ** represents a significant level at 5%, and * represents a significant level at the 10% level.

According to the empirical results, it is not difficult to find: SME index fund is the first choice for greedy and risk-averse investors.

VI. SUMMARY

This paper does not only answer the question that what kind of index fund should domestic investors choose but more importantly provides investors a new investment decision method under uncertainty. In comparison with Cheng *et al.* (2003) [8], the stochastic dominance method is particularly applicable to the fund company's product development and marketing planning (the examination based on the relevant theme or industry investment index can effectively reduce the investment risk, and relevant inspection conclusions can be used as a powerful tool to persuade investors before the launch of new products). Moreover, the stochastic dominance method also provides new insights into the rating and classification of index funds. In addition, the frontier efficiency evaluation based on stochastic dominance provides an important reference for fund performance evaluation.

REFERENCES

[1] Y. Zhao and S.S Ba, "Risk Assessment and Prediction of Indexed Investment: An Empirical Study Based on Chinese Index Funds," Systems Engineering-Theory & Practice, 2009, issue 12, pp.80-87.

[2] Levy, H., Stochastic Dominance: Investment Decision Making under Uncertainty, Berlin: Springer, 2006.

[3] Young-Hyun Cho, Oliver Linton, and Yoon-Jae Whang, "Are there Monday effects in stock returns: A stochastic dominance approach," Journal of Empirical Finance, 2007, vol. 14, pp. 736-755.

[4] Linton, O., E. Maasoumi and Y.-J. Whang, "Consistent Testing for Stochastic Dominance under General Sampling Schemes, Review of Economic Studies," 2005, vol. 72: 735-765.

[5] Z. Bai, H. Li, H. Liu and Wing-Keung Wong, "Test Statistics for Prospect and Markowitz Stochastic Dominances with Applications," The Econometrics Journal, 2011, vol. 14, pp. 278-303.

[6] Horowitz, J.L., The Handbook of Econometrics. Amsterdam: North Holland, 2000.

[7] Lahiri, S.N. "Theoretical Comparison of block bootstrap methods," The Annals of Statistics, 1999, vol. 27, pp. 386-404.

[8] B. Cheng, X.J. Zhang and T.J. Jiang, "An Index Construction Method of Weighted Equity Conversion under Uniform Scale and Its Application," Systems Engineering-Theory & Practice, 2003, issue 10, pp. 36-42.