Hybrid Factor Regression Approach to Identify Trading Styles of Thai Equity Funds and Their Attribution

Chuchai Srisansanee
Graduate School of Management and Innovation, Fl 8 Bldg CB 5
King Mongkut's University of Technology Thonburi, Thung Khru, Bangkok, 10140, Thailand.
Tel : (662) 470-9787, Fax : (662) 470-9401, Email : chuchai_srisan@yahoo.com

Abstract

Hybrid factor regression (HFR) is a hybrid multi-factor approach that can integrate both time-series and cross-sectional data to identify trading styles and calculate attribution of equity funds. In the Thai market, funds can be easily analyzed. It is found that funds employ the same trading style and hence a single equation is required to simultaneously compute attributions of all funds. A performance assessment will be more accurate in terms of both risk and return. This information is very important to retail investors, to decide whether they would invest in stocks themselves or simply purchase funds.

Keywords: Factor Models, Factor Analysis, Linear Regression, Attribution, Performance Measurement, Ranking, Trading Style, Equity Funds.

1. Introduction

Prior to the subprime crisis in mid 2008, equity funds have gained popularity as an investing alternative. One of the reasons is that their return is attractive in comparison with bank deposit rates. A common question among retail investors is "which fund should I invest in?" Return on net asset value is a common criterion employed by ranking companies to measure performance of funds [1]. A fund with this large value is claimed to have superior performance. This measure should not, however, be considered alone because it does not reveal overall behavior of funds, e.g., style and attribution. Some styles may be unsuitable for some retail investors because they can be passive or aggressive. Major sources of return can be unfavorable to investors, i.e., spike. To answer these questions, styles and attribution must be identified first, then performance measures, e.g. Sharpe ratio, Treynor ratio, and active alpha, can be fairly compared within the same style. At present, attribution is generally analyzed by multi-factor models while style identification can be analyzed by factor analysis. It will be efficient if we have a single approach that can analyze simultaneously both styles and attribution. This research has designed a special kind of hybrid model, namely, hybrid factor regression (HFR) that can nicely handle both analyses. We will solely focus on style identification and attribution. Ranking equity funds deserves a separate discussion and could be employed in the future.

2. Multi-Factor Model

The multi-factor model is a collection of statistical tools employed to formulate relationships between target returns and important drivers. Important drivers are generally called explanatory variables and target returns are called response variables. The model is useful because it can identify the source and amount of returns. Generally, a linear function is used to link response variables to explanatory variables that have financial meaning because it is easy to build and comprehend; hence, transparent for risk assessment. Among a class of multi-factor models, cross-sectional and time-series approaches are well known to practitioners [2, 3]. The first approach requires cross-sectional data. It ties explanatory variables to returns of several stocks obtained at a particular time t. It completely ignores time information. The second approach requires time-series data. It ties returns of a single stock to explanatory variables obtained at multiple times. Linear regression and factor analysis are widely used in both approaches. They share the same functional forms but have different statistical properties. Their cross-sectional form at time t can be represented as:

$$r_n = \alpha + \sum_{j=1}^{k} \beta_j x_j + \varepsilon \qquad (1)$$

where r_n is a return of investment n (n = 1, 2, …, N), N is a number of stocks, α and β_j are unknown parameters, x_j are explanatory variables, and ε is an error term. Similarly, a time-series form for stock i is:

$$r_t = \alpha + \sum_{j=1}^{k} \beta_j x_j + \varepsilon \qquad (2)$$

where r_t is return of investment t (t = 1, 2, …, T), T is a time horizon, α and β_j are unknown parameters, x_j are explanatory variables, and ε is an error term. In general, explanatory variables can be fundamental factors, e.g., P/E, leverage ratio, and/or market factors, e.g., volatility, momentum, exchange rates, but they also can be any variable. Commercial vendors that use a model of this type are Connor and Herbert [3] and Fama and French [4].

Both approaches decompose the total return into two attribute returns, e.g. α and $\beta_j x_j$. A constant term α is called alpha and usually interpreted as a return due to ability of a fund manager. A sensitivity β_j of an explanatory variable x_j measures a return of fund as percentage of a return of the variable x_j. For example, if x_j is a market return, e.g., SET50, the attribution $\beta_j x_j$ will be a return attributed to the market. This information provides fund managers a view for optimizing their asset allocation.

Either regression or factor analysis has a common drawback. They cannot integrate both cross-sectional and time-series information. Some hybrid approaches have been proposed to combine both types of information. They are based on a panel data analysis [5]. In this article, we will use a modified version of factor analysis that can combine cross-sectional data and time-series data. We call this approach "hybrid factor regression" or HFR for short.

3. Hybrid Factor Regression

HFR is a special type of hybrid model. It takes advantage of both factor analysis and linear regression. Factor analysis is an excellent tool to decompose the total risk into systematic risk and specific risk. Linear regression, however, is excellent for identifying sources of returns. HFR is constructed in two stages. First, factor analysis is applied on time-series data $R^T \in \Re^{T\times N}$ which consists of T days of N stock returns:

$$R^T = LY + E_{spec}^T \qquad (3)$$

where R is standardized stock returns ($\Re^{N\times T}$), L is loading factors ($\Re^{N\times p}$), p is a dimension of risk drivers called latent factors, Y is orthogonal latent factors ($\Re^{p\times T}$), and E^T_{spec} is specific error ($\Re^{N\times T}$). This stage exploits time-series information of stocks and breaks down the total risk of stocks into systematic risk Σ and specific risk Ψ. See eq. (4). This information is critical to measure fair performance on a unit of risk.

$$\begin{aligned}\mathrm{Var}(R^T) &= \mathrm{Var}(LY) + \mathrm{Var}(E^T_{spec}) \\ &= L \cdot \mathrm{Var}(Y) \cdot L^T + \Psi \\ &= LL^T + \Psi \\ &= \Sigma + \Psi\end{aligned} \tag{4}$$

Latent factors are drivers of systematic risk but they have no financial meaning. To find their meaning, they are linked to explanatory variables with a time-series approach. Thus, in a second stage, Y is regressed on explanatory variables X ($\Re^{T\times k}$). After this stage, HFR exploits both cross-sectional and time-series data. Since latent factors are mutually orthogonal, we can separately build p linear regression models. At the end, a model is:

$$y = \alpha + X\beta + \varepsilon_{sys} \tag{5}$$

where $y \in \Re^T$ is a latent factor corresponding to a row of Y, i.e., $Y^T = [y_1\ y_2 \ldots y_p]$, $\alpha \in \Re^T$, $\beta \in \Re^k$ is sensitivity and $\varepsilon_{sys} \in \Re^T$ is systematic error. Substitute eq. (5) into eq. (4), hybrid factor regression can be written as:

$$\begin{aligned}Y^T &= (y_1\ \ y_2\ \ \cdots\ \ y_p) \\ &= (\alpha_1\ \ \alpha_2\ \ \cdots\ \ \alpha_p) + (X\beta_1\ \ X\beta_2\ \ \cdots\ \ X\beta_p) \\ &\quad + (\varepsilon_1^{sys}\ \ \varepsilon_2^{sys}\ \ \cdots\ \ \varepsilon_p^{sys}) \\ &= \tilde{A} + X\tilde{B} + \tilde{E}_{sys}\end{aligned} \tag{6}$$

Multiply eq. (6) by L^T, we obtain

$$\begin{aligned}Y^T L^T &= (\tilde{A} + X\tilde{B} + \tilde{E}_{sys})L^T \\ &= \tilde{A}L^T + X\tilde{B}L^T + \tilde{E}_{sys}L^T \\ &= A + XB + E_{sys}\end{aligned} \tag{7}$$

Substitute eq. (7) into a transpose of eq. (3) yields:

$$\begin{aligned}R &= Y^T L^T + E_{spec} \\ &= A + XB + E_{sys} + E_{spec}\end{aligned} \tag{8}$$

where A is alpha, B is beta of stocks relative to explanatory variables, E_{sys} is systematic risk, and E_{spec} is specific risk. We can see from eq. (9) that HFR preserves the risk decomposition of original factor analysis. This feature is useful to modify standard performance measures.

$$\begin{aligned}\mathrm{Var}(R) &= \mathrm{Var}(A + XB) \\ &\quad + \mathrm{Var}(E_{sys}) + \mathrm{Var}(E_{spec}) \\ &= \mathrm{Var}(E_{sys}) + \mathrm{Var}(E_{spec}) \\ &= \Sigma + \Psi\end{aligned} \tag{9}$$

Figure 1 illustrates a structure of hybrid factor regression.

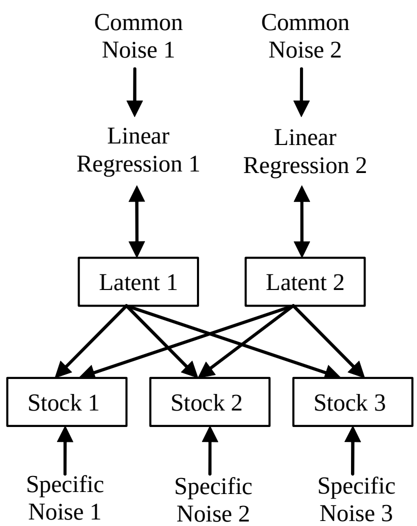

Figure 1 Hybrid factor regression.

In addition to the ability to combine cross-sectional and time-series data, and decompose risk, HFR has two other important features which are: (i) dimensional reduction, and (ii) built-in dependence structure. For instance, if we have 400 stocks, we will have to build 400 regression models using eq. (1) or (2). These models treat 400 stock returns as if they are uncorrelated. Their estimated parameters are biased and predictive power is compromised as a result. Regression models in the hybrid factor regression, on the contrary, require much fewer regression models. The number of models are equal to a number of important latent factors, which are very small in general. In some data sets, we can use a single latent factor as it will be illustrated in the subsequent section.

4. Data

To analyze styles of equity funds, we need two kinds of data sets. The first kind of data is NAV of equity funds Data which is available at www.thaimutualfund.com. The second one is a market benchmark. Data is available in the database of The Thai Bond Market Association, (ThaiBMA). Fund performance is ranked by return and published on the internet by Lipper Ltd. (www.lipperleaders.com). We use top funds with Lipper ranking in this study. As of Dec 28, 2007, there are approximately 105 funds in the Lipper database. We select the top thirty NAV-return equity funds. We believe that rational retail investors would not invest in funds with a lower rank and the number "30" is large enough for analytics to identify hidden patterns. Since the purpose of this paper is purely academic, we hide actual names of funds and label them by numbers 1, 2, ..., and 30 to avoid criticism among their bookkeepers. The choice of market benchmark can be SET, SET100, or SET50 indices. We selected SET50 index in this study with the reason explained in Section 5. The period of all data set is between Jan 3, 2006 and Dec 28, 2007 with a total of 488 days.

We construct five rolling data sets to monitor a shift in styles. They are one-year data sets (quarterly) labeled as DS1, DS2, DS3, DS4, and DS5. Because one fund is relatively new, there are 29 funds in data sets DS1 and DS2. Their information is shown in Table 1.

Table 1 Five data sets

Data Set	Start	End	Sample Size
DS1	03-Jan-06	29-Dec-06	243
DS2	03-Apr-06	30-Mar-07	242
DS3	03-Jul-06	29-Jun-07	245
DS4	02-Oct-06	28-Sep-07	246
DS5	03-Jan-07	28-Dec-07	245

5. Style Identification and Attribution

The paramount achievement of factor analysis is its ability to cluster funds into homogeneous groups unrevealed by simple statistical tools. Funds with large loading values on the same latent factors are considered to have similar characteristics. Consider Figure 2 obtained from factor analysis on the data set DS1. All funds are highly loaded on latent factor 1. Their loading on other latent factors are negligible. It implies that the return of all funds is driven by a single unknown factor. Similar patterns emerge from data sets DS2, DS3, and DS4 as shown in Figures 3-5 respectively. In Figure 6, a pattern appears to be driven by multi-unknown sources. It is possible that the optimization process prematurely stops in data set DS5. A common approach to tackle this problem is to employ a so-called factor rotation technique which transforms loading factors to achieve a simple structure. Several rotations have been experimented with and we found that a quartimax rotation can achieve this task. Quartimax rotation is a matrix T that maximizes a quantity Q [5].

$$\max\ Q = \sum_{n=1}^{N}\sum_{k=1}^{p}(l_{ik}^{*})^{2} \qquad (10)$$

where l_{ik}^{*} is a loading element in $L^{*} = LT$ from fund n and latent k. The result is shown in Figure 7. It is clear that the 1st latent factor contributes most of all fund returns. The implication of this finding is that there is only a single style for all funds and the 1st latent factor is important. Thus, our analysis will be focused solely on that latent factor. This empirical finding makes the 2nd stage of HFR much easier.

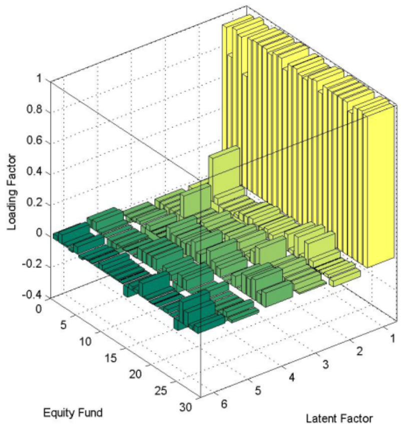

Figure 2 Loading factors obtained from a data set DS1.

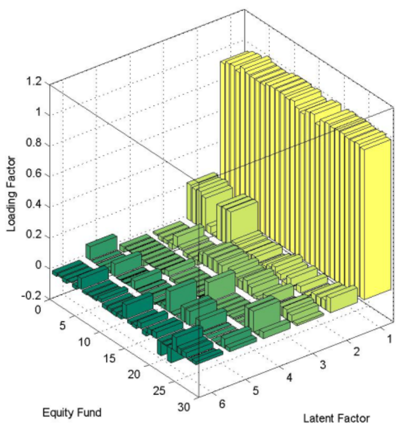

Figure 3 Loading factors obtained from a data set DS2.

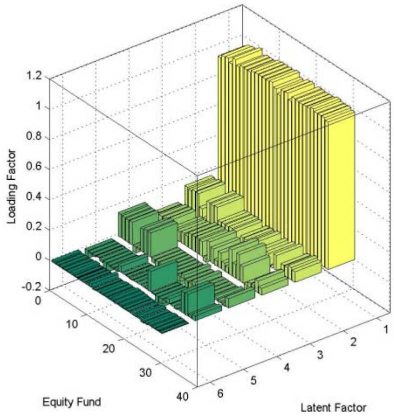

Figure 4 Loading factors obtained from a data set DS3.

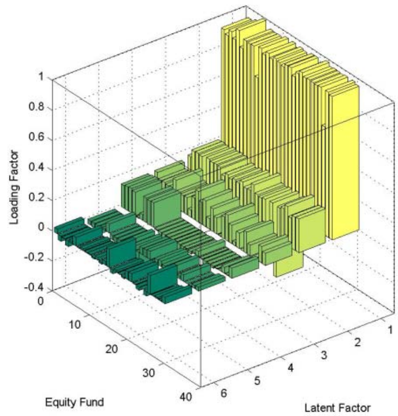

Figure 5 Loading factors obtained from a data set DS4.

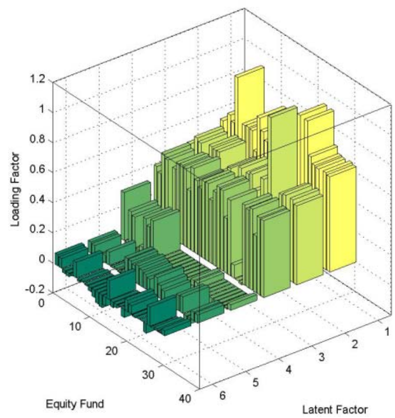

Figure 6 Loading factors obtained from a data set DS5.

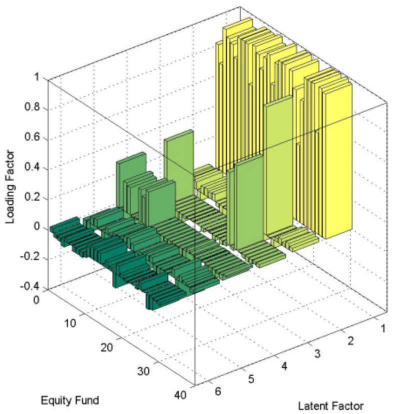

Figure 7 Loading factors obtained from a data set DS5 with quartimax rotation.

We are now ready to analyze styles and attribution of funds. HFR will be executed for each data set. Since fund 25 is found to be high-load on the 2nd latent factor only in the data set DB5, the impact of the 2nd latent factor can be temporary and we decided to keep it for further analysis.

Stage-1 HFR is run to obtain loading factors and the 1st latent factor. The latent factor is regressed on some explanatory variables in stage 2. The theory of CAPM suggests that we should test some stock indices as the first explanatory variable. Empirical findings [6] with Thai indices indicated that SET50 index has similar risk characteristics to the SET index.We choose SET50 because it comprises a smaller subset of Thai stocks; hence, retail investors can achieve a lower cost if they decide to replicate SET50. Results for five data sets are shown in Table 2.

Table 2 Regression obtained from fitting the 1st latent factor to SET50

Data Set	α	β	p-val (α)	p-val (β)	p-val (F-test)	R^2
DS1	0.0181463	56.9032259	0.0120366	0.000000	0.000000	0.9876100
DS2	0.0194231	56.5722084	0.0053087	0.000000	0.000000	0.9885668
DS3	–0.0346014	59.2108626	0.0000019	0.000000	0.000000	0.9877986
DS4	–0.0560655	54.6775332	0.0000000	0.000000	0.000000	0.9880436
DS5	–0.0916359	70.3451247	0.0000000	0.000000	0.000000	0.9768127

Overall, regression nicely fits the 1st latent factor to SET50. Residual analysis also confirms that underlying assumptions of linear regression are acceptably not violated. Figure 8 is a residual plot of data set DS1. Other data sets have similar residual patterns. Therefore, we strongly believe that SET50 is a single driver of returns and it is alone sufficient to predict return of all funds over study periods. Final models are obtained by scaling α and β with loading factors obtained from the 1st latent factor in stage 1. Since standard factor analysis applies on standardized returns, the result must be reverted by eq. (11).

$$\alpha_{F,i} = \alpha_i l_i \sigma_i + \mu_i \tag{11a}$$

$$\beta_{F,i} = \beta_i l_i \sigma_i \tag{11b}$$

where l_i, μ_i, and σ_i are a loading factor, a mean, and a volatility of fund i, respectively.

Alpha and beta for all funds of five data sets are efficiently visualized by a box-whisker plot. Figure 9 shows us that most of beta are in a range of 0.95 and 1 so it is likely that funds were constructed to track SET50. Since SET50 is a single driver of fund returns and a SET50 tracking strategy is employed, we can conjecture that funds are passive. In Figure 10, alpha is mostly positive and usually interpreted that a majority of fund managers have contribution to fund performance. Alpha itself only answers a question "apart from market return, who does perform better?" It will be more informative if we know a portion of returns attributed from fund managers themselves and market return. This process is called attribution [3] and can be simply calculated as:

$$\alpha_i^{attrib} = \frac{\alpha_i}{\mu_i} \tag{12}$$

$$\beta_i^{attrib} = \frac{\beta_i \cdot \mu_{set50}}{\mu_i} \tag{13}$$

where α_i^{attrib}, β_i^{attrib}, and μ_i are alpha attribution, beta attribution, and a return of fund i respectively and μ_{SET50} is a return of SET50. $\alpha_i^{attrib} + \beta_i^{attrib} = 1$.

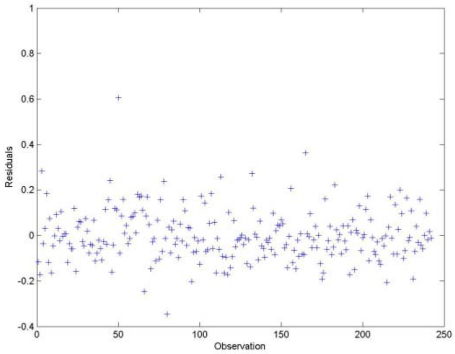

Figure 8 Residual plot of 1st latent factor and SET50 obtained from a data set DS1

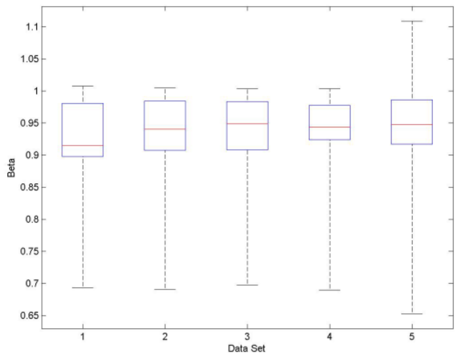

Figure 9 Box-whisker plot of beta

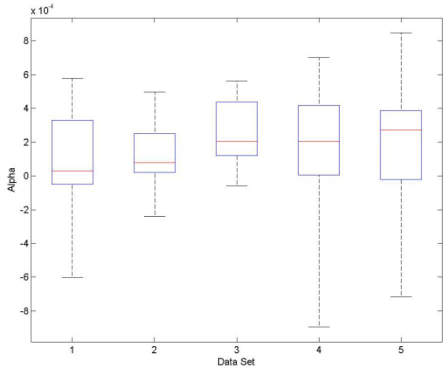

Figure 10 Box-whisker plot of alpha

Attribution is run for each data set and results are shown in Figures 11-15. It is obvious in Figures 13-15 that alpha attribution is relatively small in comparison with beta attribution. It asserts the belief that funds are passive. In Figures 11 and 12, attribution is highly volatile. Seemingly, it was caused by two major events in 2006. They are political turmoil, e.g. demonstration and coup, and BoT's 30% capital reserve requirement. They caused panic among investors and fund managers made a great deal of effort to rebalance their portfolios. The impact fades out in data sets DS3, DS4, and DS5. A sample of numerical alpha and beta attribution are shown in Table 3.

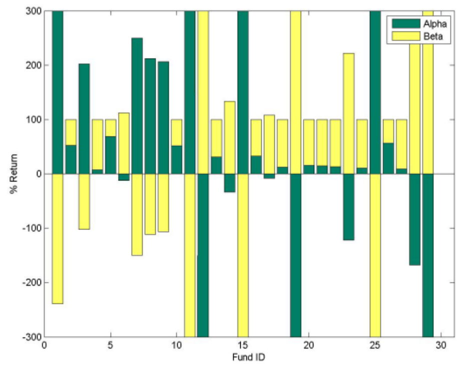

Figure 11 Alpha attribution and beta attribution from data set DS1

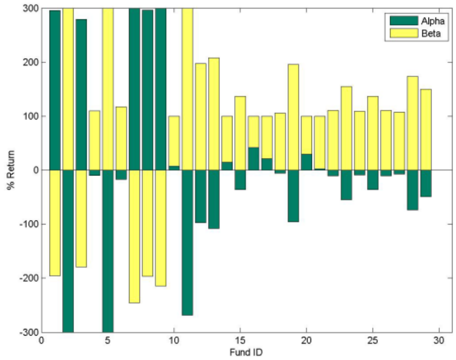

Figure 12 Alpha attribution and beta attribution from data set DS2

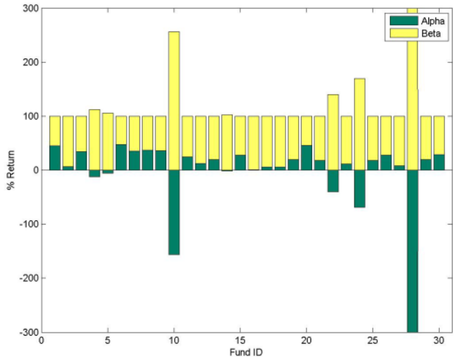

Figure 14 Alpha attribution and beta attribution from data set DS4

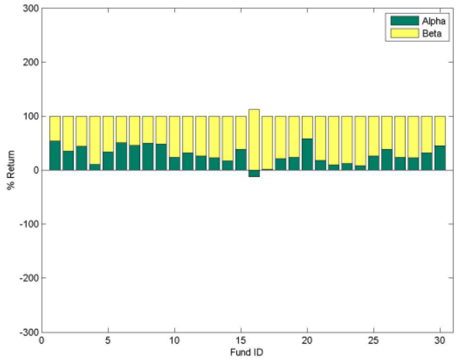

Figure 13 Alpha attribution and beta attribution from data set DS3

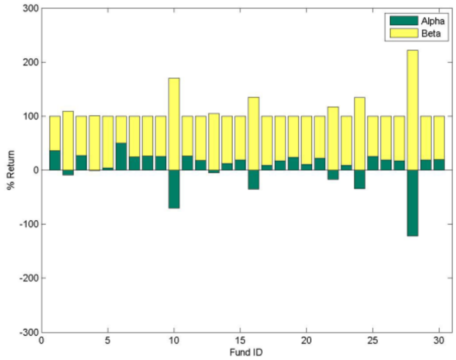

Figure 15 Alpha attribution and beta attribution from data set DS5

Table 3 Numerical alpha and beta attribution for data set DS3

Fund	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
α_i^{attrib}	53.6433	35.1753	44.1828	10.6721	33.6944	50.5320	45.8960	50.0120	47.9790	23.8988
β_i^{attrib}	46.3567	64.8247	55.8172	89.3279	66.3056	49.4680	54.1040	49.9880	52.0210	76.1012
Fund	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
α_i^{attrib}	32.0525	26.3263	23.1768	17.3217	38.5183	–11.9905	1.3779	21.3335	24.0167	58.0030
β_i^{attrib}	67.9475	73.6737	76.8232	82.6783	61.4817	111.9905	98.6221	78.6665	75.9833	41.9970
Fund	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**
α_i^{attrib}	18.0558	9.4334	12.3672	8.3603	26.0212	38.5176	24.0292	23.0558	31.6920	45.2140
β_i^{attrib}	81.9442	90.5666	87.6328	91.6397	73.9788	61.4824	75.9708	76.9442	68.3080	54.7860

The discussion of return attribution only gives a hint on the possibility of passive equity funds. Now we show that the pattern will emerge when risk attribution is analyzed along with return attribution. Risk attribution is a collection of risk components partitioned into systematic risk and specific risk using eq. (9). Results of data sets DS3, DS4, and DS5, are shown in Figures 16-18 respectively.

Monitoring a shift in Figures 16, 17, and 18, we can see that specific risk is very small in Figure 16 while more noticeable in Figures 17 and 18. Dynamic shift in risk agrees with that of returns. The evidence gives us more confidence to conclude that equity funds are passive but it is too fast to make that conclusion.

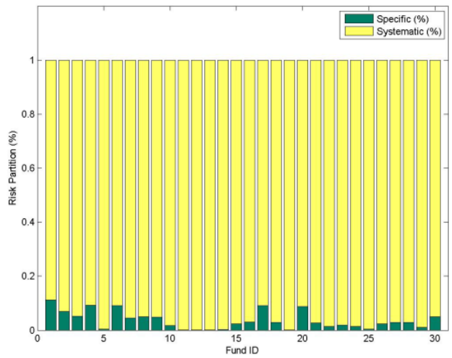

Figure 16 Risk attribution from data set DS3

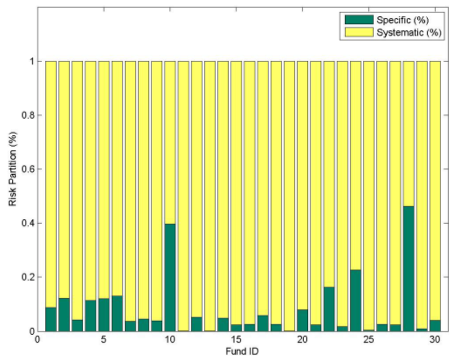

Figure 17 Risk attribution from data set DS4

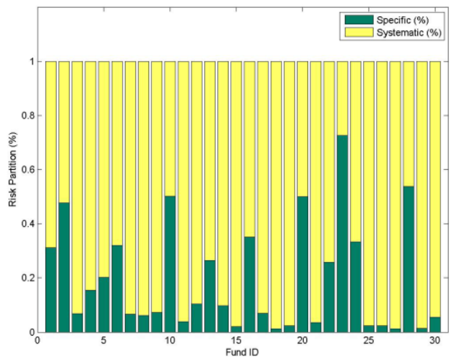

Figure 18 Risk attribution from data set DS5

It turns out that if we stack up a graph of return attribution on that of risk attribution, interesting patterns emerge. Considering stack-up Figure 19, systematic risk dominates the risk side while beta returns attribute a great deal on a return side. Apparently, funds are passive. In Figure 20, some funds have a great deal of specific risk. They include funds 10, 22, 24, and 28. Their corresponding alpha attribution becomes highly negative. In Figure 21, funds 10, 16, 22, 24, and 28 have a great deal of specific risk and also generate negative alpha attribution. Other funds that also have a great deal of specific risk, e.g., fund 20 in Figure 21, cannot generate large alpha attribution to compensate extra specific risk. Empirical evidences so far suggest a single possibility that Thai equity funds are naturally passive.

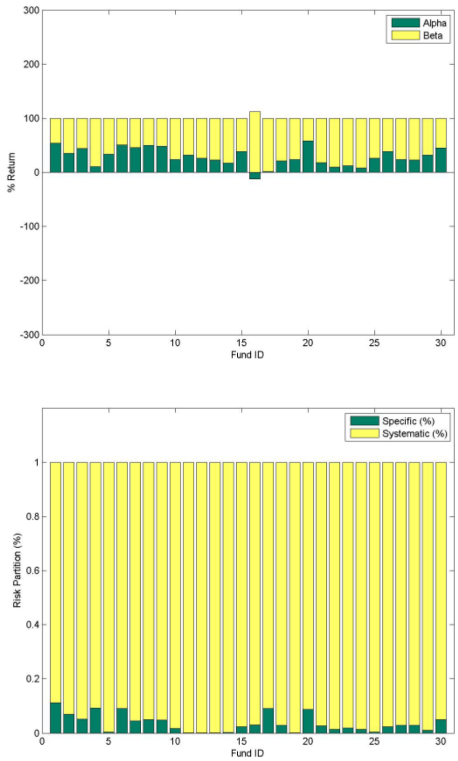

Figure 19 Stack up of risk and return attribution for data set DS3

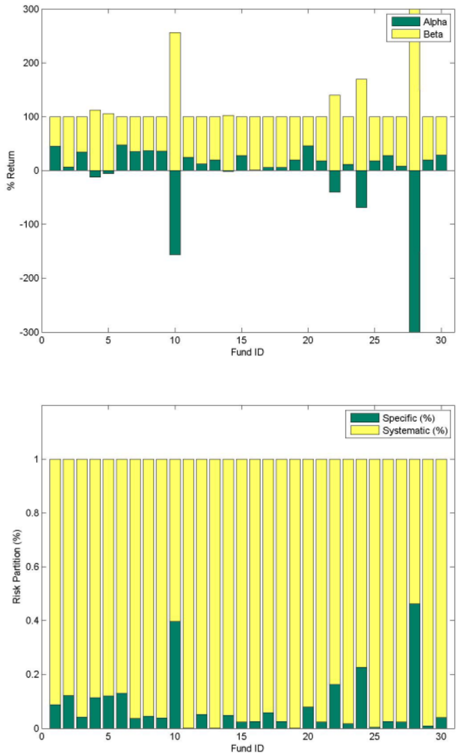

Figure 20 Stack up of risk and return attribution for data set DS4

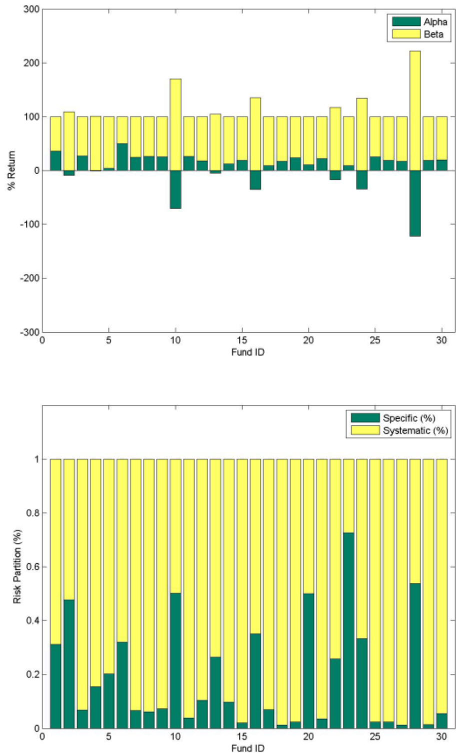

Figure 21 Stack up of risk and return attribution for data set DS5

6. Conclusion

A passive fund in a sense of practitioners usually means a fund that tracks SET50 well. It can be easily constructed just by picking major stocks listed in SET50. An active fund, on the other hand, is believed to have different composition from that of the passive fund and performs better than SET50. This is a fallacy. Good returns may be attributed to the market. The attribution analysis has shown that a seemingly active fund with different risk structure from that of the passive fund may indeed be passive. The findings will be obvious when risk-attribution structures stack up on return-attribution structure. Thus, an active fund should be active in both risk and return sides. From this perspective, all Thai equity funds under the study are passive. Retail investors may purchase blue-chip stocks listed in SET50 with appropriate weights and expect similar performance to top equity funds.

7. Acknowledgments

The author would like to thank ThaiBMA for the generous support of valuable data and access to computer facilities. Any opinions expressed in this article are those of the author and do not reflect the views of ThaiBMA.

8. References

[1] Capocci, D. and Hubner, G. 2004, “Analysis of Hedge Fund Performance”, Journal of Empirical Finance, Vol. 11, No.1, pp. 55-89.

[2] Carhart, M. M. 1997, “On Persistence

in Mutual Fund Performance", Journal of Finance, Vol. 52, No. 1, pp. 57-82.

[3] Connor, G. and Herbert, N. 1998, "Regional Modeling of the Western European Equity Market", BARRA-Working Paper, London

[4] Fama, E.F. and French, K.R. 1993, "Common Risk Factors in the Returns on Stocks and Bonds", Journal of Financial Economics, Vol. 33, No. 1, pp. 3-56.

[5] Lattin, J.M., Carroll, J.D., and Green, P.E. 2003, "Analyzing Multivariate Data", Toronto: Brooks/Cole, Thomson Learning, Inc.

[6] Srisansanee, C. 2009, "Sufficiency Economy in Globalization with Financial Engineering", The Fifth National Conference of Economists, NIDA, Thailand.