Association for Information Systems

AIS Electronic Library (AISeL)

ICEB 2001 Proceedings

International Conference on Electronic Business (ICEB)

Winter 12-19-2001

A Novel Index Portefolio Model by Minimizing the Absolute Tracking Error -- Empirical Stusies in the Taiwan Stock Market

Chieh-Yow ChiangLin

An-Pin Chen

Sue Feng Wang

A NOVEL INDEX PORTFOLIO MODEL BY MINIMIZING THE ABSOLUTE TRACKING ERROR
- EMPIRICAL STUSIES IN THE TAIWAN STOCK MARKET

*Chieh-Yow ChiangLin, **An-Pin Chen, **Sue Feng Wang
* Department of Information Management, Shuh-Chen University
** Institute of Information Management, National Chiao Tung University
1001 Ta Hsueh Road, Hsinchu, Taiwan 300, R.O.C.
e-Mail: clcy@iim.nctu.edu.tw

ABSTRACT

The index portfolio model attempts to form a portfolio whose time series in the market can trace the selected index as much as possible. The traditional index portfolio model, estimated coefficients models proposed by Salkin, established the portfolio by minimizing the square tracking error. In this paper, a novel index portfolio model formed by minimizing the absolute tracking error is proposed. In addition to preserving the characteristics of Salkin's model, the proposed model can guarantee obtaining the global optimum solution and, in contrast to Salkin's model, it can avoid the effect of the extreme value, which Salkin's model may not. Also in contrast to the traditional model, the proposed one is a linear programming model and can then include practical constraints in the models, including the transaction cost constraints and limited stock catalog constraints. How the improved models address these constraints would be discussed as well. Moreover, different empirical studies in the Taiwan Stock Market are provided to demonstrate the proposed model's effectiveness.

Key words: index portfolio, square error, absolute error, linear programming, quadratic programming

1. INTRODUCTION

The index portfolio mathematical models attempt to form a portfolio whose time series in the market can trace the selected index as much as possible. The index portfolio has been widely applied, including the formation of an index fund and establishment of an arbitrage portfolio for an index future.

The most popular portfolio model is the estimated coefficients model proposed by Salkin(1989). Salkin's model established the portfolio by minimizing the square tracking error. However, Salkin's model has several limitations. First, the global optimal solution can be obtained only under certain conditions because the mathematical programming model of Salkin's model is a quadratic programming (QP) model. Second, Salkin's model concentrates mainly on minimizing the square error. However, in practice, the mean square error may not be an adaptive proxy of the loss function. Third, extreme value data may heavily adversely impact the results from Salkin's model. Fourth, Salkin's model is not a linear programming (LP) model and then is difficult to consider complicated constraints, which would cause the original QP model to be a nonlinear programming (NP) model.

In this paper, a novel index portfolio model developed from the mean absolute deviation (MAD) model proposed by Konno and Yamazaki (1991) is proposed. In contrast to the traditional method, the proposed model forms the index portfolio by minimizing the absolute tracking error between historical returns of the portfolio and the selected index in the objective function to obtain the optimal portfolio investing weights. The MAD model was derived to replace the mean-variance (MV) model proposed by Markowitz in 1951. The MAD model attempted to transform the MV model into an LP problem from a QP problem. In this paper, the MAD model is reformulated and can be applied to forming the index portfolio. Since the proposed model is an LP model, it can overcome the limitations of Salkin' s model. Moreover, this study also provides empirical studies in the Taiwan stock market to demonstrate the effectiveness of the proposed model.

The rest of this paper is organized as follows. Section 2 reviews pertinent literature. The proposed model is then derived in Section 3. Section 4 presents the empirical study process and subsequent results. Section 5 then discusses different problems when forming the index portfolios. Conclusions are finally made in Section 6.

2. LITERATURE REVIEW

2.1 Index portfolio models

Rudd proposed two methods in 1980 to form an index portfolio: stratification model and optimization model. The optimization model was a mathematical programming model whose objective function was to minimize the residual risk as shown as in Equation (1). Their empirical results indicated that the optimization model was better than the stratification model.

$$
\begin{aligned}
& Min. w_P^2 \\
& s.t. \\
& b_P = \sum_{i=1}^{N} q_i b_i = 1 \\
& \sum_{i=1}^{N} q_i = 1 \\
& q_i \geq 0,\ for i = 1, \ldots, N
\end{aligned}
\qquad (1)
$$

Where w_P^2 denotes the residual risk of the portfolio, q_i represents the investing weight of stock i., âi is the â value of stock i, âp denotes the â value of the portfolio, and N represents the total number of stock categories.

Andrews et al. (1986) proposed three models for establishing the index portfolio: full replication model, stratified model, and sampling model. The full replication model duplicated the stock weights in the index and, thus, the conducts of the formed portfolio were the same as the index. However, the transaction costs of this model were too high to satisfy practical applications. In addition to that the stratified model maintained the same index weights in the portfolio as the index weights of the traced index, market values of the stocks determined the stocks weights. Notably, the stock would be rejected if the market value of a stock did not reach a certain level. As for the sampling model, it selected some representative stocks to form the portfolio. However, this method was somewhat subjective and inferior to the other two methods in empirical tests.

In 1989, Salkin defined the tracking error as Equation (2) to evaluate the performance of different index portfolio models. In this study, the tracking error function in Equation (2) is taken as one of the evaluation functions as well. Equation (2) is shown as follows:

$$
\begin{aligned}
p_{t,L} &= \frac{(P_t - P_{t-L})}{P_{t-L}} \\
d_{t,L} &= \frac{(I_t - I_{t-L})}{I_{t-L}} \\
R_L &= \sqrt{\sum_{t=L}^{T} (d_{t,L} - p_{t,L})^2}
\end{aligned}
\qquad (2)
$$

where P_t denotes the value of the portfolio in time instant t, I_t represents the value of the index in time instant t, $p_{t,L}$ is return of investment of the portfolio in past time interval L at time instant t, $d_{t,L}$ denotes return of investment of the index in past time interval L at time instant t, and R_L is the tracking error between the portfolio and index in past time interval L at time instant t.

Salkin proposed four index portfolio models: non-stratified estimated coefficients model, stratified estimated coefficients model, non-stratified capitalization weighted model, and stratified capitalization weighted model. In that work, empirical studies were performed by the weekly data from January 1985 to December 1986 in the Japan stock market. He concluded that the non-stratified estimated coefficients model was the best among those models.

The non-stratified estimated coefficients model can be described as follows:

$$Min.R^2 = \sum_{t=1}^{T} (d_t - \sum_{j=1}^{S} w_j r_{j,i})^2 \qquad (3)$$

where d_t denotes the return of the index at time instant t, $r_{j,I}$ represents the return of the stock i at time instant t, w_j is the invested weight of stock j which would be determined by the model. In contrast to the non-stratified estimated coefficient model, the stratified estimated coefficients model included the industry weights constraints in the model.

Basically, the estimated coefficient model is a QP model and can obtain only the local optimal solution if the quadratic matrix in the objective function is not positive-definite.

2.2 MAD model

In 1991, Yamakazi and Konno proposed the MAD model to replace the MV model of Markowitz (1952). However, the MAD model did not consider the covariance relationship between assets, thereby inducing the estimate risk. However, this problem would not arise if the MAD model were reformulated to establish an index portfolio model. The MAD model can be described as follows:

$$\begin{aligned} & Min. \sum_{t=1}^{T} \left| \sum_{j=1}^{N} (r_{j,t} - r_j) x_j \right| \\ & s.t. \\ & \sum_{j=1}^{N} r_j x_j \geq rM, \\ & \sum_{i=1}^{N} x_j \leq M, \\ & 0 \leq x_j \leq u_j, j = 1,...,N \end{aligned} \qquad (4)$$

where x_j denotes the invested capital of the asset j, $r_{j,t}$ represents the return of the asset j at time instant t, r_j is the average return of the asset j, ρ denotes the minimal return of the portfolio, and u_j represents the upper invested capital level of stock j.

The MAD model can be transformed into an LP problem by applying the deviation variables and, in doing so, the global optimal solution is obtained. In a related work, Feinstein and Thapa (1993) modified the MAD Model to enhance the solving efficiency. The model of Feinstein and Thapa is shown as follows:

$$\begin{aligned} & Min.Z = \sum_{t=1}^{T} (n_t + v_t) \\ & s.t.: \\ & n_t - w_t - \sum_{j=1}^{n} (r_{jt} - r_j) x_j = 0 \\ & n_t \geq 0, w_t \geq 0 \\ & c_1, c_2, c_3 \end{aligned} \qquad (5)$$

Furthermore, Li, Chen, and ChiangLin (1998) modified Feinstein' s model to enhance the solving efficiency. Their model is shown as follows:

$$\begin{aligned} & Minimize\, 2\sum_{t=1}^{T} w_t \\ & s.t. \\ & \sum_{j=1}^{n} (r_j - r_{jt}) x_j - w_t + s_t = 0 \\ & w_t \geq 0, s_t \geq 0 \\ & c_1, c_2, c_3 \end{aligned} \qquad (6)$$

3. PROPOSED MODEL

3.1 The proposed index portfolio model

By reformulating the MAD model, the proposed model can be described as follows which can be applied to form the index portfolio:

$$\begin{aligned} & Min. \sum_{t=1}^{T} \left| \sum_{j=1}^{N} r_{j,t} x_j - r_t \right| \\ & s.t. \\ & \sum_{i=1}^{N} x_j = 1, \\ & 0 \leq x_j \leq u_j, j = 1,...,N \end{aligned} \qquad (7)$$

where x_j denotes the invested weight of the asset j, $r_{j,t}$ represents the return of asset j at time instant t, r_t is the return of the index at time instant t, and u_j denotes the upper limitation of the invested capital of the asset j.

Exactly why the equation (7) can be applied in the constitution of the index portfolio can be described as follows.

The objective function of equation (7) is to minimize the summation of the absolute tracking error between the returns of the portfolio and the traced index at every time instant. Therefore, the weights of the resulting portfolio are the optimal weights of the assets by the mathematical programming procedure. Thus, the resulting portfolio can trace the index.

The above model can be modified according to the method of Li, Chen and ChiangLin to enhance the computational efficiency. The modified model can be as shown as follows:

$$\begin{aligned}
&Minimize\, 2\sum_{t=1}^{T} w_t \\
&s.t. \\
&\sum_{j=1}^{n} (r_t - r_{jt}) x_j - w_t + s_t = 0 \qquad (8) \\
&w_t \geq 0, s_t \geq 0 \\
&c_1, c_2
\end{aligned}$$

3.2 Comparison between the proposed model and Salkin's model

The differences between the proposed model and Salkin's model can be summarized as follows:

1. The proposed model minimizes the absolute error between the portfolio and the index during the analytical period. In contrast, Salkin's model minimizes the square error. In practical applications such as forming the index fund portfolio, the loss derived from tracking error of the index portfolio is always the absolute tracking error but not the square tracking error.
2. The proposed model is an LP model and, thus, the global optimal solution can be obtained in any case. However, Salkin's model is a QP model, which cannot ensure that the global optimal solution is obtained.
3. Salkin's model traces the index by the squared error and, therefore, would be affected by the extreme value data more seriously than the proposed model.
4. The fact that the proposed model is an LP model accounts for why more complicated constraints can be included in the model and would not hinder the results of the global optimal solution.

Besides, the proposed model contains excellent characteristics of Salkin's model, including the following:

1. The proposed model can include the industrial stratified constraints.
2. The proposed model can consider other constraints, such as â constraint or the exposure limitations for different risk factors.
3. Investors can arbitrarily select the analytical period, the analytical frequency and the number of stocks.

In the following section, some empirical tests are performed to demonstrate the effectiveness of the proposed model.

4. EMPERICAL STUDIES

4.1 Data description

This section compares the performance between Salkin's model and the proposed model using historical data from the Taiwan stock market. The daily returns of different stocks were calculated by the formula as shown in equation (2). The indexes to be traced include the Simex Index and the Taiwan Stock Weighted Index (TSWI). Stocks in the index portfolio were arbitrarily selected from different industry catalogs as listed in Table 1. This table also contains the industrial attributes of selected stocks. The analytical period ranges from 1997/1/4 to 1998/11/30. Evaluation functions for both models include the mean square error (MSE), mean absolute error (MAE), and correlation function.

Both programming models are computed by LINGO Hyper Release 4.0 (LINDO Systems, Inc., 1998) on a PC Pentium II 400 with 128 MRAM. The LINGO produced by LINDO System Inc. is a mathematical programming package widely uses in personal computers.

4.2 Comparison of the non-stratified cases

In this sub-section, we compare the non-stratified cases. The proposed model and Salkin's model are established by arbitrarily selecting twenty-nine stocks in the Taiwan stock market as listed in Table 1. Table 2 summarizes the empirical results for tracing Simex index by two models.

The in-sample period ranges from 1998/1/3 to 1998/9/30, and the out-of-sample period lasts from 1998/10/1 to 1998/11/30. Table 3 display the empirical results of TWSI index portfolio.

Table 1 Selected stocks and their industrial attributes

Industry Stock	1	2	3	4	5	6	7	8	9	10	11	12	13
1101	1	0	0	0	0	0	0	0	0	0	0	0	0
1202	0	0.36	0	0	0	0	0.01	0	0	0	0.21	0.42	0
1216	0	0.94	0	0	0	0	0	0	0.06	0	0	0	0
1301	0	0	1	0	0	0	0	0	0	0	0	0	0
1402	0	0	0	1	0	0	0	0	0	0	0	0	0
1433	0	0	0.23	0.77	0	0	0	0	0	0	0	0	0
1504	0	0	0	0	1	0	0	0	0	0	0	0	0
1605	0	0	0	0	0.98	0	0	0	0.02	0	0	0	0
1710	0	0	1	0	0	0	0	0	0	0	0	0	0
1802	0.77	0	0	0	0	0	0	0	0	0.23	0	0	0
1902	0	0	0	0	0	0.78	0.22	0	0	0	0	0	0
2002	0	0	0	0	0	0	0	0	0.01	0	0.99	0	0
2105	0	0	0.92	0	0	0	0	0	0.08	0	0	0	0
2201	0	0	0	0	0	0	0	0	0	0	0	1	0
2303	0	0	0	0	0	0	0	0	0	1	0	0	0
2306	0	0	0	0	0	0	0	0	0	1	0	0	0
2311	0	0	0	0	0	0	0	0	0	1	0	0	0
2330	0	0	0	0	0	0	0	0	0	1	0	0	0
2342	0	0	0	0	0	0	0	0	0	1	0	0	0
2506	0	0	0	0	0	0	1	0	0	0	0	0	0
2515	0	0	0	0	0	0	1	0	0	0	0	0	0
2603	0	0	0	0	0	0	0	0	0	0	0	0	1
2802	0	0	0	0	0	0	0	1	0	0	0	0	0
2803	0	0	0	0	0	0	0	1	0	0	0	0	0
2804	0	0	0	0	0	0	0	1	0	0	0	0	0
2805	0	0	0	0	0	0	0	1	0	0	0	0	0
2806	0	0	0	0	0	0	0	1	0	0	0	0	0
2903	0	0	0	0	0	0	0	0	1	0	0	0	0
9907	0	0	0	0	0	0	0	0	1	0	0	0	0
Simex	0.037	0.041	0.084	0.050	0.055	0.012	0.022	0.306	0.043	0.244	0.050	0.035	0.021

Table 2 Performance summary for tracing the Simex index by two models

Portfolio	In sample			Out of sample		
Evaluation Function	Simex	Salkin’s Model	Proposed Model	Simex	Salkin’s Model	Proposed Model
Return	-0.1120	-0.1015	-0.0892	0.2759	0.3085	0.3024
Risk	2.2848	2.3799	2.3737	3.4341	3.6441	3.5473
Correlation	1	0.9704	0.9710	1	0.9927	1
MSE	0	0.0224	0.0259	0	0.0239	0.0191
MAE	0	0.1109	0.1089	0	0.1235	0.1058

Table 3 Performance summary for tracing TWSI index by two models

Portfolio	In sample			Out of sample		
Evaluation Function	Simex	Salkin’s Model	Proposed Model	Simex	Salkin’s Model	Proposed Model
Return	-0.0738	-0.1015	-0.0892	0.1306	0.3085	0.3024
Risk	2.0553	2.3799	2.3737	2.9079	3.6441	3.5473
Correlation	1	0.9726	0.9718	1	0.9890	0.9894

MSE	0	0.1118	0.1141	0	0.1838	0.1628
MAE	0	0.2582	0.2574	0	0.3253	0.3032

According to the empirical results, regardless of whether in Simex index portfolio case or in TWSI index portfolio case, the performances of Salkin' s model in both cases were better than the proposed model in the sample if the MSE was taken as the evaluation function. However, if the MAE was taken as the evaluation function, the performances of the proposed model in both cases were better than Salkin' s model in the sample regardless of whether in Simex portfolio case or in TWSI portfolio cases. These results are acceptable because the proposed model and Salkin' s model are designed to minimize the MAE function and MSE function, separately.

4.3 Comparison of the stratified cases

In this sub-section, the industry weight constraints were included in both the proposed model and Salkin' s model. Table 1 lists the industry weights of the selected stocks and the Simex index at time instant 1998/9/30. Table 4 summarize the empirical results of these cases.

Table 4 Performance summary for tracing Simex index by two models

– Stratified cases

Portfolio Evaluation Function	In sample			Out of sample		
	Simex	Salkin' s Model	Proposed Model	Simex	Salkin' s Model	Proposed Model
Return	-0.1120	-0.0997	-0.0966	0.2759	0.3210	0.3252
Risk	2.2848	2.4050	2.4243	3.4341	3.6798	3.6463
Correlation	1	0.9896	0.9892	1	1.0234	1.0242
MSE	0	0.0272	0.0295	0	0.0269	0.0225
MAE	0	0.1228	0.1220	0	0.1224	0.1130

According to the empirical results, the performances of both models are inferior to the non-stratified cases as shown in Table 2. This is reasonable because more constraints in the model would reduce the solution space and, in doing so, the optimal solution would be worse than that in the non-stratified case. However, industrial stratification is an attempt to obtain the better results in the forecasting period. According to the results of the out-of-sample in Table 2 and Table 4, the stratified cases are not inferior to the non-stratified ones.

4.4 Empirical tests of different analytical periods

In this sub-section, we present cases of different analytical periods. As Table 2 indicates, the interval of the analytical period is 9 months. Table 5 summarizes the performances from two other cases.

Table 5 Performance summary for another two cases

Portfolio Evaluation Function	1998/4/1-1998/9/30 (164 data)			1998/7/1-1998/9/30 (71 Data)		
	Simex	Salkin' s Model	Proposed Model	Simex	Salkin' s Model	Proposed Model
Return	-0.2140	-0.1971	-0.1871	-0.1513	-0.1301	-0.0991
Risk	2.0948	2.1790	2.2075	2.5176	2.5685	2.8872
Correlation	1	0.9889	0.9879	1	0.9816	0.9771
MSE	0	0.0225	0.0281	0	0.0229	0.0626
MAE	0	0.1153	0.1069	0	0.1894	0.1087

According to the empirical results, when the MAE function is taken as the evaluation function, different analytical periods do not affect the results. However, when the MSE function is taken as the evaluation function, different analytical periods affect the results. This phenomenon is owing to that longer analytical periods include more extreme value data.

4.5 Empirical results of different selected stocks

In this sub-section, we select seventy-seven stocks to form the index portfolio for the sake of comparing the performances of the twenty-nine stocks case. Table 6 summarizes the empirical results.

Table 6 Performance of tracing Simex index by two models

- Seventy-seven selected stock case

Portfolio Evaluation Function	In sample			Out of sample		
	Simex	Salkin' s Model	Proposed Model	Simex	Salkin' s Model	Proposed Model
Return	-0.112	-0.0895	-0.0742	0.2759	0.2749	0.2592
Risk	2.2848	2.3082	2.3042	3.4341	3.4443	3.4284
Correlation	1	0.9886	0.9888	1	1.0000	1.0000
MSE	0	0.0010	0.0106	0	0.0117	0.0112
MAE	0	0.0568	0.0552	0	0.0821	0.0829

According to the empirical results in Table 2 and Table 6, the performances of seventy seven selected stock portfolio are better than those of twenty nine selected stock portfolios. This is owing to that more selected stocks extend the solution space and, in doing so, a better solution can be obtained. However, the portfolio from more selected stocks implies higher transaction costs. Moreover, when the total invested capital is fixed, the portfolio from more selected stocks may conduce odd size investment.

5. DISCUSSION

1. Owing to the decrease of the solution space, the performance in the sample worsens more than that in the non-stratified case when forming the index portfolio by the industrial stratified method. The stratified method attempts to catch the industrial characteristic of the index. If the stratified methods were adequate, the subsequent performance of the forecasting period is better than that of the non-stratified model. However, stratification is relatively difficult because a company may contain many industrial attributes as shown in Table 1.
2. The industry-stratified constraints in the model include three methods. One method is to limit industry weights of the index portfolio the same as those of the traced index. This method significantly decreases the solution space and may cause no solution. The second method limits the industrial weights of the index portfolio between certain levels, which are derived from the traced index. The third method penalizes the industrial weight biases between the index portfolio and the traced index in the objective function of the models. However, determining the levels in the second method and the penalized parameters in the third method is rather difficult and may be a subjective task.
3. Selecting more stocks to form the index portfolio may lead to higher transaction costs. However, selecting fewer stocks to form the index portfolio may lead to investment of too much capital in a single stock, ultimately increasing the liquidity cost.
4. The index portfolio can be applied to form an arbitrage portfolio or an index fund. On the other hand, the index portfolio model can include the views from the valuation model to form a portfolio combining passive and active investment.
5. Before forming the index portfolio, a pre-process filter can be applied to select stocks, thereby enhancing the stability of the portfolio.

6. Selection of the evaluation function should consider the loss function of the practical application.

6.CONCLUSION

This study presents a novel index portfolio model. In contrast to the traditional Salkin’s model, the proposed model can obtain a global optimal solution and is unaffected by extreme value data. Owing to that the proposed model is a linear programming model, more constraints can be included into the model to consider practical limitations when forming the index portfolio. These constraints include the transaction cost constraint and limited stock catalog constraint.

REFERENCES

[1] Andrews.C., W., Ford D. and Mallison K. “The Design of Index Fund and Alternative Methods of Replication.”, The Investment Analyst, 82:pp 13-16, 1986

[2] Anthoney, Herbest, “Capital Budgeting, Theory, Quantitative Methods, and Applications”, 1982

[3] Evans, John and Stephen H. Archer, “Diversification and Reduction of Dispersion: An Empirical Analysis”, Journal of Finance, vol 23, No.5, p.765, 1968

[4] Feinstein, C. D., and M. N. Thapa, “Mean-Absolute Deviation Portfolio Optimization.”, Management Science, (38)1993, 1552-1553

[5] Markowitz, Harry M., “Portfolio Selection”, Journal of Finance, pp, 77-91, March 1952

[6] Meade N. and Salkin G.R. “Developing and Maintaining an Equity Index Fund”, Journal of Operational Research Society, 41:pp.599-607, 1990

[7] Meade N. and Salkin G.R. “Index Fund-construction and Performance Measurement”, Journal of Operational Research Society, 40:pp.871,-879, 1989

[8] Rudd, A. “Optional Selection of Passive Portfolio”, Financial Management, Spring, pp. 57-66, 1980

[9] Scott Mason, Robert Merton, Andre Perold, Peter Tufano, “Cases in Financial Engineering : Applied Studies of Financial Innovation”, 1995

[10] Yamakozi, H. and H. Konno, “Mean Absolute Deviation Portfolio Optimization Model and Its Application to Tokyo Stock Market.”, Management Science., 37(1991), 519-531