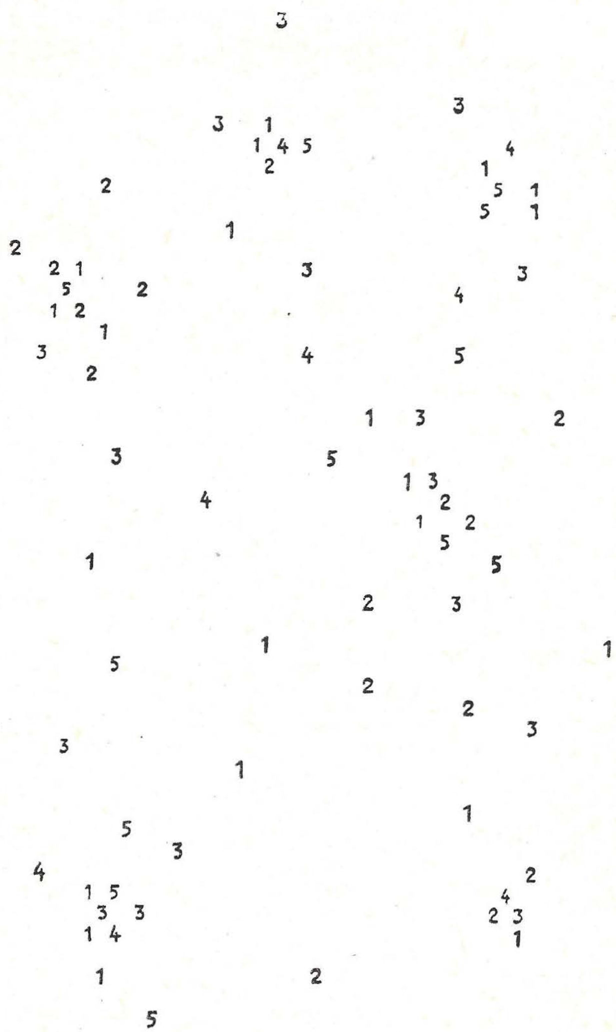
summarized in Table 1. For each example this table contains the minimal (A), average (B), and maximal (C) values of the objective function divided by ten and rounded to four digits for those 20 initial partitions and the minimal (D) and maximal (E) value for the corresponding final partitions after applying the modified exchange method. Table 2 contains, again for the same four examples, the minimal (F), average (G) and maximal (H) percentage gain obtained in this way and calculated



```
M= 73   N=5  NR= 1   IT= 2   D= 0.2787E+05   (TRW)

1 : (21)    0.8249E+04       2 : (16)    0.5868E+04
3 : (16)    0.6272E+04       4 : ( 8)    0.3236E+04
5 : (12)    0.4240E+04
```

Fig. 1

Table 2

Example	n	2	3	4	5	6	7	8	9
1	F	0.0	0.0	1.1	4.7	7.0	6.6	6.7	8.9
	G	4.6	7.5	9.8	13.8	19.9	23.1	26.2	30.1
	H	16.7	30.0	20.3	29.0	32.2	36.0	39.9	45.6
2	F	0.1	1.4	1.4	3.3	6.1	4.1	6.9	10.5
	G	2.5	5.8	8.3	12.5	16.4	17.5	21.9	24.5
	H	10.6	15.4	25.5	35.0	35.2	40.9	38.4	44.7
3	F	0.0	1.2	2.0	2.4	6.8	8.9	10.6	12.3
	G	2.3	4.6	8.2	11.8	14.4	18.2	20.3	24.1
	H	13.7	13.9	24.0	43.5	35.7	50.2	37.7	42.6
4	F	0.0	0.0	1.1	1.5	2.4	3.7	5.0	5.2
	G	1.5	3.2	4.0	5.8	7.8	9.9	10.9	12.9
	H	5.2	7.6	8.4	13.4	21.1	17.4	22.6	24.3

according to the figures of Table 1. For illustration, Fig. 1 contains the group memberships, for example 4 and $n=5$. (The position of the points (objects) is given by the values of the two variables).

4. Conclusion

As it can be seen from the Tables 1 and 2 the modified exchange method works very well. The values of the final partitions are nearly equal in all cases. (This is different when applying the exchange method for (2)). For $n=2, ..., 9$ the values of the objective function decrease very slowly, i.e., the value of (7) is indeed nearly zero all the time. For larger n, of course, this would not have to be that way. Finally the gain in the objective function value as against random partitions is remarkable and is increasing with the number of groups.

References

[1] BRAUN H. Strukturanalyse eines Tankstellennetzes. In: Späth H. Ed. Fallstudien Operations Research. Bd. 1. Munich, R. Oldenburg, 1978.

[2] SPÄTH H. Cluster dissection and analysis. Theory, examples, and FORTRAN programs Chichester, Horwood, 1985.

Antyklustering: kryterium maksymalizacji wariancji

W pewnych zastosowaniach zbiór m obiektów scharakteryzowanych przez wartości s zmiennych x_{ij} $(i=1, ..., m, j=1, ..., s)$ powinien być rozbity na n możliwie podobnych podzbiorów. Po-

kazano, że w tym przypadku odpowiednim podejściem jest maksymalizacja wariancji, tzn. szukanie takiego rozbicia $C_1, \ldots, C_n$ zbioru $\{1, \ldots, m\}$, dla którego

$$\sum_{j=1}^{n} \sum_{i \in C_j} \|x_i - \bar{x}_j\|^2$$

osiąga wartość maksymalną. Podano heurystyczną metodę rozwiązania tego zagadnienia i przytoczono wyniki obliczeń dla kilku przykładów.

Антикластеризация: критерий максимизации дисперсии

В некоторых приложениях совокупность m объектов характеризируемых значения s переменных x_{ij} $(i=1, \ldots, m,\ j=1, \ldots, s)$ должна быть разбита на n, возможно сходных подсовокупностей. Показано, что в этом случае надлежащим подходом является максимизация дисперсии, то есть поиск такого разбиеиия $C_1, \ldots, C_n$ совокупности $\{1, \ldots, m\}$, для которого

$$\sum_{j=1}^{n} \sum_{i \in C_j} \|x_i - \bar{x}_j\|$$

достигает максимального значения. Предложено эвристический метод решения этой задачи и приведено результаты вычислений для нескольких примеров.

Control
and Cybernetics
VOL. 15 (1986) No. 2

Selecting optimal specimens of stock populations

by

JOHN M. MULVEY
THOMAS S. NADBIELNY
Princeton University

A systematic approach is proposed for selecting a representative subset — called a "specimen" — of an identified time-varying population of stocks. The procedure is based upon a mathematical programming model which minimizes loss of information between the target population and the specimen. It is shown that, for two distinct populations and several independent forecast periods, the movements of the optimal specimen and the population are almost identical — generally within 1%.

1. Introduction

Investors are often faced with the task of selecting from among a pre-screened population of promising stocks. Many investment services, for example Merrill Lynch and Value Line, publish lists of "best buys" on a weekly or monthly basis. Since historical evidence indicates that several of these lists have outperformed the overall market averages on a risk-adjusted basis? (ignoring transaction costs), Black (1975), the investor may focus on one of these groups in an attempt to form a representative well diversified portfolio. This report describes an optimization approach for choosing such a subset, called a specimen.

As another example, the investor may wish to mirror desirable industries (e.g., computer companies) or the markets of particular countries (e.g., Japan) in which stock index funds, Babson (1976), Ehrbar (1976), are lacking. In most of these instances, transaction and informational costs prevent the investor from purchasing the entire (or a large segment of the) population. Again, the problem is selecting a representative specimen.

The generic problem is constructing a portfolio to mirror movements of a target population of n stocks. In constructing a specimen, a critical concern is the loss of information resulting from those stocks excluded from the specimen. Of course, there are innumerable ways to measure informational loss. One simple way is to compare the summary statistics that are generated by the population with the summary statistics of the specimen.

Letting

$\{I\}$: set of population variables with values a_{ik} for $k=1, ..., l$ attributes, $i \in I$

$\{J\}$: set of variables eligible for inclusion in specimen $\{J\} \subseteq \{I\}$

$\{J^*\}$: set of variables chosen for the specimen $\{J^*\} \subseteq \{J\}$

$$y_j = \begin{cases} 1 & \text{if } j \in J^* \\ 0 & \text{otherwise} \end{cases} \quad \text{all } j \in J$$

w_i: coefficient weight for variable $i \in \{I\}$ in unabridged population (e.g. number of shares)

and w_j^*: revised weights for variables in specimen $j \in J^*$

The objective then translates into choosing $\{J^*\}$ and the accompanying weights $\{w_j^* \mid j \in J^*\}$ such that the summary statistics of the population and the specimen are approximately equal:

$$\text{[MSP]} \quad \varphi(I) \approx \varphi(J^*)$$

where $\varphi(\cdot)$ depicts a vector operation, e.g., ordinary summation or variance/covariance calculations. To illustrate, the obvious formula for the total market value of the population and the specimen at time period k is shown below:

$$\varphi_k(I) \triangleq V_k = \sum_{i \in I} w_i a_{ik} \quad \text{for all } k,$$

$$\varphi_k(J^*) \triangleq W_k = \sum_{j \in J^*} w_j^* a_{ik} \quad \text{for all } k,$$

where a_{ik}=price of security i, time period k.

One approach for solving this problem is statistical sampling. Numerous techniques including simple random, stratified, and clustered sampling, Cohran (1977), have been used for decades when forming a "random sample". These statistical techniques are effective when the target population is unavailable for comparison. Also, the dynamic aspects of the problem contribute to the difficulties in using these techniques, Mosteller and Tukey (1978). As discussed Mulvey (1980), however, the random statistical techniques may be inferior to deterministic models when the population or its underlying characteristics is available for testing.

In the next section, we describe a mathematical programming model for solving this problem. Although the resulting model is quite large, it has a special structure which allows for efficient solution via a combined heuristic/relaxation method Geoffrion (1970), Mulvey and Crowder (1979). The technique generally locates solutions within 1% of the optimum for practical size problems (e.g., n=300).

The empirical results, presented in section 3, show that the optimization model's performance exceeds those of randomly generated portfolios.

2. Mathematical Model and Stock Selection Procedure

The following simplified mathematical programming model often renders high-quality solutions to [MSP] over numerous performance measures:

$$\text{[DAM]} \quad v_q(\bar{x}) \stackrel{\Delta}{=} \min \sum_{i \in I} \sum_{j \in J} d_{ij} x_{ij}$$

subject to

$$\sum_{j \in J} y_j = q \tag{1}$$

$$\sum_{i \in I} w_i x_{ij} \leqslant W_j \quad j \in J \tag{2}$$

$$\sum_{j \in J} x_{ij} = 1 \quad i \in I \tag{3}$$

$$x_{ij} \leqslant y_j \quad i \in I,\ j \in J \tag{4}$$

$$x_{ij} \in \{0, 1\},\ y_j \in \{0, 1\} \quad i \in I,\ j \in J \tag{5}$$

where

$$d_{ij} = \Big[\sum_{k=1}^{l} (a_{ik} - a_{jk})^r\Big]^{1/r} \quad i \in I,\ j \in J \tag{6}$$

w_i=weight for $i \in I$
W_j=capacity for $j \in J$
and $|I|=n$ $|J|=m$.
q=number of stocks in specimen portfolio

Again, $\{I\}$ is the population consisting of n stocks (variables) and l attributes per stock, whereas $\{J\}$ indicates those variables eligible for the specimen*. In testing, relative price (price plus reinvested monetary and stock dividends) was employed as the relevant stock attribute a_{ik}. Constraint (1) fixes the number of variables in the specimen at q. Binary variable x_{ij} designates whether variable $i \in I$ will be associated with variable $j \in J$*. Constraints (3) require that all population stocks have a single association. Distance metric d_{ij} measures the similarity of stocks i and j across l time periods. After solving [DAM], the revised weights are calculated by the equation

$$w_j^* = \sum_{i \in I} w_i x_{ij}^* \quad \text{for all } j \in J \tag{7}$$

where x_{ij}^* is the solution to [DAM]; these weights identify the proper proportion of stock to purchase for the specimen.

The objective of model [DAM] is to select a subset of q stocks out of m in the eligible population such that the total loss of information, as measured by $v_q(\cdot)$, is minimized.

* In most instances, $\{I\}=\{J\}$

This approach is based upon the optimal clustering of stocks into q homogeneous groups, followed by a stock selection procedure in which one stock from each compact cluster, the median, is selected for the specimen. First, the n population stocks are clustered into q homogeneous groups, as defined by the pairwise sum of distances** for all stocks within a cluster. Clusters can be formed to identify industries or risk classes, Lintner (1965), Rosenberg (1974). Since an optimization model is used, the resulting clusters are guaranteed to be as compact as possible. Once these clusters are identified, a single stock is chosen from each cluster and placed into the specimen. Weights are then transferred from all stocks in a cluster to their associated median. The approach has been used successfully within several contexts, including the U.S. Treasury Department to select a representative population of taxpayers for micro-economic tax studies, Mulvey (1980).

Model [DAM] has several inherent advantages over the usual heuristic cluster analysis methods Anderberg (1973). It has a well defined objective function — $v_q(\cdot)$ — that measures the amount of compactness in each cluster and the total loss of information. Thus the specimen is directly comparable with the population for multiple values of q; this sensitivity analysis is valuable when setting the number of stocks in the specimen. (See Section 3.3). Second, the solution is generally robust with respect to errors in the data, since a guaranteed optimal solution is generated. Third, model [DAM] is able to limit the size of any single cluster by means of constraint (2), thereby preventing the specimen from being dominated by any single stock. This last feature reduces the risk of purely random movements due to idiosyncratic stock behavior.

In comparison with Sharpe's classical single index model, Sharpe (1967, 1970), a multi-index strategy is adopted in which a cluster corresponds to a group of stocks with similar price comovements. Arnott 1980 uses a heuristic cluster analysis for identifying extra-market risk factors, also see Farrell, Jr. (1974). Instead of employing a quadratic programming procedure, a la Markowitz, small highly diversified (heterogeneity across clusters) subsets are constructed which in aggregate mirror the target population. The construction of the specimen can be viewed as a data aggregation problem with dynamic aspects. Orcutt, Watts and Edwards (1968), Zipkin (1980).

It should be emphasized that [DAM] was unsolvable for practical size problems ($n \geqslant 200$) until recently since the number of decision variables grows as a function of n^2+n (e.g., 40, 200 variables for $n=200$ securities). Thus the approach's usefulness had been greatly limited. A similar situation had occurred with Markowitz's quadratic programming [QP] model, Markowitz (1959) until Perold (1981) developed an efficient computer program for solving large [QP] problems. Over the past few years, Lagrangian relaxation methods, Geoffrion and McBride (1978), Held, Wolfe and Crowder (1979), have been proven effective for solving highly structured integer programming models such as [DAM]. Mulvey and Crowder (1979) specialized these concepts for the uncapacitated optimal clustering problem.

** Dynamic aspects are taken into account by including time dependent attributes in the distance function. See Section 3.

3. Empirical Results

The investment strategy was tested for two populations of stocks. Stocks comprising the S & P 500, representing almost 80% of the total market value of the New York Stock Exchange; were chosen as the first target population. The second target population was composed of 33 "best buy" stocks from a recent list of a well-known investment advisory service.

In assigning stocks to clusters, the following distance function was incorporated:

$$d_{ij}=\left\{\sum_{t=1}^{T}\left\{\left[\log(p_{i,t})-\sum_{k=1}^{T}\log(p_{i,k})/T\right]-\left[\log(p_{j,t})-\sum_{k=1}^{T}\log(p_{j,t})/T\right]\right\}^{2}\right\}^{1/2}$$

where d_{ij}=relative similarity of stock i and j's price movement
$p_{i,t}$=adjusted price of stock i, time period t
$[1, T]$=calibration period.

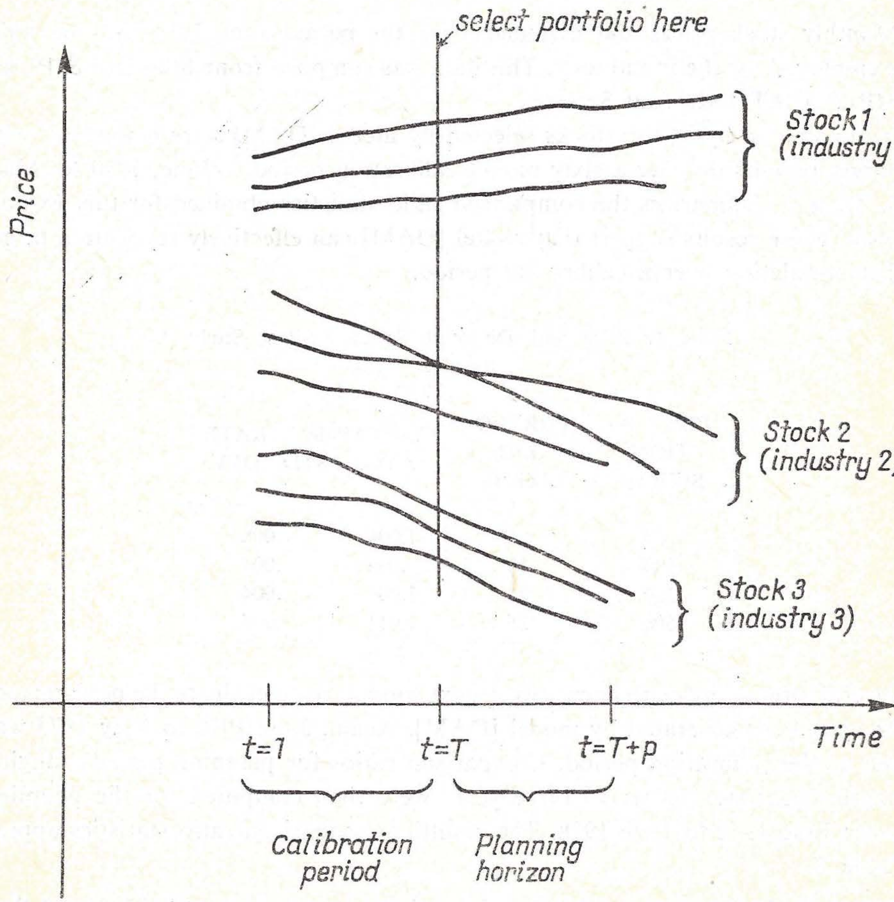


Fig. 1. Dynamic aspects of portfolio selection

In contrast to static cluster analysis, here the problem is complicated with the introduction of the time dimension. Essentially the model attempts to capture the time invariant aspects of a dynamic population. In practice, given time period T, the model calibrates over the period $[1, T]$ to determine the pairwise distances between all stocks i and j. Figure 1 depicts the situation.

Using these distances in conjunction with model [DAM], the optimal specimen is then chosen at period T. The stocks comprising the specimen $\{J^*\}$ are held throughout the planning horizon — period $[T, T+p]$. Transactions are disallowed during this period.

A "comparison ratio" statistic was developed to evaluate differences between the target population $\{I\}$ and the specimen $\{J\}$. At period k, the market value of the population (V_k) and the specimen (W_k) (both normalized as of time period T) are compared. Comparison rations for an ideal specimen would equal unity for all k.

3.1. Population #1: the S&P 500

Monthly stock prices and dividends* for the period June 1970 to June 1978 were employed for the initial tests. This data was compiled from Standard & Poor's COMPUTSTAT Industrial Service.

In the first analysis, the stocks selected by model [DAM] were compared with the target population over a sixty month calibration period — June, 1970 to May. 1975. Table 1 summarizes the comparison ratio statistics obtained for this ex-post analysis. These results suggest that model [DAM] can effectively replicate a target market population over a calibration period.

Table 1. Price and Dividend Series Ex-Post Study Results

POPULATION SIZE (n)	PORTFOLIO SIZE (q)	COMPARISON RATE AVG.	STD. DEV.
50	5	1.006	.008
100	10	.996	.004
250	25	1.000	.004
500	25	1.011	.005

Several forecasting experiments were next conducted to evaluate the performance of the specimens generated by model [DAM]. Again, June 1970 to May 1975 was chosen as the calibration period. Comparison ratios for planning periods ranging in length from one quarter to three years were then computed for the planning horizon: June 1975 to June 1978. The resulting comparison ratio statistics appear in Table 2.

* Here, dividends were assumed to be reinvested at an opportunity rate of 10% and these dividends taxed at 40%. Prices were adjusted for stock splits and stock dividends.

Table 2. S&P 500 Forecasting Results

Portfolio Size	Forecast Length	Comparison Ratios Average	Std. Dev.
10	1 Qtr	1.000	.001
10	2 Qtr	.999	.001
10	1 Yr	.998	.001
10	3 Yr	.991	.005
25	1 Qtr	.999	.001
25	2 Qtr	.999	.001
25	1 Yr	.999	.001
25	3 Yr	.995	.003
50	1 Qtr	1.001	.001
50	2 Qtr	1.000	.001
50	1 Yr	.998	.002
50	3 Yr	.997	.002
100	1 Qtr	1.000	.001
100	2 Qtr	1.001	.001
100	1 Yr	1.002	.002
100	3 Yr	1.001	.003

Note that the majority of comparison ration statistics lie close to the ideal value of 1. Not surprisingly, comparison ratios for shorter intervals are found to be closer to 1 than those for longer intervals; also the variance of the comparison ratios increases with forecast length.

3.2. Population #2: 33 "Best Buy" Population

A list of the 33 "best buy" stocks is presented in Table 3. Daily data on prices and dividends for the period August 23, 1982 to February 8, 1983, which corresponded to a major bull market, was used to form this target population. This data was compiled from the Dow Jones New Retrieval Service (DJNS). Prices were again adjusted for dividends which were assumed to be reinvested and stock splits. It should be noted that during this time period an investor would have realized a sizable return if he had been able to mirror the results of the target population. The relative performance of the best buy population to that of the S & P 500 over the planning period is exhibited in Figure 2.

Ex-post and forecasting tests were conducted using specimen sizes ranging from 2 to 6 (q) elements. Results for a 60 day calibration period and a 59 day forecast period are presented in Table 4. Summary statistics from 1000 equally-weighted randomly-generated porfolios are also presented. Using the average absolute difference between the comparison ratio and one as a performance measure, the specimens are much more highly associated with the target population than the equally weighted randomly generated portfolios. Table 5 presents the statistics for a 90 day calibration period and a 29 day forecast period. Once again, the optimally selected

Table 3. List of "Best Buys"

STOCK NO.	STOCK NAME	ABBREV.	BETA	P/E	STARTING PRICE (8/23/82)
1	American Express	AXP	1.15	10.0	45.13
2	Edwards (A. G.) & Sons	AGE	1.50	10.5	18.63
3	American Stores Co.	ASC	.85	8.3	47.50
4	Brown Group Inc.	BGG	.85	9.2	37.75
5	Carter-Wallace	CAR	.90	8.2	13.38
6	Carling O'Keefe	CKB	1.05	7.2	6.38
7	Clorox Co.	CLX	1.00	10.0	15.25
8	First National Boston	FBB	.80	4.7	23.25
9	Financial Corp. of America	FIN	1.55	9.6	17.25
10	First. Va. Banks, Inc.	FVB	.70	5.7	8
11	Giant Foods	GFS	1.00	8.8	30.25
12	General Housewares	GHW	1.15	9.6	7.88
13	Integrated Resources	IRE	1.30	8.9	16
14	Jamesway Corp.	JMY	1.25	8.2	9.13
15	Kellwood Co.	KWD	.85	6.7	13
16	Leucadia Nat'l.Corp.	LUK	.95	9.1	12.25
17	Merrill Lynch & Co.	MER	1.70	8.3	30.88
18	Mercantile Stores	MST	.75	9.8	72.88
19	NCNB Corp.	NCB	.95	5.9	13
20	Oxford Ind.	OXM	.80	6.9	30
21	Proctor & Gamble	PGG	.75	10.9	93.75
22	Pueblo International	PII	.85	7.2	4.88
23	Paine Webber, Inc.	PWJ	1.90	8.7	20.75
24	Stop & Shop Cos.	SHP	.95	10.7	33.13
25	Security Pacific	SPC	1.00	4.7	29.88
26	Thompson Medical	TMM	1.00	9.3	15.88
27	V. F. Corp.	VFC	.75	6.9	55
28	Conair Corp.	CAC	1.20	10.8	13.50
29	Dunkin' Donuts	DUN	1.15	10.9	17.25
30	First Boston, Inc.	FBO	1.05	4.8	42.50
31	Lincoln First Banks	LFB	.70	4.5	25.25
32	Philadelphia Nat'l Corp.	PHN	.85	4.9	36.25

specimens are, on average, much more closely associated with the target population than are the 1000 randomly generated portfolios.

A graphical view of a typical cluster is shown in Figures 3 and 4. In particular, the log-transformed and mean adjusted prices (Figure 4) represent the comovements of the four stocks to a far greater degree than the absolute prices (Figure 3).

Figure 5 presents the specimen's performance to that of the S & P 500 over the entire 119 day period. The time series of five portfolios ranging in size from 2 to 6 (q) are depicted in relationship to the 33 best buy target populations. Except for the portfolio consisting of two stocks, the movements of the target population and the specimens are almost identical.

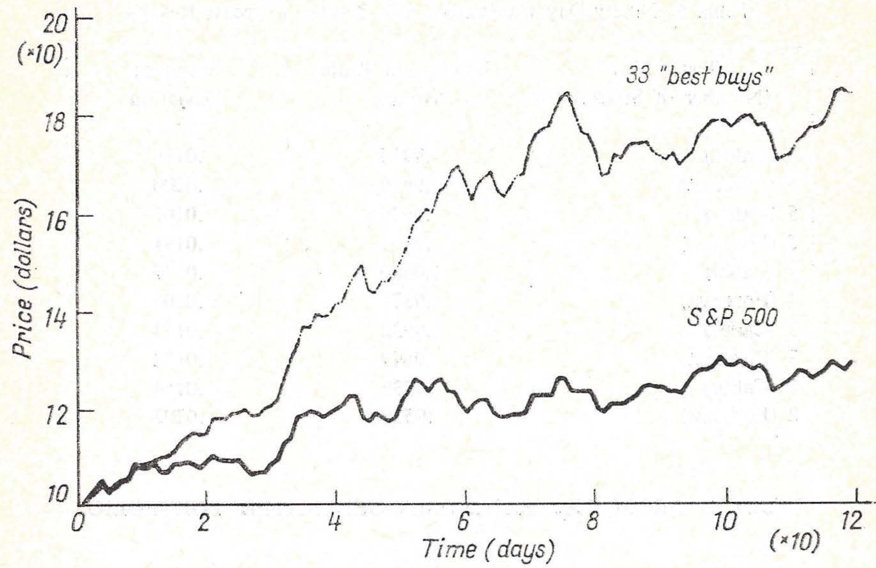


Fig. 2

Table 4. Sixty-Day Calibration and 59-Day Forecast Results

Cluster Size (Number of Stocks)	Comparison Ratio Average	Standard Deviation
6 (Calib.)	1.0010	.0119
6 (Forecast)	.9990	.0190
5 (Calib.)	1.0010	.0126
5 (Forecast)	.9900	.0183
4 (Calib.)	.9882	.0112
4 (Forecast)	1.0011	.0174
3 (Calib.)	.9572	.0131
3 (Forecast)	.9987	.0292
2 (Calib.)	.9703	.0154
2 (Forecast)	.9628	.0323

RESULTS FROM 1000 RANDOMLY-GENERATED PORTFOLIOS					
Cluster Size	6	5	4	3	2
Avg. Difference Between C. R. and 1	.0422	.0472	.0522	.0642	.0827
Standard Deviation	.0319	.0344	.0396	.0476	.0550

Table 5. Ninety-Day Calibration and 29-Day Forecast Results

Cluster Size (Number of Stocks)	Comparison Ratio Average	Standard Deviation
6 (Calib.)	.9916	.0116
6 (Forecast)	.9610	.0200
5 (Calib.)	.9929	.0106
5 (Forecast)	.9808	.0191
4 (Calib.)	1.0000	.0127
4 (Forecast)	.9871	.0205
3 (Calib.)	.9932	.0173
3 (Forecast)	.9649	.0172
2 (Calib.)	.9739	.0154
2 (Forecast)	.9530	.0207

RESULTS FROM 1000 RANDOMLY-GENERATED PORTFOLIOS

Cluster Size	6	5	4	3	2
Avg. Difference between C. R. and 1	.0430	.0464	.0548	.0632	.0783
Standard Deviation	.0316	.0344	.0389	.0420	.0523

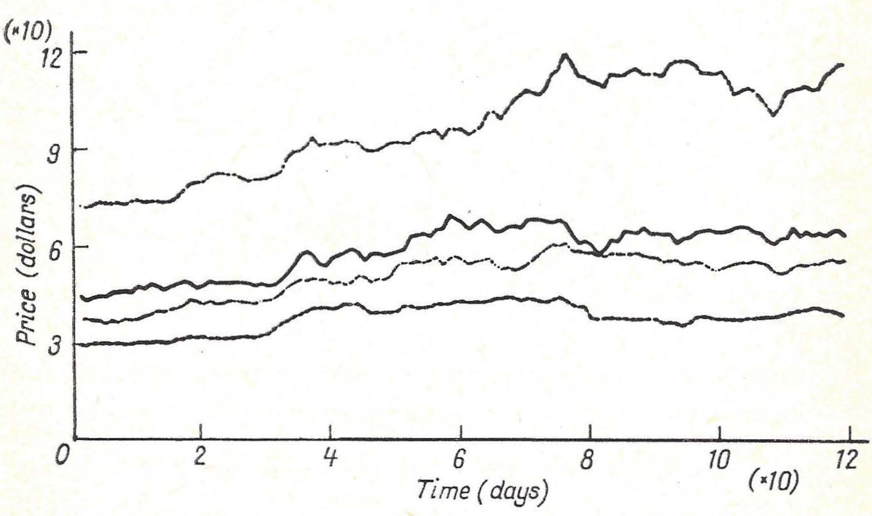


Fig. 3

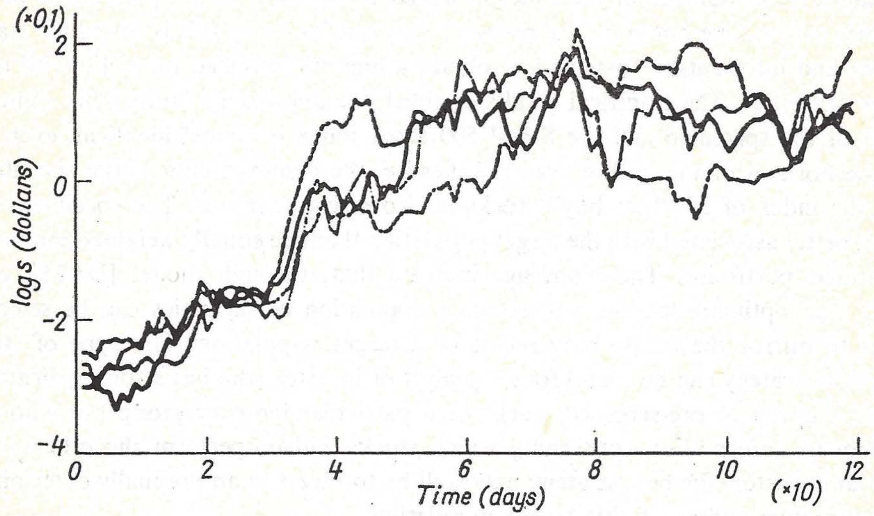


Fig. 4

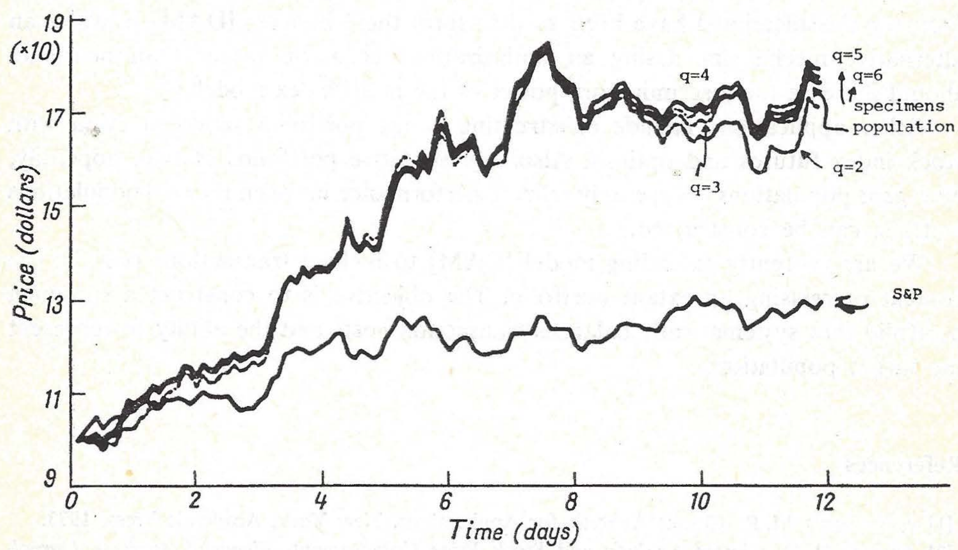


Fig. 5. Specimen and Sa.P values over 119-day period

4. Conclusions and Future Research Activities

The purpose of the data aggregation model [DAM] is to form a portfolio containing a small number of securities designed to mirror a target population of securities over a time horizon. The prime motivation for doing this is to reduce transaction and information costs and to obtain a highly diversified portfolio with only a few securities. The empirical results establish the approach's utility. The comovement of the specimen and the S & P 500 stock index is almost identical, even for a time horizon of up to three years. Likewise, the comovements of the specimens and the index of 33 "best buy" stocks is also nearly identical. The specimens are much better associated with the target population than are equally weighted randomly generated portfolios. These findings indicate that, through model [DAM], it is likely that optimally chosen subsets of a population of securities can be selected so as to mirror the future movements of a target population. This type of stock selection strategy can be useful for an analyst or investor who has strong indications about a group of pre-screened stocks or a particular industry group but who has inadequate information concerning which stocks will outperform the others. For such an investor, the best strategy may well be to invest in an optimally determined representative subset of this target population.

There are several obvious applications to the ideas presented here. One is the identification of indices for a multi-index portfolio-selection model Cohen, Zinburg and Zeikal (1977), Markowitz and Perold (1981). Largely, heuristic method (e.g., industry classifications) have been used to form these indices; [DAM] provides an alternative mechanism. Using an optimization technique rather than heuristics should increase the discriminatory power of the multi-index model.

Other applications include constructing hedge portfolios in conjunction with stock index futures and options. Also representative portfolios of two, hopefully, divergent populations, where only relative performance between the two populations matters, can be constructed.

We are currently extending model [DAM] to include transactions costs in the context of revising an extant portfolio. The objective is to construct a specimen portfolio that systematically balances transaction costs and the ability to represent the target population.

References

[1] Anderberg M. R. Cluster Analysis for Applications. New York, Academic Press, 1973.

[2] Arnott R. D. Cluster Analysis and Stock Price Comovement. *Financial Analysts Journal*, (November–December 1980), pp. 56–62.

[3] Babson D. L. Index Funds: Why Throw in the Towel? *Journal of Portfolio Management*, 1976, pp. 53–55.

[4] Black F. Yes, Virginia, There is Hope: Tests of the Value Line Ranking System. In: How to Use the value Line Investment Survey, A Subscriber's Guide, New York: Bernard & Co., Inc., 1975, pp. 20–23.

[5] COCHRAN W. G. Sampling Techniques. John Wiley & Sons New York, 1977.

[6] COHEN J. B., ZINBURG E. D., ZEIKAL A. Investment Analysis and Portfolio Management. Homewood, IL, Richard D. Irwin, Inc., 1977.

[7] COHEN K. J., POGUE J. A. An Empirical Evaluation of Alternative Portfolio-Selection Models. *Journal of Business*, 1967, pp. 166–193.

[8] EHRBAR A. F. Index Funds — An Idea Whose Time is Coming. *Fortune* 1976, pp. 145–154.

[9] FARRELL J. L. Jr. Analyzing Covariation of Returns to Determine Homogeneous Stock Groupings. *Journal of Business*, April 1974, pp. 186–207.

[10] GEOFFRION A. M. Elements of Large-Scale Mathematical Programming. *Management Science* **16** (11) 1970, 652–691.

[11] GEOFFRION A. M., MCBRIDE R. C. Lagrangian Relaxation Applied to Capacitated Facility Location Problems. *AIIE Transactions*, **10** (1978), 40–47.

[12] HAMMER A. R. Small Companies Score Best. *The New York Times*, April 4, 1983, pp. D1-4.

[13] HELD M., WOLFE P., CROWDER H. P. Validation of Subgradient Optimization. *Mathematical Programming*, **6** (1974), 62–88.

[14] LINTNER J. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *The Review of Economics and Statistics*, **47** (1965), 13–37.

[15] MARKOWITZ H. M. Portfolio Selection, Efficient Diversification of Investments. Cowles Foundation Monograph 16, Yale University Press, 1959.

[16] MARKOWITZ H. M., PEROLD A. F. Portfolio Analysis with Factors and Scenarios. *Journal of Finance*, **36** (1981), 871–877.

[17] MOSTELLER F., TUKEY J. W. Data Analysis and Regression Addison-Wesley, 1978.

[18] MULVEY J. M., CROWDER H. P. Cluster Analysis: An Application of Lagrangian Relaxation *Management Science*, **25** (1979), 329–340.

[19] MULVEY J. M. Reducing the U. S. Treasury's Taxpayer Database by Optimization *Interfaces*, **10** (1980) 5, 101–108.

[20] ORCUTT G., WATTS H. W., EDWARDS J. B. Data Aggregation and Information Loss *American Economic Review*, **58**, September 1968, pp. 773–787.

[21] PEROLD A. F. Large Scale Portfolio Optimization Harvard Business School Working Paper No. HBS-82-24, 1981.

[22] ROSENBERG B. Extra-Market Components of Covariance in Security Returns *Journal of Financial and Quantitative Analysis*, 1974, pp. 263–274.

[23] SHARPE W. F. A Linear Programming Algorithm for Mutual Fund Portfolio Selection *Managament Science*, **7** (1967), 499–510.

[24] SHARPE W. F. Portfolio Theory and Capital Markets New York. McGraw-Hill, 1970.

[25] ZIPKIN P. H. Bounds for Row-Aggregation in Linear Programming *Operations Research*, **28** (1980), 903–916.

Wybór optymalnych wzorców populacji rynkowych

Zaproponowano systematyczne podejście do wyboru reprezentatywnego podzbioru (zwanego „wzorcem") identyfikowanej, zmiennej w czasie populacji rynkowej. Procedura jest oparta na modelu programowania matematycznego, który minimalizuje straty informacji między populacją docelową a wzorcem. Pokazano, że dla dwóch różnych populacji i szeregu niezależnych okresów przewidywane zmiany optymalnego wzorca oraz populacji są prawie identyczne — ogólnie z dokładnością do 1%.

Нахождение оптимальных образцов рыночных популяций

Предложен систематический подход для нахождения представительной подсовокупности (называемой „образцом") идентифицированной, переменной во времени рыночной популяции. Метод базирует на модели линейного програмирования, котора минимизирует потери информации между целевой совокупностю и образцом. Показано, что для двух различных совокупностей и ряда независимых периодов, предвидены изменения оптимального образца и совокупности, почти идентичны — в общем с точностю до 1%.