5

*Stock Market Indices: A Principal Components Analysis**

GEORGE J. FEENEY and DONALD D. HESTER

This chapter investigates a widely quoted stock market index, the Dow Jones Industrial Average (hereafter DJI), and constructs some alternative indices. Their performances are compared to the DJI. The question of applying the indices to problems of portfolio selection is explored when investors' utility functions are quadratic in the rate of return. By constructing indices from data collected in different time periods, some conclusions are drawn about the constancy of price and rate of return covariance and correlation matrices of the 30 Dow Jones industrial stocks over time.

I

In evaluating indices it is necessary to examine the purpose for which an index is to be used. As investors are concerned with earning a high rate of return, it is perplexing to find that the most quoted indices are price indices, not indices of rate of return.[1] Investors of course are concerned with price

---

* The authors are respectively employed by the General Electric Company and Yale University. We are indebted to the Cowles Foundation and the RAND Corporation for supporting this research and to James Keaton for research assistance. In part this paper reports on research carried out under a grant from the National Science Foundation. The paper was presented at the TIMS meeting in Pittsburgh, March, 1964.

[1] Specifically, the rate of return is defined to be a stock's quarterly dividend plus change in price over a quarter divided by the stock's previous quarter closing price, all figures adjusted for stock dividends or splits.

appreciation which serves to increase the rate of return. Perhaps more plausible explanations for emphasis on prices derive from low marginal rates of taxation of capital gains and from interest on the part of brokerage houses in encouraging trading commissions. This chapter reports indices of both prices and rates of return.

What do investors seek from a stock market index? One possibility is that investors desire to know the values of their portfolios. In this view the best index will have weights equal to the percentage of an investor's portfolio in each stock; a second portfolio requires a second index. Unless investors agree to restrict their asset bundles to a small set of alternatives, there is little hope of constructing an optimal set of weights.

A second possibility is that investors desire information about alternatives to their portfolio or "the market." This information may be viewed as a norm by which to evaluate the performance of their portfolios. If, in addition, investors implicitly agree to view a set of common stocks as describing the market, a basis exists for constructing an optimal price index. No pretense of sampling is made. For purposes of the present paper we adopt this interpretation.

Historians may have good reason for specifying *a priori* a set of weights to apply to their ideal index. For example, Cowles Commission indices weight stock prices by the volume of shares outstanding in order to portray the experience of the representative investor at different points in time.[2] The present purpose of constructing a norm makes no appeal to such arguments; *a priori* weights are not of interest. The weights will be seen instead to depend on which of a number of naive theories best describes investors' desired information about the market.

Stock price indices have a notoriously unsophisticated past. Failures to adjust for stock splits, stock dividends, and warrants, naive sampling schemes, and highly arbitrary and undefended sets of weights are conspicuous examples.[3] The DJI is an excellent representative of this tradition. The index may be thought of as a weighted average of prices of 30 widely held common stocks. All stock prices are weighted equally; weights are changed only when (*a*) a large stock dividend or split occurs or (*b*) a new stock is substituted for one of the 30. The principle for determining weights is simply continuity; if a split occurs, the weight of each price increases so that the previous day's closing DJI will be unchanged if the implied post-split price is used.[4] Of course, many other sets of weights exist which will

[2] Alfred Cowles, *Common Stock Indices*, 1871–1937, Bloomington: Principia Press, Inc., 1938.

[3] *Op. cit.*, pp. 33–50.

[4] *Basis of Calculation of the Dow-Jones Averages*, a short description published by Dow Jones and Co., Inc., 1963.

assure continuity. The defense for the present procedure appears to be computational simplicity. In the day of the computer, do such arguments apply?

The Dow Jones' weighting scheme implies that when a stock splits it is of less consequence to investors; at least it will be given a smaller voice in the average. If stocks split only when they are rising relative to other stocks, the DJI will exhibit a downward bias. While we know of no investigation of corporate share splitting policy, this appears to be the practice. Further, by ignoring minor stock dividends, the average incorporates a downward bias over time. A more sophisticated stock price index seems to be called for.

What criteria should be employed in constructing an index? To develop these criteria consider a one dimensional index a. In this index investors are believed to measure the market by movements in stock prices (adjusted only for splits and stock dividends) or rates of return. An index which reports this information is assumed to be a (linear) combination of, for example, adjusted prices. The index will be most sensitive (informative) if weights are assigned in a way which captures the maximum variance of the set of reference stock prices over time. An algebraic technique for obtaining an index with this property is the extraction from a covariance matrix of stock prices of the largest characteristic root, or the principal component.[5] The value of its associated vector, normalized in some arbitrary manner, is the index $\bar{a}$. Symbolically, investors wish an index which reports the maximum variance in some $(N \times I)$ column vector of N stock prices X_t over time. For T periods, successive vectors define a $(N \times T)$ matrix M. The largest λ solving the determinantal equation $|MM' - \lambda I| = 0$ is the variance of the index $\bar{a} = \sum \bar{\alpha}_i x_{it}$ when $\sum \bar{\alpha}_i^2 = 1$. For purposes of exposition, here the normalization is $\sum \alpha_i^2 = \lambda$. This normalization implies that the resulting α_i's are the correlation coefficients between index a and the corresponding stock price or rate of return.

A number of objections can be raised against index a. Suppose two stocks each fall one point. An investor may well not regard these two pieces of news as being equally informative. Perhaps one of the stocks is high priced; a one point reduction may reflect considerably different

[5] For discussion see: (1) H. Theil, "Best Linear Index Numbers of Prices and Quantities," *Econometrica*, Vol. 28, No. 2 (April 1960), pp. 464–480, (2) M. G. Kendall, *A Course in Multivariate Analysis* (New York: Hafner Publishing Co., 1957), pp. 10–36, (3) T. W. Anderson, "Asymptotic Theory for Principal Component Analysis," *Annals of Mathematical Statistics*, Vol. 34, pp. 122–148, and (4) H. Hotelling, "Analysis of a Complex of Statistical Variables in Principal Components," *Journal of Educational Psychology*, September–October, 1933.

percentage declines in the two stock prices. If investors are concerned with percentage declines, rates of return are the best representation of stocks; they are studied below. An alternative to rates of return is a component analysis of logarithms of stock prices.[6]

Second, one of the stocks may have a much larger variance in its price movements; an investor may subjectively adjust for this stock's behavior when evaluating the informational content of the market news. We might therefore transform the stock price into a standardized variate with zero mean and unit standard deviation. In terms of the discussion above we have another time dated column vector of stocks Y_t having elements related to their corresponding elements in X_t by the simple rule

$$y_{it} = \frac{x_{it} - \bar{x}_i}{s_i}$$

where $\bar{x}_i$ is the stock's mean price and s_i is an estimate of its standard deviation. Index *b* is constructed by extracting characteristic roots from the $N \times N$ correlation matrix corresponding to MM'.

Third, is the investor likely to view equal positive and negative movements of a stock price, standardized or raw, as being equally informative? In view of the long history of rising stock prices, we believe downward movements are likely to be recorded as more informative. More precisely, investors may extrapolate a past price trend and view equal movements of standardized variables about this trend as conveying equivalent amounts of information. That is, an index *c* is constructed by the technique described above from a vector Z_t having element

$$z_{it} = \frac{x_{it} - \bar{x}_i}{s_i} - bt$$

In each of these three indices we make a guess about what conveys the most information to decision makers, the investors. We know of no study which allows an investigator to consider any one of them as most useful, although such research is elemental in the construction of an optimal index. This problem (having to guess) can be viewed differently. Investors have a set of beliefs at time $\hat{t}$ about the future stream of returns from the market. How, if at all, are investors' beliefs changed by having perceived certain price movements in an interval after $\hat{t}$? If the set of beliefs remain unchanged, we infer that price movements in the interval conveyed no

[6] The *Value Line* 1100 *Stock Average* is a geometric average of stock prices.

information.[7] What transformation of stock prices best describes changes in investor beliefs?

Fourth, in order that indices constructed in this framework be equally informative in different time periods; it is necessary that investors assume that the "structure" of stock prices remains constant over time; each index makes a different, but obvious, assumption about the structure of prices. This means that similar movements in the index at different dates cause the same change in investor beliefs. Alternatively, investors may derive information only from movements in linear combinations of stock prices; deviations in the observed correlations of prices are strictly ignored. In Section IV some evidence about the stationarity of the relevant correlation and covariance matrices is presented.

Fifth, a stock price may move, but contain no information about the market. This would happen if a conspicuous explanation for the movement exists, e.g., news of an impending strike or failure of a company sponsored legal action.[8]

Finally, a stock price movement may result from a stock's going ex dividend. This movement clearly conveys no information; it will not affect rates of return.

The criterion for constructing an ideal index is, therefore, not surprisingly the information which investors plan to derive from it. Indices are estimated by extracting characteristic roots from an appropriate covariance matrix or more familiarly by the method of principal components.

One question remains; precisely how many numbers should be used to describe the movements of a set of N stocks? So far we have spoken about only one number, the value of a vector associated with the largest λ; this is a one dimensional representation of stock prices. Most price indices are one dimensional. If individuals do subjectively represent information in more than one dimension, there is reason to report the value of vectors associated with other λ's. Of course, if individuals subjectively represent information in N dimensions there is little point in constructing an index.

---

[7] The reader should recognize that movements of stock prices may convey no information. In this case, beliefs are a function of income statements and other data collected from outside the stock market. This suggests that descriptions of stock market action permit one to say nothing about future stock prices. The observable similarity of opening and closing prices of stocks would appear to contradict this view; surely one can make probability statements about the range of a stock's closing price on the basis of its opening that morning.

[8] In a factor analysis of a set of stock prices, adjustment for this fifth consideration is the estimation of communalities. In future studies, perhaps independent estimates of communalities can be constructed. Research in progress by J. Bossons at the Carnegie Institute of Technology may provide the raw material for this further extension of the present study.

We suggest that this is highly unlikely; investors can evaluate only a very few numbers at one time. In the next section indices *a*, *b*, and *c* are each reported in two dimensions.

II

Data utilized in this study were collected from standard sources, *The New York Times*, *The Wall Street Journal*, *Standard and Poor's Stock Guide*, etc. They include cash and stock dividends, end of quarter closing prices, and stock splits for the 30 stocks in the DJI on December 31, 1961. The period covered includes 50 consecutive quarters subsequent to December 31, 1950.[9]

The data were adjusted for stock dividends and splits prior to calculations. Thus, if a stock splits 3 for 1, its price is weighted in the postsplit period by the factor 3, etc. Individual stocks were normalized, so that after adjustment their second quarter 1963 price used in the calculations was the observed second quarter price.

Table 1 reports weights estimated for the three stock price indices from data from all 50 quarters. Index *a* computed from raw price data is perhaps the closest to the spirit of the DJI. Its first component (dimension) accounts for 76 per cent of the generalized variance of the 30 stocks; its second component explains 14 per cent.[10] Not surprisingly, stocks which were high priced in 1963 tend to be weighted heavily, i.e., duPont, General Foods, AT & T, Eastman Kodak, Owens-Illinois, and Union Carbide. Other securities which for various reasons are highly erratic during the period also carry large weights, e.g., U.S. Steel and General Electric. While the DJI makes no effort to exploit the observed variance of individual stocks, it is heavily influenced by high price stocks as is index *a*. A comparison of the weights of index *a* with those of the DJI is not illuminating; the latter are time subscripted. Index *a* has been normalized to have the same mean and variance as the DJI in Figures 1 and 2. The high correlation between the first component of *a* and the DJI in Figure 1 confirms the suspicion that their similar emphasis on high price securities makes the two very similar. To put it differently, the DJI weighting scheme

[9] Stock substitutions were made in the DJI stocks during the 1950's; no adjustment is attempted in the present study. End of quarter price quotations refer to closing prices on the last trading day in March, June, and September. Fourth quarter price quotations are the closing prices on the last trading day prior to Christmas. This different treatment is intended to avoid noise resulting from investors realizing capital gains or losses before the year's end.

[10] "Generalized variance" refers to the trace of the corresponding covariance or correlation matrix.

Table 1 Weights for First Two Components of Price Indices *a*, *b*, and *c*.

	a		*b*		*c*	
	(1)	(2)	(1)	(2)	(1)	(2)
Allied Chemical	0.137	0.039	0.820	−0.251	0.837	−0.049
Alcoa	0.405	0.308	0.757	−0.596	0.424	−0.851
American Can	0.067	0.035	0.647	−0.399	0.895	0.340
AT & T	0.361	−0.272	0.800	0.559	−0.773	0.013
American Tobacco	0.123	−0.081	0.792	0.498	0.485	0.487
Anaconda	0.101	0.168	0.423	−0.734	0.911	−0.164
Bethlehem	0.217	0.104	0.849	−0.416	0.761	−0.566
Chrysler	−0.042	0.025	−0.305	−0.199	0.765	0.493
duPont	0.951	0.217	0.931	−0.235	−0.255	−0.865
Eastman Kodak	0.587	−0.231	0.900	0.334	−0.779	−0.505
General Electric	0.376	0.047	0.932	−0.141	−0.015	−0.896
General Foods	0.391	−0.254	0.841	0.511	−0.841	−0.126
General Motors	0.236	0.009	0.922	−0.080	0.510	−0.405
Goodyear	0.222	−0.012	0.958	0.044	0.562	−0.453
International Harvester	0.123	−0.051	0.836	0.326	0.636	0.390
International Nickel	0.297	−0.061	0.927	0.159	−0.123	−0.465
International Paper	0.156	0.054	0.917	−0.335	0.943	−0.156
Johns-Manville	0.150	−0.004	0.872	−0.001	0.732	0.008
Owens-Illinois	0.374	−0.012	0.943	0.025	−0.104	−0.811
Procter and Gamble	0.342	−0.238	0.823	0.532	−0.733	−0.030
Sears	0.319	−0.185	0.845	0.455	−0.694	−0.030
Standard Oil (Cal.)	0.195	0.007	0.863	−0.089	0.646	−0.075
Esso	0.198	0.068	0.793	−0.334	0.680	−0.270
Swift	0.037	0.019	0.344	−0.194	0.813	0.360
Texaco	0.251	−0.090	0.889	0.271	−0.371	−0.024
Union Carbide	0.422	0.131	0.903	−0.286	0.193	−0.910
United Aircraft	0.186	0.155	0.636	−0.592	0.721	−0.321
U.S. Steel	0.407	0.136	0.866	−0.283	0.207	−0.886
Westinghouse	0.145	0.004	0.670	−0.006	0.585	−0.016
Woolworth	0.175	−0.124	0.786	0.537	0.117	0.348
Percent variance	75.76	13.93	65.67	13.70	39.90	23.20

is a very informative index if investors are concerned with raw stock prices; virtually it is the best price index available!

Index *b* is estimated from standardized stock prices and thus will not necessarily accord high weights to stocks simply because they carry a high price tag. Inspection of the weights suggests that time trends dominate the weights. Stocks which had a low rate of price appreciation during the twelve year period, Chrysler, Swift, United Aircraft, American Can, Anaconda, and Westinghouse, have low weights. All other stocks are weighted more or less equally. The first component accounts for 66 per cent of the generalized variance. Figures 3 and 4 show plots of index *b*

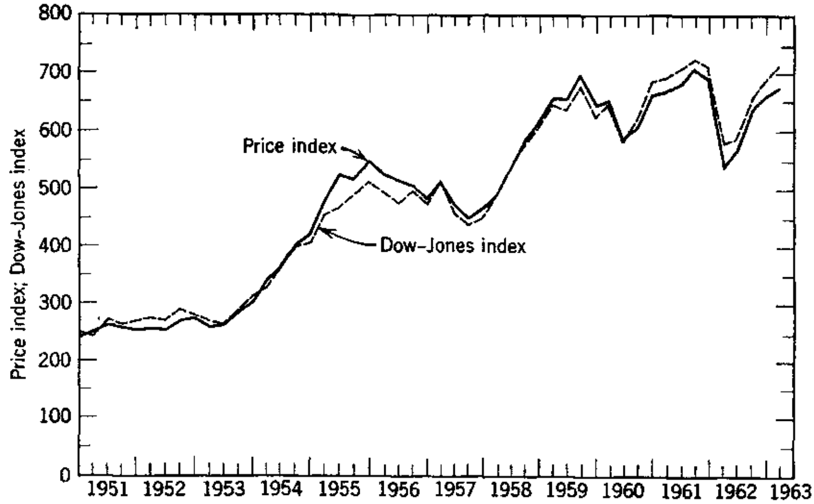

Figure 1 Index *a* component 1, $R = .992$. R denotes the correlation between the component and the Dow-Jones Index.

against the DJI. It is a bit surprising to find that the largest component of the normalized index *b* also correlates highly with the DJI.

One would expect differences between index *b* and index *a* (and the DJI) unless variances of individual stock prices are identical. Variances are not identical; Table 2 reports means and variances of end of quarter adjusted prices and of quarterly rates of return of the 30 DJI stocks over the 50

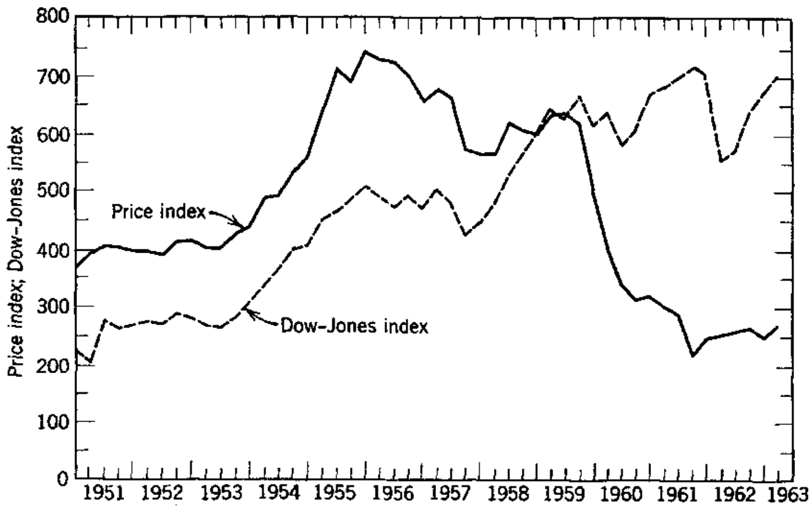

Figure 2 Index *a* component 2, $R = -.111$.

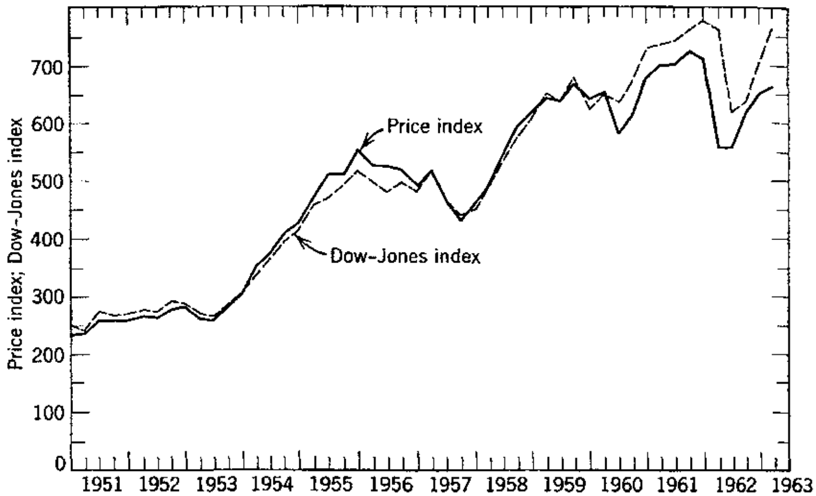

Figure 3 Index b component 1, $R = .994$.

quarter period. The trend in end of quarter prices is also reported in this table. The weights of index a and b differ greatly but the principal component of each is highly correlated with the DJI. It is tempting to conclude that a large number of positive weighted indices would correlate well with the DJI; nevertheless, for the period studied, the DJI is again nearly a best price index for individuals who subjectively adjust for stock variances.

The reason that both first components are positive (except Chrysler) is the strong positive trend in stock prices over the 50 quarter period. As

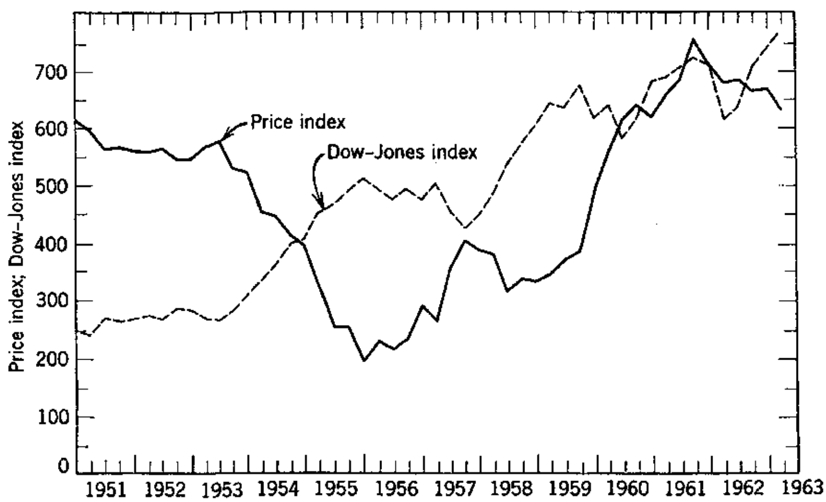

Figure 4 Index b component 2, $R = .083$.

Table 2 Means, Standard Deviations, and Trends of Adjusted End of Quarter Prices and Quarterly Rates of Return

	$\bar{p}(\$)$	$\sigma_p(\$)$	$\frac{dP}{dt}\left(\frac{\$}{\text{quarter}}\right)$	$\bar{R}(\%)$	$\sigma_R(\%)$
Allied Chemical	45.12	9.47	0.387	2.34	9.09
Alcoa	62.32	29.55	1.163	4.75	14.68
American Can	40.62	5.83	0.238	2.37	7.37
AT & T	73.62	25.97	1.557	3.11	6.19
American Tobacco	23.05	9.05	0.503	2.96	8.46
Anaconda	51.65	13.26	0.114	2.79	12.41
Bethlehem	33.04	14.28	0.687	3.90	11.77
Chrysler	33.46	8.40	−0.198	3.12	14.22
duPont	177.44	56.28	3.182	3.60	8.64
Eastman Kodak	59.21	36.70	2.374	5.00	9.78
General Electric	55.38	22.50	1.359	4.52	11.39
General Foods	37.13	26.59	1.670	5.38	8.24
General Motors	38.71	14.27	0.879	4.77	10.17
Goodyear	23.55	13.11	0.819	5.73	12.78
International Harvester	39.12	8.40	0.466	2.81	8.42
International Nickel	42.46	18.33	1.127	4.23	10.18
International Paper	26.02	9.50	0.500	3.50	10.74
Johns-Manville	44.54	9.99	0.505	2.70	9.25
Owens-Illinois	66.67	22.33	1.315	2.82	8.54
Procter and Gamble	38.02	23.95	1.451	4.23	8.89
Sears	38.93	21.45	1.341	4.66	9.44
Standard Oil (Cal.)	39.18	12.65	0.771	3.97	8.39
Esso	42.69	13.91	0.740	4.51	8.31
Swift	40.02	6.04	0.066	1.96	8.32
Texaco	31.50	15.96	1.054	5.41	8.77
Union Carbide	100.25	26.23	1.322	2.50	9.12
United Aircraft	42.72	16.32	0.609	4.42	15.22
U.S. Steel	53.39	26.39	1.283	3.84	12.94
Westinghouse	33.55	12.12	0.480	3.10	14.61
Woolworth	53.49	12.86	0.698	2.39	8.22

economists are all too frequently aware, time series variables or indices with trends tend to be highly correlated. The high correlations between first components of indices a and b and the DJI probably are mostly owing to this trend.

The second components of a and b and the first component of c also have an interpretation. As the components are unique only up to linear transformations we may reverse the signs of the second component of index *b* without loss of information. Then the reader may verify that stocks in all three components having substantial negative signs are AT & T, Eastman Kodak, General Foods, Procter and Gamble, and Sears. Woolworth has large negative signs in two of the components. These are perhaps the most consumer oriented of the Dow Jones stocks; apparently these components discriminate between producer and consumer goods industrial stocks. An explanation for this might be that profits of producer goods firms and consumer goods firms reach peaks at different points in a business cycle. A simple accelerator model might yield such a result. In

Table 3 Weights for Rate of Return Indices *a*, *b*, and *c*

	a		*b*		*c*	
	(1)	(2)	(1)	(2)	(1)	(2)
Allied Chemical	0.455	−0.095	0.783	0.041	0.780	0.046
Alcoa	0.663	0.023	0.628	0.369	0.599	0.404
American Can	0.185	−0.044	0.422	0.026	0.414	−0.001
AT & T	0.184	−0.108	0.529	−0.457	0.538	−0.473
American Tobacco	0.176	0.097	0.373	−0.581	0.375	−0.591
Anaconda	0.648	0.096	0.776	0.205	0.773	0.215
Bethlehem	0.597	0.151	0.723	0.200	0.706	0.225
Chrysler	0.478	0.264	0.466	0.158	0.494	0.130
duPont	0.438	0.0	0.780	0.159	0.780	0.153
Eastman Kodak	0.399	−0.164	0.636	−0.038	0.633	−0.033
General Electric	0.475	−0.322	0.619	0.106	0.605	0.121
General Foods	0.217	−0.108	0.455	−0.573	0.462	−0.575
General Motors	0.489	0.221	0.715	0.207	0.716	0.198
Goodyear	0.637	−0.076	0.748	0.076	0.732	0.100
International Harvester	0.389	0.039	0.712	−0.047	0.724	−0.061
International Nickel	0.491	0.076	0.736	−0.082	0.730	−0.071
International Paper	0.548	−0.117	0.793	0.142	0.785	0.153
Johns-Manville	0.385	−0.097	0.646	−0.141	0.638	−0.133
Owens-Illinois	0.360	−0.097	0.685	−0.267	0.682	−0.264
Procter and Gamble	0.313	−0.105	0.580	−0.589	0.592	−0.582
Sears	0.378	0.046	0.645	−0.368	0.658	−0.372
Standard Oil (Cal.)	0.302	0.032	0.579	0.405	0.576	0.381
Esso	0.305	0.052	0.557	0.586	0.544	0.576
Swift	0.306	−0.161	0.585	−0.156	0.583	−0.165
Texaco	0.411	−0.020	0.744	0.206	0.745	0.193
Union Carbide	0.477	−0.102	0.798	0.131	0.790	0.145
United Aircraft	0.398	0.651	0.335	0.065	0.313	0.075
U.S. Steel	0.738	0.097	0.842	0.072	0.830	0.099
Westinghouse	0.348	−0.600	0.338	−0.092	0.331	−0.087
Woolworth	0.251	−0.002	0.530	−0.461	0.538	−0.470
Percent total	41.23	8.67	41.10	8.77	40.69	8.92

order for this explanation to hold, it must also be assumed that stock prices are closely related to current earnings.

That trend is largely responsible for the signs of weights in the first components of *a* and *b* can be seen by examining the largest component of index *c* which is adjusted for trend. One third of its weights are negative and the component accounts for only 40 per cent of the generalized variance. Allied Chemical, American Can, Anaconda, Bethlehem, Chrysler, International Paper and Swift enter the index with strong positive weights. These securities would seem to support only weakly the accelerator argument suggested above. Perhaps in the case of this component variations in individual stock trends about the market trend mask the accelerator relation. The seven securities above have a comparatively slow rate of price appreciation over the twelve year period. The consumer stocks with the exception of Woolworth tended to have much higher rates of price appreciation.

One final point should be made about price indices. In Section I, it was argued that there were reasons to expect that the DJI had a downward bias over time. Inspection of Figures 1 and 3 lend no empirical support to this view. If a downward bias exists, it is of small consequence over the recent decade.

In Table 3 the first two components of *a*, *b*, and *c* rate of return indices are reported. The first two components of index *a* account for much smaller fractions of the generalized variance, 41 per cent and 9 per cent, than was the case for prices. A major reason for this is the absence of any sharp trend in rate of return during the period under study. All stocks are positively correlated with the index; Alcoa, Anaconda, Bethlehem, Goodyear, International Paper, and U.S. Steel have the largest correlations with the component. For purposes of comparison we have constructed an index of rates of return for Dow Jones stocks.

$$RI = \frac{DJI_t - DJI_{t-1}}{DJI_{t-1}} + \sum \frac{DJI_t}{DJI_{t-1}} \cdot \frac{D_{i,t-1}}{P_{i,t-1}}$$

where:

$P_{i,t-1}$ is the adjusted price of the *i*th stock at the beginning of period $t-1$.

$D_{i,t-1}$ is the amount of dividends paid by the *i*th stock during period $t-1$.

This pseudo DJ return index is plotted against the first two components of index *a* in Figures 5 and 6. The correlation between this index and our first component is again very high, 0.931.

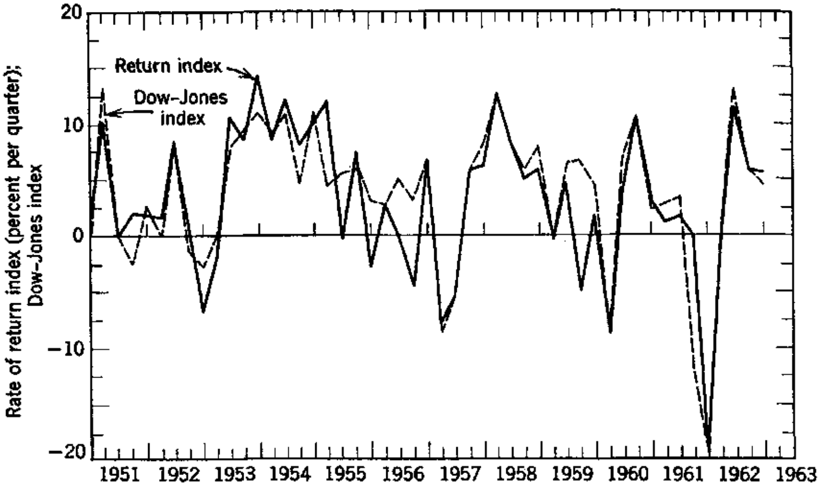

Figure 5 Index *a* component 1, $R = .931$.

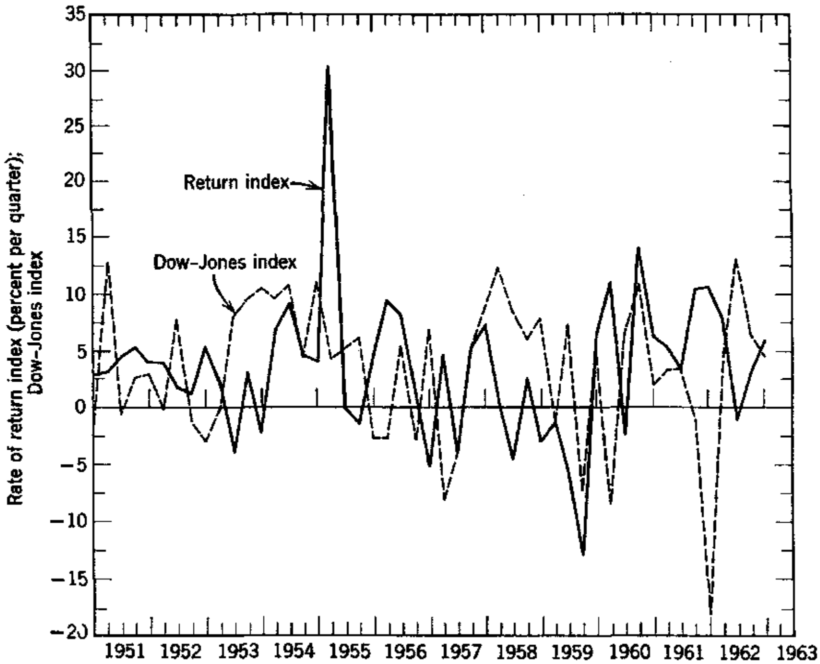

Figure 6 Index *a* component 2, $R = -.100$.

The first two components of index *b*, which refer to standardized rate of return data, account for the same amount of the generalized variance as index *a*, 41 per cent and 9 per cent respectively. All stocks are again positively correlated with the first component. Allied Chemical, Anaconda, duPont, International Paper, Union Carbide, and U.S. Steel have correlations exceeding 0.75. The correlation between the first component and the pseudo index is 0.952.

As can be seen from Table 3, the weights of index *b* and *c* are virtually identical owing to the absence of trend in rate of return.

The following comments summarize this section. The first component of stock price indices is much better in representing information about the 30 Dow Jones stock prices than the first component of the rate of return index in representing their rates of return. This is due to a strong positive postwar trend in stock prices. Once this trend is removed from stock prices the first component accounts for about 40 per cent of the generalized variance. The first component of any of the rate of return indices also accounts for about 40 per cent of the generalized variance.

The first components of the three rate of return indices are defined by positive vectors.[11] This pattern might be expected if the "attitude of the market" really is the most important determinant of stock price movements. Apparently then movements of individual stock prices are dominated by the tone of the market over the 50 quarter period. The significance of this observation for investors choosing portfolios is suggested in the next section.[12]

Finally, the DJI is highly correlated with the first components of indices *a* and *b*. In a sense it is a very good and informative index; whether it will retain its properties in periods where the strong trend is absent is hypothetical. Judging from the methods used in calculating it, an equivalent performance in a trendless world is dubious. Once trend is removed, the weighting of stocks in the largest component bears very little resemblance to those used in calculating the DJI. No evidence of a downward bias in the DJI was detected.

---

[11] Of course components are only unique up to a multiplicative constant, which may be positive or negative. A more accurate description of the vectors would be that they are isosign; that is, their elements are either nonnegative or nonpositive. Rather than apply novel terminology we suffer the inaccuracy in the text. It is trivial to demonstrate that a correlation matrix which contains no negative elements has a largest characteristic vector which is isosign. Other correlation matrices, like the matrix of Dow Jones stock rates of return, which have a few small negative elements may also have isosign largest characteristic vectors if their ranks exceed two.

[12] A similar observation has been made by K. Borch, "Price Movements in the Stock Market," mimeographed Econometric Research Program Paper No. 7, Princeton University, April 30, 1963.

III

In this section the relation between the computed indices and the problem of portfolio selection is examined. It is assumed that the utility function of investors is quadratic in the rate of return, i.e.,

$$u = f(r, \sigma_r^2), \qquad \frac{\partial u}{\partial r} > 0, \qquad \frac{\partial u}{\partial \sigma_r^2} < 0.$$ [13]

It is convenient also to view the 30 Dow Jones industrials as constituting the universe of available assets.

Inspection of the utility function indicates that rate of return, not price, components will be of interest in this section. However, owing to higher marginal rates of taxation on dividend income than capital gains, the utility function of high marginal tax rate investors is probably better viewed as a quadratic function of $(P_t - P_{t-1})/P_{t-1}$. The reader is cautioned that we have not used this argument in our calculations; strictly speaking, our conclusions apply only to investors whose behavior is independent of the tax structure.

Investors will be concerned with the components of covariance matrices; they will find little of interest in the components of correlation matrices, indices b and c. This is because components are not invariant to change in units of variables, except in the trivial case of a scalar multiplication of elements of a covariance matrix. Consequently this section refers only to components of index a.

It is worth noting that these elementary observations have been totally ignored in a recently published study of portfolio behavior of mutual funds. We restrict comments to Farrar's empirical work.[14]

In Chapter 2, he correctly specifies Markowitz's utility function.[15] In Chapter 3 he states:

> In order to reduce the population to feasible dimensions, therefore, a good deal of aggregation is immediately necessary. Despite its obvious shortcomings, the absence of a practicable alternative serves to justify aggregation according to industry and asset groupings. The presence of convenient and reliable Standard and Poor's price indices for such classifications is also a non-negligible factor in the choice.[16]

---

[13] Portfolio selection has been analyzed previously in this framework by H. M. Markowitz, *Portfolio Selection*, New York: John Wiley and Sons, Inc., 1959, and J. Tobin, "Liquidity Preference as Behavior Towards Risk," this volume, Chapter 1.

[14] D. E. Farrar, *The Investment Decision Under Uncertainty* (Englewood Cliffs: Prentice-Hall, Inc., 1962), Chs. 2–4.

[15] *Ibid.*, p. 27.

[16] *Ibid.*, p. 39.

Using monthly industrial stock price indices he extracts principal components from a 47 × 47 matrix of correlation coefficients. Although he professes to be concerned with reducing "redundancy" or "collinearity," a legitimate objective of principal components analysis, Farrar inappropriately applies factor analytic techniques which assume the existence of a particular *model.*[17] He finds that the 47 × 47 correlation matrix of indices is nearly represented by three nonorthogonal vectors. These vectors along with other pseudo assets or assets are assigned rates of change of prices based on a set of expectations assigned to the 23 mutual funds studied. Using a quadratic programming algorithm, Farrar estimates a risk return efficiency locus and concludes that the funds are behaving reasonably efficiently.[18]

Three points deserve mention. First, it is difficult to justify making a factor analysis of stock prices, rather than rates of return or at least percentage change in stock prices. Farrar has used a different variable than the rate of return, the argument of Markowitz's quadratic utility function. Second, when estimating principal components, he extracted roots from the correlation, not the covariance, matrix. It seems likely that considerable differences in the variance of different industrial stock indices exist, to say nothing about stocks used in constructing the indices. For both reasons Farrar's vectors are not appropriate for studying the Markowitz model.

Finally, the power of Farrar's evaluation of the efficiency of mutual fund portfolios appears to be extremely low; for common stocks nearly all the variance and the highest return is associated with his largest industrial factor in which most widely held common stock groups receive a heavy weight. All funds' portfolios lie relatively close to his estimated efficiency frontier because all of the following are nearly on it and on a straight line from the origin of the risk-return axes; cash, the two largest bond components, a utility stock index, and the largest common stock factor. These assets or pseudo assets probably account for over 90 per cent of the investments eligible for mutual funds. Farrar's random portfolio selection test does not take this last consideration into account.

Returning to the main argument, there are two important reasons for applying quadratic utility functions to problems of portfolio selection: (1) diversification is a reasonable form of behavior and (2) the problem, in principle, can be solved. Quadratic programming routines exist for large digital computers; these routines with a set of expected rates of return and a known covariance matrix should reduce the problem of

---

[17] Kendall, *op. cit.*, p. 37.

[18] Markowitz, *op. cit.*, p. 20.

portfolio selection to an uninteresting manipulation. Why then should this problem be reconsidered here?

If risk avoidance is very important to investors an illuminating representation of a perceived covariance matrix is simply its set of characteristic roots and vectors. Elsewhere one of us has argued that, in a risky world, if the covariance matrix is singular and short sales are permitted, then a riskless portfolio with nonnegative return always exists.[19] Further, even if the covariance matrix is not singular, a minimum risk portfolio is a

Table 4 A Representation of Six Securities in Three Dimensions

Security	Component 1	Component 2	Component 3	$E(R)$
x_1	0	0	0	0.05
x_2	1.00	0	0	0.08
x_3	0.25	0.25	−0.5	0.06
x_4	1.0	0	0	0.04
x_5	0.42	−0.75	0	0.04
x_6	0.5	0.5	0.25	0.05
λ	2.49	0.88	0.31	

weighted average of normalized characteristic vectors, where weights are reciprocals of the associated characteristic roots. In the present study we did not find that the covariance matrix is singular, although the smallest root was very small indeed.

Do investors really have well formulated expectations about future returns to an asset? This question cannot be studied with our data. Is a covariance matrix constant over time? This question is investigated in the next section; in the present section the answer is assumed to be affirmative. To what extent do investors perceive the covariance matrix? A covariance matrix for the Dow Jones industrials is, of course, a 30 dimensional vector space. We believe it to be highly unlikely that ordinary mortals can have very strong convictions about each of the 435 covariances among these stocks, to say nothing about the universe of available assets. They must somehow simplify their beliefs about the covariation of a set of assets. We propose that in effect investors summarize the covariance matrix in the largest few principal components.

It would be neat to demonstrate that investors actually thought in terms of components of covariance matrices; we don't try. However, investors

[19] Hester, D., "Efficient Portfolios with Short Sales and Margin Holdings," in this volume, Chapter 3.

or at least financial newspapers do reveal a set of beliefs about the structure of the covariance matrix of rates of return. We will demonstrate that these beliefs have empirical substance.

First, Table 4 suggests how a set of beliefs about six securites might be viewed in a three dimensional vector space. Table 5 reports the first five components of the rate of return covariance matrix of the 30 Dow Jones industrials.

Table 5 Weights for the Five Largest Components of Index *a*

	Component 1	Component 2	Component 3	Component 4	Component 5
Allied Chemical	0.455	−0.095	−0.038	0.015	−0.050
Alcoa	0.663	0.023	0.128	0.462	0.209
American Can	0.185	−0.044	−0.033	−0.004	−0.056
AT & T	0.184	−0.108	−0.071	−0.163	−0.037
American Tobacco	0.176	0.097	−0.242	−0.315	0.022
Anaconda	0.648	0.096	−0.090	0.108	−0.104
Bethlehem	0.597	0.151	0.025	0.074	0.094
Chrysler	0.478	0.264	0.473	−0.108	−0.527
duPont	0.438	−0.000	−0.012	0.038	−0.126
Eastman Kodak	0.399	−0.164	−0.114	0.011	0.095
General Electric	0.475	−0.322	0.146	0.092	0.149
General Foods	0.217	−0.108	−0.083	−0.266	0.036
General Motors	0.489	0.221	0.028	0.019	−0.195
Goodyear	0.637	−0.076	−0.144	0.015	0.111
International Harvester	0.389	0.039	−0.003	−0.081	−0.084
International Nickel	0.491	0.076	−0.091	−0.117	0.025
International Paper	0.548	−0.117	−0.155	0.114	−0.007
Johns-Manville	0.385	−0.097	−0.006	−0.101	0.050
Owens-Illinois	0.360	−0.097	−0.180	−0.078	0.044
Procter and Gamble	0.313	−0.105	−0.019	−0.366	0.097
Sears	0.378	0.046	−0.062	−0.268	−0.103
Standard Oil (Cal.)	0.302	0.032	−0.194	0.178	−0.136
Esso	0.305	0.052	−0.063	0.268	−0.094
Swift	0.306	−0.161	0.043	−0.087	−0.045
Texaco	0.411	−0.020	−0.150	0.093	−0.030
Union Carbide	0.477	−0.102	0.000	0.074	0.026
United Aircraft	0.398	0.651	0.330	−0.176	0.463
U.S. Steel	0.738	0.097	−0.040	−0.013	0.006
Westinghouse	0.348	−0.600	0.561	−0.138	0.063
Woolworth	0.251	−0.002	−0.216	−0.243	−0.018
Cumulative Percent of Variance	41.23	49.90	56.90	63.59	68.79

The coefficients in Table 4 are rough subjective estimates of the correlation between each of the components and the individual securities. All assets are assumed to have the indicated expected rates of return and identical variances. Perhaps the components can be interpreted as (1) the market, (2) the defense sector, and (3) the farm sector, respectively, all of which are believed to be orthogonal. Although knowledge of the structure of the covariance matrix is crude, it is not without value. A portfolio of a risk averting investor might include x_1, x_2, and x_3. Such a portfolio will capture a considerable amount of uncorrelated variation (which of course is invaluable in reducing portfolio variance!), while earning a good rate of return. If short sales are allowed, then a riskless portfolio with positive rate of return is holding x_2 and selling x_4 short. If x_3 were not available, x_6 would appear more attractive, for it would serve to capture uncorrelated variation in components 2 and 3. A more subtle investor might recognize that x_5 and either x_3 or x_6 make attractive pairs; by suitably purchasing from the set of securities much of the variance of the second component can be eliminated from the portfolio. Similarly, holding x_3 and x_6 together will serve to eliminate variance of the third component.

Assuming that investors represent the Dow Jones industrials in five dimensions we may draw some conclusions about stocks which would have been useful in reducing portfolio variance during the 50 quarter period. We shall not venture to name these components. From Table 5, one may see that stocks whose rate of return was either relatively uncorrelated with any of the components or was highly correlated with any of the last four components are attractive; most stocks were highly correlated with the first component. Likely candidates from the 30 are Alcoa, American Can, AT & T, Chrysler, United Aircraft, and Westinghouse. Whether or not these securities will prove useful for purposes of diversification in future years depends upon whether the covariance matrix is unchanging over time. Evidence concerning this issue is presented in the next section.

While we do not presume to summarize the vast literature concerning the New York Stock Exchange or the 30 Dow Jones industrials, one common theme is evident in this material: individual stocks are frequently summarized by industries. The suggestion is that rates of return of stocks in an industry are positively correlated; this is a hypothesis about the structure of the covariance matrix. It is by no means obvious that this hypothesis is valid. If GM is having a good year is that because all automotives are doing well or is it because GM is cutting into sales by Ford, American Motors, and/or Chrysler? A test of whether it is useful to aggregate into industries is whether firms in the same industry enter different components of the market with similar weights.

A null hypothesis is that the correlations of two stocks with a component will be of the same sign with probability 0.5. By examining the frequency with which a pair of stocks have the same sign of correlation coefficient with a number of different components we may test the hypothesis underlying the aggregation of stocks into industry classifications. The meaning of acceptance of the hypothesis for portfolio selection is clear. If stocks of the same industry have highly correlated rates of return, unless one of the stocks is sold short, there is no advantage to having a position in more than one. Rejection of the hypothesis implies that portfolio risk is reduced by spreading out one's automotive holdings among the different firms.

Table 6 reports the frequency of same sign component correlations for 13 pairs of stocks for arbitrarily defined industries. While the distribution

Table 6 Number of Industry Similar Signs among the Largest 5 and the Largest 10 Rate of Return Components 1951–1963

Industry and Firm	Largest 5	Largest 10
Automobiles		
Chrysler and GM	4	6
Chemicals		
duPont and Allied	5*	7
Electricals		
GE and Westinghouse	4	6
Foods		
Swift and General Foods	3	5
Nonferrous Metals		
Anaconda and Alcoa	3	6
Anaconda and International Nickel	3	8*
Alcoa and International Nickel	3	6
Oils		
Esso and Standard Oil (Cal.)	5*	9*
Esso and Texaco	4	7
Texaco and Standard Oil (Cal.)	4	5
Packaging		
American Can and Owens-Illinois	4	6
Retailing		
Sears and Woolworth	4	8*
Steels		
U.S. Steel and Bethlehem	3	8*
Actual (sum)	49	88
Expected	32.5	65
Standard deviation	4.0	5.7
	Significant	Significant

* The asterisk indicates significance at .05 in the binomial approximation.

of some signs is not binomial, a rough test can be applied by assuming that they are binomially distributed. Tests are performed roughly at the 0.05 level for each pairwise comparison and for a pooled sum of the 13 pairwise comparisons. The results appear to support the hypothesis rather impressively. In all 13 cases signs of correlations of firm stock rates of return with either the first five or the first ten components are more often similar than not. In 5 of the 13 industry groupings one or both of the pairwise comparisons were significant. The pooled sum of similar signs in pairwise comparisons is overwhelmingly significant.

Finally we note that although the "industries" representation of the market has empirical support, it may not be too useful a representation for investors. Different industries may move together; a portfolio of an automotive stock and a steel stock may not afford much greater protection against risk than an alternative consisting of GM and Ford, assuming no short sales. Alternatively, orthogonality may be a useful property to impose on representations of covariance matrices.

In summary, the Dow Jones rate of return covariance matrix is not singular; there is considerable variation about the first or market component which accounts for only 40 per cent of the generalized variance. The variance explained by the next four components individually is 9%, 7%, 7%, and 5% of the generalized variance. Many opportunities existed during 1951–63 for portfolio diversification by risk averting investors among these 30 stocks, even when short sales were outlawed. Second, to think that investors really do think in terms of 30 dimensional vector spaces seems a trifle unreasonable. We suggest that a compression of beliefs about covariance matrices into a much smaller number of dimensions is more likely to approximate their evaluation of a portfolio's risk. A very rough example suggests how such compressed information might be utilized.

Finally, we noted that investors or rather the financial press often refers to groups of stocks, firms in an industry, when discussing portfolio selection problems. This appears to reflect a belief about the structure of the covariance matrix, i.e., stocks of the same industry have positively correlated rates of return over time. We reported evidence which strongly supports this belief about the covariance matrix, but observed that this representation may not be too helpful to investors.

IV

In this section intertemporal constancy of indices, the correlation matrix, and the covariance matrix are studied by breaking the 50 quarter period in half, January 1951 to December 1956 and January 1957 to June

1963. Calculations are restricted to a and b indices; index c is not sufficiently interesting to justify the expense of further calculations. A number of different comparisons are made for both price and rate of return indices. First, are the weights associated with indices the same when indices are estimated from each of the subperiods? Second, is the percentage of the generalized variance explained by components estimated from each of the subperiods the same? Third, do components estimated in the early period accurately describe stock price movements in the second period? Of more interest to problems of portfolio selection is the degree to which rate of return covariance matrix components retain their property of orthogonality. Finally, is the popular view of structure, intra-industry similarity of movements in rates of return, empirically valid in subperiods?

Table 7 reports correlations between the largest components of price indices a and b and adjusted prices of each of the 30 Dow Jones industrials. The indices are estimated from each of the two subperiods. The correlation between the weights of the two estimated first components of indices a and b were respectively -0.086 and $+0.063$. The structure of both the covariance and the correlation matrices of stock prices was quite different in the two subperiods. This result should prove very discouraging to advocates of stock price indices.

Table 8 reports correlations between the largest components of rate of return indices a and b and rate of return of each of the 30 stocks. The correlation between the weights of the two estimated first components of the two indices were respectively $+0.561$ and $+0.552$; both differ significantly from zero using Fisher's z transformation. Although changes in the weights are undeniable, apparently some constant structure exists in both the covariance and the correlation matrices so far as the first components go. Examination of the signs of weights in the second through fifth components does not provide support for a hypothesis that these weights are constant as well. Approximately half the weights changed signs between the two subperiods for both indices. Evidently the structure of the rate of return matrices did in fact change between the two periods; some further details about the nature of the change are presented below. At least as far as the first component goes, there does seem to be some hope for constructing an interesting rate of return index.

Table 9 summarizes a different, not independent, comparison of the structure of the covariance and correlation matrices of the two subperiods. The percentage of the generalized variance explained by the five largest components is shown for each index, for both price and rate of return in each subperiod. Both price indices changed markedly between the two periods. In the early subperiod, the first component accounted for a very large proportion of the generalized variance. Partly this was owing to the

Table 7 First Components of Price Indices Estimated from Subperiods

Stock	Index *a*		Index *b*	
	1951–1956	1957–1963	1951–1956	1957–1963
Allied Chemical	0.159	0.173	0.937	0.780
Alcoa	0.575	−0.158	0.945	0.052
American Can	0.093	−0.015	0.807	−0.173
AT & T	0.062	0.861	0.918	0.806
American Tobacco	0.026	0.268	0.723	0.777
Anaconda	0.274	−0.054	0.888	0.142
Bethlehem	0.221	−0.045	0.953	0.115
Chrysler	0.019	0.015	0.216	0.001
duPont	0.982	0.661	0.965	0.699
Eastman Kodak	0.173	0.837	0.967	0.879
General Electric	0.284	0.185	0.950	0.586
General Foods	0.080	0.800	0.949	0.829
General Motors	0.209	0.210	0.968	0.575
Goodyear	0.115	0.189	0.957	0.847
International Harvester	0.044	0.264	0.685	0.870
International Nickel	0.200	0.390	0.913	0.804
International Paper	0.165	0.062	0.969	0.559
Johns-Manville	0.119	0.194	0.907	0.760
Owens-Illinois	0.227	0.370	0.951	0.825
Procter and Gamble	0.071	0.752	0.946	0.814
Sears	0.111	0.671	0.942	0.785
Standard Oil (Cal.)	0.168	0.088	0.963	0.117
Esso	0.226	−0.044	0.943	−0.294
Swift	0.087	0.132	0.659	0.836
Texaco	0.119	0.331	0.966	0.604
Union Carbide	0.385	0.231	0.962	0.610
United Aircraft	0.331	−0.195	0.936	−0.622
U.S. Steel	0.321	0.088	0.942	0.379
Westinghouse	0.073	0.153	0.625	0.495
Woolworth	0.034	0.464	0.629	0.903

larger trend component in stocks in that period. The DJI rose from 248 in March 1951 to 499 in December 1956; at the end of June 1963 it was 707. This difference is not sufficient to explain the variation. A more important consideration is that the market had no major slumps in the first period of the magnitude of those in 1957–1958 or 1962. This changed structure reinforces the previous conclusion about the lack of any stationary structure in the price matrices.

Table 8 First Components of Rate of Return Indices Estimated from Subperiods

Stock	Index *a*		Index *b*	
	1951–1956	1957–1963	1951–1956	1957–1963
Allied Chemical	0.277	0.460	0.711	0.808
Alcoa	0.521	0.514	0.576	0.677
American Can	0.131	0.152	0.366	0.517
AT & T	0.057	0.257	0.486	0.632
American Tobacco	0.036	0.257	0.057	0.575
Anaconda	0.527	0.524	0.800	0.748
Bethlehem	0.550	0.422	0.723	0.772
Chrysler	0.302	0.514	0.478	0.480
duPont	0.348	0.376	0.796	0.750
Eastman Kodak	0.271	0.394	0.553	0.669
General Electric	0.345	0.415	0.606	0.610
General Foods	0.083	0.290	0.264	0.613
General Motors	0.439	0.389	0.737	0.693
Goodyear	0.469	0.554	0.705	0.768
International Harvester	0.273	0.395	0.706	0.772
International Nickel	0.292	0.487	0.589	0.793
International Paper	0.419	0.466	0.804	0.769
Johns-Manville	0.216	0.397	0.536	0.687
Owens-Illinois	0.253	0.338	0.722	0.663
Procter and Gamble	0.304	0.284	0.784	0.560
Sears	0.315	0.368	0.637	0.714
Standard Oil (Cal.)	0.227	0.250	0.589	0.551
Esso	0.200	0.256	0.493	0.588
Swift	0.165	0.338	0.514	0.628
Texaco	0.304	0.370	0.679	0.785
Union Carbide	0.269	0.490	0.648	0.854
United Aircraft	0.549	0.146	0.365	0.318
U.S. Steel	0.570	0.639	0.766	0.894
Westinghouse	0.283	0.326	0.532	0.241
Woolworth	0.143	0.298	0.469	0.610

The story differs considerably for rates of return; the first five components appear to explain similar percentages of the generalized variance in the two periods. This statement is more accurate with respect to index *a* which concerns the covariance matrix. The explanation for the failure of signs of weights for components two through five to remain unchanged in rate of return indices is not a change in the proportion of variance which the components individually explain. Covariance and correlation matrices of rates of return thus appear to exhibit a second intertemporal constancy.

What do these intertemporal constancies of the first component of the rate of return index mean for risk averting investors? If investors only wish to reduce variation in the rate of return of these portfolios relative to the generalized variance of the 30 stocks, they could have done so in period two by purchasing stocks with low weights in the first period's

Table 9 Percent of Variance Explained by First Five Components of Indices *a* and *b*

	Price				Rate of Return			
	a		*b*		*a*		*b*	
Component	1951–1956	1957–1963	1951–1956	1957–1963	1951–1956	1957–1963	1951–1956	1957–1963
1	91	57	78	42	41	43	38	45
2	3	23	7	16	14	11	12	10
3	2	9	5	14	9	10	8	7
4	1	4	3	6	7	7	6	6
5	1	1	1	3	5	5	5	5
First 5	98	94	94	81	76	76	69	73

principal component. However, they would have no assurance that a portfolio so selected would be particularly immune to variance of the rate of return absolutely! If investors had knowledge about the path of the generalized variance, these constancies would be of more interest. Incidentally the generalized variance of rate of return index *a* increased 30 per cent in the second subperiod over its value in the first subperiod.

Figure 7 shows the first component of price index *a* plotted against the DJI when the component has been estimated from data in the first subperiod. Figure 8 shows the corresponding rate of return index plotted against our pseudo DJ rate of return index. The high correlation between the first component of the price index and the DJI no doubt is largely attributable to the considerable trend component. The explanation for the high correlation between the rate of return indices is apparently owing to the fact that the principal component and the DJ index have captured the constancy of the covariance matrix. Investors would have obtained a roughly equivalent description of the market whether they were thinking in terms of a covariance or a correlation matrix in the first subperiod.

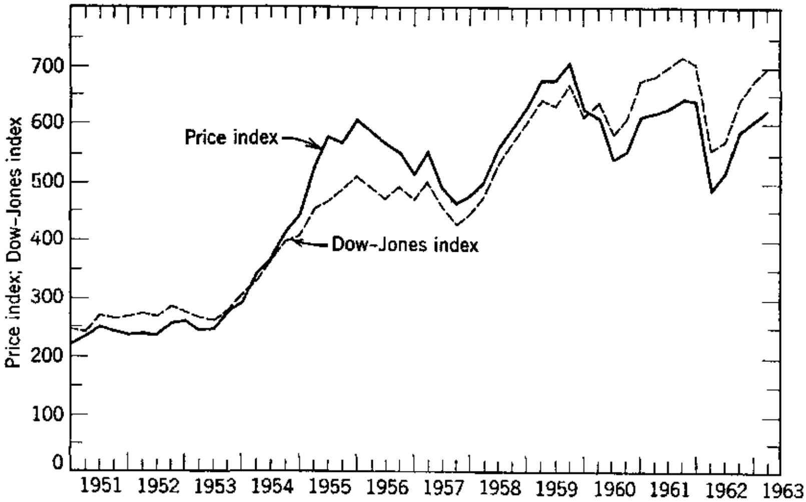

Figure 7 Index *a* component 1, $R = .946$.

A property of principal components is that they are orthogonal. If components are estimated in a subperiod, they need not retain this property over the whole period of observations. It would be convenient for portfolio selection if at least the large components of a return index estimated from a subperiod retained this property. Table 10 reports the correlation matrix in the second period of the five largest components of index *a*

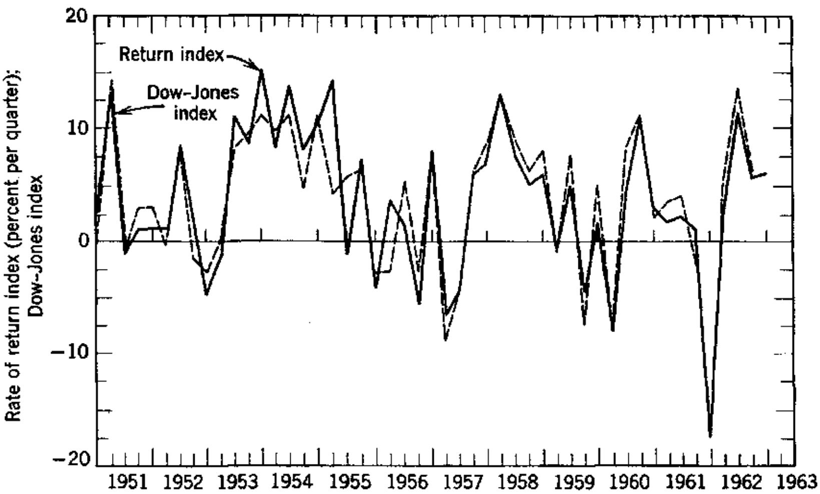

Figure 8 Index *a* component 1, $R = .915$.

estimated from the first subperiod. It is evident that the change in the structure of the covariance matrix between the two periods destroyed orthogonality.

Finally, in Section III evidence was found which supported a popular hypothesis about the structure of the covariance matrix, i.e., that rates of return of stocks of firms in an industry are positively correlated. Would this hypothesis be accepted in each of the subperiods? Table 11 reports the relevant information; it is analogous to Table 6. Evidence in favor of the hypothesis is conspicuously weaker in the second subperiod. In

Table 10 Correlation Matrix in Subperiod 2 of the Largest Five Components of Index *a* Estimated in Subperiod 1

	c_1	c_2	c_3	c_4	c_5
c_1	1				
c_2	−0.534	1			
c_3	−0.504	0.156	1		
c_4	−0.061	−0.034	0.484	1	
c_5	0.265	−0.168	0.286	0.026	1

the first 24 quarters 6 of 13 pairwise comparisons are significant in one or both of the sets of components; only two are in the second period. In 11 of the 13 comparisons the number of similar signs declined. One of the two exceptions, steel, was undoubtedly influenced considerably by the tumultuous clash between the industry and the White House in 1962. The sums of the 13 pairwise comparisons are highly significant during the first 24 quarters; only one of the two sums is significant in the second period (and it is barely so at that). The structure of the covariance matrix changed so that rates of return of stocks of firms in an industry were much less strongly correlated.

An interpretation is that during the buoyant first subperiod all firms in an industry were profiting by expanding industry sales. In the second subperiod, profits of firms were more at the expense of rivals in an industry. In a sense, the situation changed from a "seller's" market to a "buyer's" market. In the seller's market, products of firms were being disposed of in rapidly increasing quantities; the principal determinant of firms' profits and the related rate of return on their stock was the rate of growth in industry sales. In the buyer's market, although the rate of growth of industry sales continued to influence firm profits, it was joined by a second determinant, the extent to which firms were able to cut in on rivals' sales. In view of the

Table 11 Number of Industry Similar Signs among Largest 5 and Largest 10 Components of Index a, Estimated from Subperiods

	First 24 Quarters		Last 25 Quarters	
Industry-Firm	of 5	of 10	of 5	of 10
Automobiles				
Chrysler and GM	4	6	2	5
Chemicals				
duPont and Allied	4	7	3	5
Electricals				
GE and Westinghouse	5*	8*	3	5
Foods				
Swift and General Foods	5*	7	3	6
Nonferrous Metals				
Anaconda and Alcoa	4	7	3	5
Anaconda and Intnl. Nickel	4	8*	3	7
Alcoa and Intnl. Nickel	3	5	1	4
Oils				
Esso and Standard Oil (Cal.)	5*	9*	4	8*
Esso and Texaco	5*	8*	3	6
Texaco and Standard Oil (Cal.)	5*	9*	3	6
Packaging				
American Can and Owens-Illinois	3	5	2	4
Retailing				
Sears and Woolworth	3	5	4	6
Steels				
U.S. Steel and Bethlehem	4	6	5*	10*
Actual sum	54	90	39	77
Expected sum	32.5	65	32.5	65
Standard deviation	4.0	5.7	4.0	5.7
Deviation	Significant	Significant	Nonsignificant	Significant

* The asterisk denotes significance at the 0.05 level in the binomial approximation.

fact that similar signs of weights slightly exceeded the null hypothesis expectation, we infer that the first determinant was slightly stronger than the second in the second subperiod. For portfolio selection, the conclusion is that in the second subperiod investors could diversify to a greater extent by investing in different firms in an industry.

V

We conclude this chapter with a few suggestions for future research. First, serious efforts to describe empirically investors' portfolio behavior will need to make explicit the set of information which investors have on hand. An interesting experiment would be to invite a sample of investors to make estimates of the covariance between various securities. We hypothesize that the resulting subjective covariance matrix could be essentially represented in a very small number of dimensions—say, three.

Second, a formal analysis of how investor beliefs are related to news in the market is necessary if the approach of this chapter is to be carried further. What data are important or significant in the minds of investors? The Dow Theory and other such constructs argue that some information is contained in past performance of aggregative indices. An alternative theory is that information is acquired only by studying income statements, balance sheets, and the rate of growth of firms. The relative informational value of these two sources can only be appraised by studying a sample of investors.

Third, from our investigation of a 50 quarter period we believe that an informative one dimensional index of rate of return may be constructed. Further sophistication of this index requires that noise owing to situations peculiar to a particular firm be eliminated from the index. This means that some attempt must be made to estimate communalities, before extracting the largest root of the covariance matrix. Alternatively some technique must be devised by which the informational content (for the market) of a stock's rate of return movements can be separated from its total rate of return variance. We do not think that simple (linear) stock price indices are as promising a guide for investors, because weights of even the largest components appear to change considerably over relatively short periods of time.

Finally, the conspicuous weakening in the "all firms in an industry move together" hypothesis deserves more extensive investigation. Can it be shown that changes in either firms' profits or sales were more at the expense of rivals in the second subperiod than in the first? What can be said about the sensitivity of rates of return to such changes in current profits or sales, if true? In particular, what rate of discount of an earnings stream is implied by the seeming sensitivity of stock rates of return to earnings? Both the formulation and testing of these questions is a major research endeavor; its returns may well justify the investment.