by Robert D. Arnott

Cluster Analysis and Stock Price Comovement

By far the most important part of the correlation between stock price movements is explained by the effects of broad market movements. On the other hand, a model based only on overall market betas would predict a higher correlation between GM and AT&T than between GM and Ford.

The author develops five clusters of stocks corresponding to major extra-market factors, monitoring the explanatory power of each cluster statistically to determine the point in the clustering process at which the value of a cluster as a descriptor of extra-market stock price movement peaks. The clustering process terminates when clusters begin to be diluted by irrelevant stocks—for example, at the point where food companies join the utilities cluster. The explanatory power of the resulting clusters for the evaluation of extra-market risk represents better than a 30 per cent improvement over the single-index model.

The clusters embody important fundamental characteristics. The utilities cluster is dominated by interest rate sensitivity, while the cyclical cluster is strongly influenced by the economic outlook. Thus, although factor betas change over time, they are actually more stable than market betas.

The multiple factor risk model can be applied to stock classification, portfolio optimization and performance measurement, where, for example, its use simply eliminates the pesky benchmark selection problem.

INCREASING attention has been given to the comovement of security prices. It is obvious that most stocks decline in a bear market, while most stocks rise in a bull market. Thus a very important part of the correlation between stock price movements can be explained by the effects of broad market movements on each issue. On a more specific level, it is equally clear that the stock of General Motors will behave more like Ford Motor stock than AT&T stock. For a more complete understanding of how stocks behave, however, we have to look at the specific factors involved in stock price movements.

The need to understand stock comovements is particularly important in the context of portfolio analysis. A portfolio of 20 stocks from a broad range of industries will behave much the same as the market as a whole. A portfolio of 20 oil and oil service companies, on the other hand, will be exposed not only to market swings, but also to swings in the oil industry. Thus a more concentrated portfolio will be inherently riskier than a broadly diversified portfolio, regardless of the number of issues held.

In the 1960s, William Sharpe addressed the issue of portfolio risk with a simple but effective model often called the single-index model.[1] He divided the price movement of stocks into market-related and issue-specific components. In other words, stock price behavior is described solely by market behavior plus price behavior unique to the individual company. In this model, the portfolio manager cannot avoid market-related risk; he can reduce the portfolio's issue-

1. Footnotes appear at the end of article

Robert Arnott is Vice President of The Boston Company.

specific risk, on the other hand, by broadening the diversification of the portfolio so that the issue-specific price movements tend to cancel each other out.

Sharpe's model effectively describes the most significant part of stock price comovement. Unfortunately, it disregards the more subtle aspects. Indeed, the single-index model would predict a higher correlation between GM and AT&T than between GM and Ford. Consequently, it is important to evaluate security price comovements beyond the simple market-related movements in order to isolate those factors that are shared by large numbers of securities. A systematic evaluation of stock price behavior can lead to both a better understanding of stock price movements and a more rigorous classification of stocks, not according to their industry groups, but according to how the stocks actually behave.

Methodology

Our examination of extra-market risk employs a method that is closely related to one developed by James Farrell.[2] Farrell analyzed the price histories of 100 issues, removing the effects of the market in order to examine only the extra-market price movements. He combined the two issues with the highest (extra-market) correlation into a "cluster," recomputed a new set of correlations and combined the two issues or clusters with the highest correlation and repeated the process until no positive correlations remained. At that point, four "primary" clusters were present; these were loosely termed "growth," "stable," "cyclical" and "oil."

Our analysis differs in several important ways from the Farrell study. First, we studied almost 600 issues.[3] Second, we used a three-sigma filter to eliminate extraordinary data. Third, we did not carry the combination process to the point where no positive correlations remained, since we felt that doing so might result in clusters valuable as "comovement classes" in their own right being combined into a less useful "super cluster."

The combination process was closely monitored on two counts. First, the value of a cluster in describing stock price behavior was statistically monitored. This was done by measuring the correlation between each cluster and every issue in the stock universe (including members of the cluster). The mean-square correlation measures the average R^2 of all securities with the cluster (or, more specifically, of the extra-market price activity of all securities with the extra-market price activity of an equal-weighted cluster index). This mean-square correlation measures the fraction of extra-market security price activity explained by the extra-market activity of the cluster index. This provides a good indication of how significant the cluster is as a descriptor of extra-market stock price movement: The peak mean-square correlation represents the maximum value of a cluster as a "factor" in a multifactor model.

Second, the purity of each cluster was subjectively evaluated so the analysis would not reach the point where clusters were diluted by irrelevant stocks. For example, when food companies joined the "utilities" cluster, it was subjectively judged that the cluster was beginning to be diluted with weakly related issues. Experience showed that clustering terminated at about the same point under either the statistical or subjective rule. Tables I through V show the resulting clusters, which can be labeled (1) "quality growth," (2) "utilities," (3) "oil and related," (4) "basic industries" and (5) "consumer cyclical."

Continuing the clustering process until no positive extra-market correlations remained would have resulted in roughly the same four groups identified by Farrell. We judged these super clusters to be less homogeneous (i.e., diluted with weakly related issues); objectively, they were found to have about half the explanatory power (mean-square correlation) of the five groups we ended up with. On the other hand, even though we terminated the clustering process early, there were many small clusters aggregated away in arriving at the final five groups. All these smaller groups were obviously concentrated in a single industry—e.g., gold mining stocks or banking stocks—and were objectively found to have little broad explanatory power; the mean-square correlation values for these smaller clusters were found to be generally less than half the values for the five larger clusters.

Table VI summarizes our findings. Prior to eliminating market influences, the correlations between the clusters were (not surprisingly) high, as part A of Table VI shows. Part B shows, however, that eliminating market effects sharply reduced the correlations; in fact (as in Farrell's work) the correlations became largely negative.[4] The explanatory power of the five clusters for the evaluation of extra-market risk (part C of Table VI) represents better than a 30 per cent im-

Table I Growth Cluster

AMP Incorporated
Abbott Laboratories
American Express Co.
American Home Products Corporation
Anheuser-Busch, Inc.
Avon Products, Inc.
Baxter Travenol Laboratories Inc.
Betz Laboratories, Inc.
Black & Decker Mfg. Co.
Bristol-Myers Co.
Burroughs Corporation
Chesebrough-Ponds Inc.
The Coca-Cola Company
Walt Disney Productions
Eastman Kodak Co.
Emerson Electric Co.
International Business Machines Corp.
International Flavors & Fragrances Inc.
Johnson & Johnson
Eli Lilly and Company
McDonald's Corp.
Merck & Co., Inc.
Minnesota Mining & Mfg. Co.
Pepsico Inc.
Pfizer Inc.
Philip Morris, Inc.
Procter & Gamble Co.
Schering-Plough Corp.
Sperry Corp.
Squibb Corp.
Sterling Drug Inc.
Warner-Lambert Company
Xerox Corp.

Table II Utilities Cluster

American Electric Power Co., Inc.
American Natural Resources Co
American Telephone & Telegraph Co.
Baltimore Gas & Electric Co.
Brooklyn Union Gas Co.
Carolina Power & Light Co.
Central & South West Corp.
Central Telephone & Utilities Corp.
The Columbia Gas System, Inc.
Commonwealth Edison Co.
Consolidated Edison Co. of New York, Inc.
Consolidated Natural Gas Co.
Continental Telephone Corporation
Detroit Edison Co.
Duke Power Co.
El Paso Company
Enserch Corporation
Florida Power & Light Co.
General Telephone & Electronics Corporation
Gulf States Utilities Co.
Houston Industries Incorporated
Illinois Power Co.
Middle South Utilities, Inc.
New England Electric System
Niagara Mohawk Power Corp.
Northern States Power Co. (Minn.)
Ohio Edison Co.
Pacific Gas & Electric Co.
Pacific Lighting Corp.
Panhandle Eastern Pipe Line Company
Peoples Gas Company
Philadelphia Electric Co.
Property Capital Trust
Public Service Co. of Indiana, Inc.
Public Service Electric & Gas Co.
Southern California Edison Co.
Southern Co.
Tampa Electric Co.
Texas Gas Transmission Corp.
Texas Utilities Company
United Telecommunications, Inc.
Virginia Electric & Power Co.
Wisconsin Electric Power Co.

provement over the single-index model. While issue-specific risk remains the dominant source of risk for single securities, and market risk remains the dominant source of risk in portfolios, the multifactor risk model contributes significantly to our understanding of price movement.

The Multifactor Model

The final groups can, of course, be treated as indexes, with a security's price behavior described in terms of a series of "betas," one for each cluster or group. Table VII shows sample betas for several issues from different market sectors. IBM's growth beta of 0.45 means that, for every percentage point that the growth cluster appreciates *relative to the market as a whole,* IBM can be expected to rise 0.45 per cent *relative to the market as a whole.* Its negative cyclical beta indicates a similar but opposite relation with the cyclical cluster.

This analysis of extra-market price behavior makes it possible to isolate sources of risk for a company's stock. For example, Burroughs moves with the growth cluster and counter to the utilities cluster. Thus both the growth and utilities factors contribute to the risk of Burroughs' stock. Other factors play only a minor role in its price movement.

Like market betas, a stock's factor betas change over time. IBM's growth beta, a measure of its relation to quality growth stocks, was considerably higher in 1970 than in 1980. Also, statistical instability introduces a random error into any estimate of beta. However, these problems do not detract from the value or effectiveness of multifactor risk evaluation, any more than they detract from the value of market beta measures. Indeed, factor betas are actually more stable over time than market betas. This is not really surprising. We would not expect IBM's growth factor beta to change from positive to negative in the foreseeable future, nor would we expect the stock to swing from being growth-dominated to being a utilities stock.

Table III Oil and Related Cluster

Amerada Hess Corp.
Atlantic Richfield Co.
Baker International
Cities Service Co.
Conoco Inc.
Dresser Industries, Inc.,
Exxon Corporation
General American Oil Co. of Texas
Getty Oil Corp.
Gulf Oil Corp.
Halliburton Co.
Houston Oil & Minerals Corp.
Hughes Tool Company
Kerr-McGee Corp.
Louisiana Land & Exploration Co.
Marathon Oil Company
J. Ray McDermott & Co.
Mesa Petroleum Co.
Mobil Corporation
Phillips Petroleum Company
Reading & Bates Offshore Drilling Co.
Santa Fe International Corp.
Schlumberger Ltd.
Sedco, Incorporated
Shell Oil Company
Standard Oil Co. of California
Standard Oil Co. (Indiana)
Standard Oil Co. (Ohio)
Sun Company, Inc.
Superior Oil Co.
Texaco Inc.
Texas Eastern Corp.
Union Oil Co. of California
The Williams Companies

Table IV Basic Industries Cluster

Alcan Aluminum Limited
Aluminum Co. of America
AMAX Inc.
Armco Inc.
Asarco Inc.
Bethlehem Steel Corp.
Boise Cascade Corp.
Crown Zellerbach Corporation
Cyprus Mines Corp.
Dow Chemical Co.
E.I. du Pont de Nemours
Georgia-Pacific Corp.
Hercules Incorporated
Hudson Bay Mining & Smelting Co., Ltd.
INCO Limited
Inland Steel Co.
International Paper Co.
Kaiser Aluminum & Chemical Corp.
Kennecott Copper Corp.
Louisiana-Pacific Corp.
Mead Corp.
Monsanto Company
National Steel Corp.
Newmont Mining Corp.
Phelps Dodge Corporation
Potlatch Corp.
Pullman Incorporated
Republic Steel Corp.
Reynolds Metals Co.
St. Regis Paper Co.
Stauffer Chemical Company
Texasgulf Inc.
Union Camp Corp.
Union Carbide Corp.
United States Steel Corp.
Westvaco Corporation
Weyerhaeuser Co.

It is reassuring to note that these factors, although based on objective price behavior characteristics, embody important fundamental characteristics. The utility factor is dominated by interest-rate sensitivity, while the cyclical factor is strongly influenced by the economic outlook. Thus the fundamental characteristics of a stock will normally exert a stabilizing influence on its factor betas.

Factor betas can be indispensable in evaluating the risk characteristics of a portfolio. When a portfolio is heavily concentrated in utility issues, for example, random price movements will tend to cancel out issue-specific risk. Market risk remains, since all issues are exposed to it. Because the portfolio is concentrated in utility stocks, utility risk also remains. In fact, a substantial portion of the total risk of the portfolio will derive from exposure to utility factor risk — risk that a single-index model virtually ignores. Thus total portfolio risk may be as much as 50 to 70 per cent greater than a single-index model would predict.

Multifactor risk analysis can be used to reduce the risk of such a portfolio. If the manager shifts enough of the utilities portfolio into *anti-utility* issues, the utilities beta can be driven to zero.[5] The portfolio's issue-specific risk dwindles away and the market risk remains unchanged, but the utilities factor risk is systematically canceled out by balancing the utility stocks with anti-utility issues. This new portfolio—which a single-index model would show to have the same risk as the old one—would in fact be substantially less risky.

Applications

The first and simplest application of our findings is to stock classification. Stock classification is often done on the basis of industry groups, but classification on the basis of actual price behavior can be enlightening, since that behavior often differs significantly from expected results.

We would classify stocks into six classes, the first five being precisely our five clusters. We established a sixth class ("other stocks") for those stocks for which no factor was dominant. The method used in classifying each issue is straightforward: A stock is classified on the basis

Table V Consumer Cyclical Cluster

Allied Stores Corp.
American Standard Inc.
M.F. Ammanson & Co.
Armstrong Cork Co.
Associated Dry Goods Corp.
Burlington Industries, Inc.
Centex Corp.
Chrysler Corp.
Cluett, Peabody & Co., Inc.
Dayton-Hudson Corp.
Evans Products Co.
Federal Natl. Mortgage Assn.
Federated Department Stores, Inc.
First Charter Financial Corp.
Fleetwood Enterprises, Inc.
The Flintkote Co.
Ford Motor Co.
General Motors Corp.
Golden West Financial Corp.
The Goodyear Tire & Rubber Company
Great Western Financial Corp.
Imperial Corp. of America
Jim Walter Corp.
Johns-Manville Corp.
K Mart Corp.
Kaufman & Broad, Inc.
Kraft Inc.
Libbey-Owens-Ford Co.
M. Lowenstein Corp.
Lowe's Companies, Inc.
Lucky Stores Inc.
R.H. Macy & Co., Inc.
Marshall Field & Company
Masonite Corp.
May Department Stores Co.
Maytag Co.
Melville Corporation
Mohasco Corp.
National Gypsum Co.
Outboard Marine Corp.
Owens-Corning Fiberglas Corp.
PPG Industries, Inc.
J.C. Penney Company, Inc.
Redman Industries Inc.
Royal Crown Companies, Inc.
Scovill Inc.
Sears, Roebuck and Co.
Skyline Corp.
Standard Brands Paint Co.
United States Gypsum Co.
U.S. Home Corp.
F.W. Woolworth Co.

of the largest positive factor beta; those issues with no substantial positive beta (at least 0.3) are classified as "other." Any factor can have either a positive or negative effect on the price movement of a stock. Bethlehem Steel, for example, moves with the basic industries cluster and against the growth cluster, so it can be described as primarily a basic industries company and secondarily an anti-growth company.

Most of our sample stocks fit in the primary class one would expect them to. However, some curious patterns emerge in the secondary factors:

1. Retailing companies, airlines, automotive companies and broadcasters are consistently cyclical and anti-oil in their behavior.
2. Utilities and oil companies are consistently anti-growth in their price behavior.
3. Computer and high technology companies (IBM, Amdahl, Control Data, Digital Equipment, Intel, Hewlett-Packard, etc.) are consistently growth and anti-utility.
4. Banks have no consistent primary factor, ranging across all categories, while savings and loans are consistently cyclical (with pro-utility and anti-growth secondary factors).
5. Insurance companies consistently move as pro-cyclical, pro-utility companies, with either of these two factors being dominant.
6. Drug stocks tend to be growth-dominated and to have no secondary factors.
7. Many of the most popular cyclical issues, which have performed like "growth stocks" for several years, actually have neutral or negative growth betas.

Optimization

The multifactor model can also be applied to portfolio optimization. Several studies have shown no dramatic performance differences between portfolios selected with a single-index model, a multi-index model and a full covariance model. However, our research suggests that portfolios designed by using an optimizer with a multifactor risk model will be superior to portfolios designed by a single-index model. A multifactor approach is particularly well suited to a number of special optimization applications.

A pension fund, for example, has a need to diversify away from the risks that affect the underlying company. AT&T's ability to fund its pension plans is closely tied to the health of the utilities industry in general and AT&T in particular. If AT&T's pension plan holds utility stocks, and utility stocks fall, the funding obligations of the sponsor are increased precisely when the sponsor can least afford it. The portfolio managers might want to steer the pension portfolio away from issues with risk characteristics similar to those of the sponsor.

A multifactor model can readily handle this problem of "implicit investment." Assume, merely for the sake of argument, that due to the special risks of the utilities industry, the pension plans of AT&T have an implicit investment in the stock of AT&T amounting to twice the size of the entire portfolio, for a 200 per cent implicit in-

Table VI Characteristics of Clusters for Risk Evaluation

A. Cluster Correlations

S&P 500	1.00					
Growth	0.85	1.00				
Utility	0.71	0.51	1.00			
Oil	0.77	0.55	0.52	1.00		
Basic	0.84	0.60	0.55	0.65	1.00	
Cyclical	0.86	0.68	0.71	0.52	0.73	1.00

B. Extra-Market Cluster Correlations (before orthogonalization)

Growth	1.00				
Utility	−0.31	1.00			
Oil	−0.34	−0.05	1.00		
Basic	−0.37	−0.11	0.02	1.00	
Cyclical	−0.17	0.25	−0.44	0.08	1.00

C. Explanatory Powers of Multifactor Risk Model (mean-square correlations—average for all stocks in S&P 500)

	R^2 (Variance Explained)
Market Beta	0.295
Growth Beta	0.033
Utility Beta	0.018
Oil Beta	0.015
Basic Beta	0.014
Cyclical Beta	0.012

Table VII Sample Factor Betas for Selected Companies

	Market	*Growth*	*Utilities*	*Oil*	*Basic*	*Cyclical*
IBM	0.93	0.45	−0.24	−0.24	−0.33	−0.26
Digital Equipment	1.28	0.64	−0.80	−0.37	−0.17	0.07
Minnesota Mining	1.16	0.69	−0.43	−0.19	−0.12	−0.30
MGIC	1.70	−0.30	0.06	−0.16	0.33	1.01
AT & T	0.72	−0.31	0.33	−0.15	−0.15	−0.19
Phillips Petroleum	1.14	−0.41	−0.42	0.77	−0.22	−0.20
Sundstrand	1.35	−0.20	0.11	0.05	−0.02	0.84

vestment. Adding this substantial investment to the portfolio would prevent any optimizer from committing additional funds to AT&T stock. However, a single-index optimizer wold not recognize the difference between General Telephone and IBM, so a single-index optimizer will encourage investment in virtually the same issues whether or not the implicit investment in AT&T is considered. An optimizer based on a multifactor risk model would recognize the substantial similarity between the price activity of AT&T and that of all other utilities (and, indeed, all interest rate sensitive issues) and steer the pension assets away from such issues.

As another example, consider a portfolio manager who anticipates an economic upturn and wants to concentrate heavily on cyclical issues. Multifactor analysis can help; a single-index model cannot. Using multifactor analysis, the investor can constrain the optimizer to select portfolios with a cyclical beta of, for example, 0.25 or higher. Alternatively, he can adjust the estimated return on securities by an amount proportional to their cyclical betas; this will systematically steer the portfolio towards cyclical issues and away from anti-cyclical issues.

Theoretically, since the factor betas measure extra-market risk, they measure risk the market as a whole doesn't have to bear. Hence no risk premium is associated with factor betas. Bets placed on factor betas imply an attempt to time the factors. From a practical standpoint, the economic outlook does change. Multifactor betas permit a portfolio manager with economic foresight to design a portfolio that will benefit from such changes.

As a final example, consider an index fund. An index fund can readily be constructed using several hundred issues with carefully balanced security weights. If factor betas are all zero, however, then the remaining portfolio risk is predominantly issue-specific. Thus, by forcing all

factor betas to equal zero and the market beta to equal one, an optimizer can generate a good index fund from fewer issues.

A frequent problem in the index fund industry is the faulty assumption that an index of the largest issues behaves like the market as a whole. Multifactor risk analysis reveals that focusing on just the largest issues will result in a pro-growth, pro-oil, anti-utility and anti-cyclical bias. The deviations of such a portfolio from the behavior of the market will be much greater than would be predicted by a single-index risk model. The fund would fail to perform as intended.

Performance Measurement

A final area in which multifactor risk analysis can be of value is in performance measurement. If a manager has a particular style, favoring for example growth or value, that contributes substantial extra-market risk in excess of the usual issue-specific risk. If a manager is hired because of expertise in selecting, say, growth stocks, then that manager should not be expected to match the performance of a broad market indicator such as the Standard & Poor's 500. Rather, his performance should be compared with an index of factors weighted in the same proportions as the manager's actual portfolio.

Since the multifactor risk model is designed to capture and explain the dominant aspects of extra-market price comovement, it lends itself to the construction of a proper performance measurement benchmark. Suppose a portfolio has a growth beta of 0.5. Suppose also that securities with a growth beta of one outperformed a diversified market index with the same overall market beta by 10 per cent. All else equal, this portfolio should beat the market by five per cent (0.5 times 10 per cent). Failure to provide such performance (even if the portfolio beats the market) is indicative of poor issue selection.

Performance measurement is a sufficiently complex and controversial topic that any brief discussion such as this cannot adequately address its subtleties. However, the application of multifactor risk analysis to performance measurement is of substantial value. For any managers other than highly specialized "boutique" managers, multifactor risk analysis can very simply and directly eliminate the benchmark selection problem. The style issue, which has been one of the most controversial areas in performance measurement, is dealt with in a more realistic and simple manner by multifactor risk analysis than by any currently available performance measurement method.

In summary, multifactor analysis of investment risk through cluster analysis is not an academic exercise. The fuller understanding of the sources of stock price comovement conveyed by cluster analysis can significantly improve the investment decision-making process, hence investment performance. ■

Footnotes

1. William F. Sharpe, *Portfolio Theory and Capital Markets* (New York: McGraw-Hill, Inc., 1970).
2. James L. Farrell, Jr., "Analyzing Covariation of Returns to Determine Homogeneous Stock Groupings," *Journal of Business*, April 1974, pp. 186–207.
3. The stock universe consisted of the Standard & Poor's 500 plus The Boston Company's research list, and covered weekly price movements from 1968 to 1978.
4. A final orthogonalization step is necessary to eliminate all correlations for such purposes as optimization. The orthogonalization process is as follows: The effects of the market on each cluster are extracted. Then the effect of each cluster on all other clusters is similarly extracted. This process results in five factor indexes that are orthogonal (uncorrelated) with respect to the market and all other factor indexes. This process is necessary if one is using univariate calculations in a multivariate context.
5. The market as a whole, by definition, must have a market beta of one and factor betas of zero, since these factors describe *extra-market* price movement. Roughly speaking, therefore, there will be as many anti-utility issues as there are utilities.