DIVERSIFICATION AND THE REDUCTION OF DISPERSION: AN EMPIRICAL ANALYSIS*

JOHN L. EVANS** AND STEPHEN H. ARCHER**

I. INTRODUCTION

THE PROBLEMS ASSOCIATED with portfolio analysis have been the subject of intense discussion in recent years, especially after the introduction of a practical solution algorithm by Sharpe.[1] However, inspection of the literature suggests that relatively little attention has been given to one of the fundamental relationships upon which the efficacy of portfolio theory relies: the relationship between the extent of portfolio diversification and the reduction in the variation (risk) associated with portfolio returns.

Sharpe suggests that the total variation of a portfolio return may be segregated into two forms:[2] (1) Systematic variation, resulting for covariation of the returns on the individual securities with the market return; and (2) unsystematic variation, attributable to the peculiarities of the individual securities themselves—in other words, that portion of the variation of a security return not attributable to the variation of the market return. If the *co*variation between individual security returns arises solely as a result of their common correlation with the market return, it follows that the reduction in variation of a portfolio return resulting from increased diversification must be entirely a function of the reduction of unsystematic portion of the total variation. Thus, if the number of securities included in a portfolio were to approach the numbers of securities in the market, one would expect the variation of the portfolio return to approach the level of systematic variation—that is, the variation of the market return, suggesting a relationship which behaves as a decreasing asymptotic function.

This paper examines the rate at which the variation of returns for randomly selected portfolios is reduced as a function of the number of securities included in the portfolio. We demonstrate that a relatively stable and predictable relationship exists between these variables, and subsequently that the nature of this relationship, if coupled with marginal cost considerations, throws considerable light on the decision as to the optimal number of securities to be included in a portfolio.

II. THE PROBLEM

The investor's goal is to select a portfolio from among n securities, such that, regardless of the shape of his utility function with respect to risk and return,

* The authors would like to thank Professor William F. Sharpe of the University of Washington for his valuable comments and advice. Needless to say, any errors are the sole responsibility of the authors.

** University of Washington.

1 William F. Sharpe, "A Simplified Model for Portfolio Analysis," *Management Science*, Vol. IX, No. 2 (January 1963), pp. 277-93.

2. *Ibid.*

he will maximize expected return for a given level of risk, or conversely, minimize risk for a given expected return. To accomplish the above goal, the investor may rely upon a portfolio analysis model to provide him with a set of efficient portfolios.[3] If the costs associated with acquiring a portfolio are a function of the *number of shares* of the securities held, then the portfolio analysis models currently found in the literature will satisfy the necessary *and* sufficient conditions for selection of an optimal portfolio. However, if these costs are also a function of the *number of different securities* held, then the standard portfolio analysis models satisfy only the necessary conditions for selection of an optimal portfolio; to satisfy the sufficient conditions, the investor must perform a marginal analysis with respect to the costs incurred and the benefits derived from incremental increases in the number of securities included in the portfolio.[4] Marginal analysis is necessary for the following reasons: (1) Portfolio analysis *per se* does not discriminate between two portfolios identical in all respects except for the number of securities included; and (2) portfolio analysis *per se* selects between two portfolios with exactly the same expected return, the one which has the least variation of return, irrespective of how small the difference in variation is between them. Consider, for example, the following: Portfolio A, with an expected return of 14%, a standard deviation of .1231, and containing 20 securities; and Portfolio B, with an expected return of 14%, a standard deviation of .1238, and containing 15 securities.[5] Portfolio analysis alone would indicate A as optimal. However, is the addition of 5 securities, with their associated marginal cost, justified on the basis of a marginal return of .0007 reduction in portfolio standard deviation?[6] The answer is uncertain without some form of marginal analysis.

The subsequent examination will concentrate on the marginal reduction in portfolio variation resulting from successive increases in the number of securities held in a portfolio—that is, the marginal benefits to be derived from increased diversification. We will assume the following: (1) The investor is a random buyer of common stocks; (2) dividends from securities are not reinvested; and (3) equal dollar amounts are invested in each security in the portfolio.

III. The Analysis

The data used in estimating the relationship between diversification and the level of variation of portfolio returns were compiled on 470 of the securities listed in the Standard and Poor's Index for the year 1958. Observations on each security were taken at semi-annual intervals for the period January 1958-July

3. A portfolio is said to be efficient if none other has either (1) a lower variation of return for a given expected return or (2) a greater expected return for a given variation of return. See Sharpe, *Ibid.*, p. 278.

4. An example of a cost which is a function of the number of different securities included in a portfolio is the opportunity cost of search. Costs of this nature cannot be incorporated into the standard linear or quadratic portfolio selection models.

5. The standard deviation of portfolio A is the mean standard deviation for portfolios with 20 securities. The standard deviation of portfolio B is the mean standard deviation for portfolios with 15 securities.

6. The analysis is the same for two portfolios with exactly the same standard deviation, but slightly different expected returns.

1967.[7] The statistics employed in the calculation of ex post returns and the dispersion of these returns were the geometric mean and the standard deviation of the logarithms of the value relatives. The selection of these measures was a function of the following considerations. First, the geometric mean return is the yield, with continuous compounding, from holding a security (or portfolio) for a given period. Second, examination of the data indicated that the distribution of the logarithms of the changes in stock prices was more closely "normal" than was the distribution of the arithmetic changes, which led the authors to prefer the measures of central tendency and dispersion indicated above. Third, for changes less than ± 15 per cent, the change in the logarithm of price closely approximates the percentage price change, and for our purposes it is fruitful to examine percentage price changes.[8] Fourth, the standard deviation of the logarithms of the value relatives has the desirable property that it is a relative as opposed to an absolute measure of dispersion—that is, it takes into consideration the magnitude of the expected value.[9]

The ex post semi-annual return for period *i* was calculated for *each* security from the following formula, where R_i^k is the value relative (computed return) for security *k* in period *i*, P_i is the price of security *k* at the beginning of period *i*, P_{i+1}^k is the price of security *k* at the end of period *i*, and d_i is the dividend paid on security *k* during period *i*:[10]

$$R_i = \left(\frac{P_{i+1}^k + d_i^k}{P_i^k} \right) \qquad \begin{array}{r} \text{for } i = 1 \text{ to } 19 \\ k = 1 \text{ to } 470 \end{array}$$

The average return over the 19 periods for each security is given by the following, where $\bar{R}^k$ is the geometric mean return for security *k*:

$$\bar{R}^k = \exp\left(1/n \cdot \sum_{i=1}^{n} \log_e \frac{P_{i+1}^k + d_i^k}{P_i^k} \right) \qquad \begin{array}{r} \text{for } n = 19 \\ k = 1 \text{ to } 470 \end{array}$$

The standard deviation of the logarithms of the value relatives, our measure of risk, for security *k* over the 19 periods was computed as follows:

$$SD^k = \sqrt{\frac{\sum_{i=1}^{n} (\log_e \bar{R}^k - \log_e R_i^k)^2}{n-1}} \qquad \begin{array}{r} \text{for } n = 19 \\ k = 1 \text{ to } 470 \end{array}$$

Given these values for each security, 40 portfolios were selected as follows:

7. The 1958 Index contained 500 securities. Satisfactory data were available on 470 of these. All values were adjusted for stock splits and stock dividends during the period.

8. Eugene F. Fama, "The Behavior of Stock Market Prices," *Journal of Business*, Vol. 38, No. 1, (January 1965), pp. 45-6.

9. Robert A. Levy, "Measurement of Investment Performance," *Journal of Financial and Quantitative Analysis*, Vol. 3, No. 1 (March 1968), pp. 35-57.

10. The reader will note that $R_i^k - 1$ is the standard "rate of return" as found in the literature. The "rate of return" was not employed here as a result of the logarithmic formulation of our problem and the fact that the logarithm of a negative number is undefined.

a security was selected at random from among the 470 available, and its mean return and standard deviation recorded. This became an observation of a portfolio with 1 security. Next, two securities were selected at random from among the 470 and 469 available respectively,[11] and the return for each period ($\overline{R}_i$) computed from the following:

$$\overline{R}_i = 1/m \cdot \sum_{k=1}^{m} R_i^k \qquad \text{for } m = 1 \text{ to } 40$$

where *m* is the number of securities in the portfolio.

Next, the portfolio geometric mean return for the entire period considered ($\overline{R}_p$) was computed:

$$\overline{R}_p = \exp\left(1/n \cdot \sum_{i=1}^{n} \log_e \overline{R}_i \right) \qquad \text{for } n = 19$$

where *n* is the number of sub-periods.

Finally, the portfolio standard deviation (SD_p) was computed:

$$SD_p = \left[1/n - 1 \cdot \sum_{i=1}^{n} (\log_e R_p - \log_e \overline{R}_i)^2 \right]^{1/2}$$

These became an observation of a portfolio with 2 securities. The process was continued for 3, 4 . . . , 40 securities, with mean portfolio return ($\overline{R}_p$) and standard deviation (SD_p) computed for each resulting portfolio.[12] Consequently, the run produced 40 portfolios ranging in size from 1 to 40 securities (see Table 1). In all, there were 60 such runs, resulting in 2,400 portfolios (40 x 60).

The portfolios were next examined for the hypothesized relationship—decreasing portfolio standard deviation to an asymptote (at some positive level) as diversification increases. Regression analysis was performed fitting by least-squares the regression function:

$$Y = B(1/X) + A$$

to the set of values (X_i, Y_{ij}), (i = 1 to 40; j = 1 to 60), where X_i are the portfolio sizes, and the Y_{ij} are the computed mean portfolio standard deviations at each level of X_i.[13] This function yielded an extremely good fit, as indicated by a coefficient of determination of .9863 (see Figure 1). Further, the regression indicated that the mean standard diviation decreased to an asymptote, and that

11. The size of the population from which the samples were drawn was considered large enough, and the samples small enough, such that sampling without replacement would not significantly affect the accuracy of the statistic.

12. Note that once a security was selected in a particular run, this security was ineligible for further consideration in this run. Further note that each of the 40 portfolios selected is in and of itself independent from the others; i.e., each run involves an independent random selection, thus eliminating spurious results arising from serial correlation between portfolios selected in any particular run.

13. The parameters of the equation are A, the asymptote, and B, the regression coefficient. X is the value of the independent variable—number of securities included in the portfolio. Note that this function describes a rectangular hyperbola with a positive asymptote.

TABLE 1
The following describes in tabular form the portfolio selection process for one computer run:

	.	1	2	3	4	5	.	.	.	.	40
Securities selected	1	113	44	11	137	55	.	.	.	.	244
(by identification	2		219	269	27	200	.	.	.	.	88
number)	3			94	28	74	.	.	.	.	94
	4				235	71	.	.	.	.	19
	5					199	.	.	.	.	37
	.							.	.	.	.
	.								.	.	.
	.									.	.
	.										.
	40										222
Mean Portfolio Return ($\bar{R}_p$)		.117	.109	.087	.070	.071	.	.	.	.	.0608
Portfolio Standard Deviation (SD_p)		.269	.191	.170	.151	.141	.	.	.	.	.121

* Note: The *mean portfolio return* above represents the "geometric mean semi-annual return," and the *portfolio standard deviation* represents the "standard deviation of the logarithms of the value relatives."

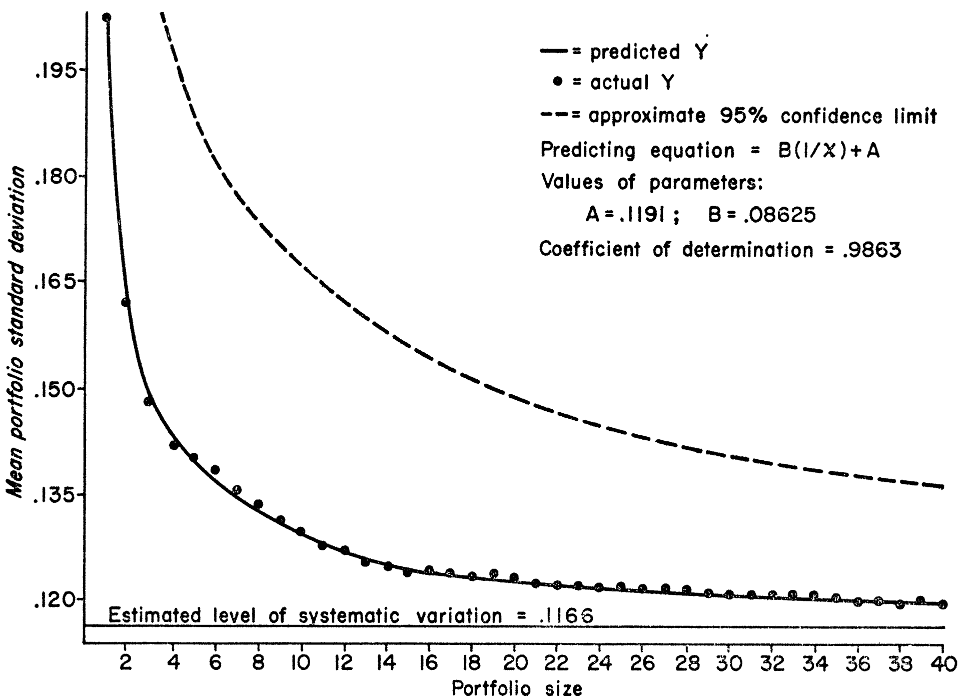

FIGURE 1

this asymptote approximated the average estimated systematic variation over the period considered.[14]

To derive some estimate of the manner in which unsystematic variation was reduced as the number of securities in the portfolio increased, two different tests were performed: (1) t-tests on successive *mean* portfolio standard deviations, which indicated on the average the significance of successive increases in portfolio size; and (2) F-tests on successive standard deviations about the *mean* portfolio standard deviation, which tend to indicate convergence of the individual observations on the mean values. As the number of securities in the portfolio approach the number of securities in the market, the dispersion about the *mean* portfolio standard deviation should approach zero, since, at the limit, all observations will include the same securities. The purpose of the F-test is to provide information as to the rate of this convergence.[15]

As is apparent from Figure 1, much of the unsystematic variation is eliminated by the time the 8th security is added to the portfolio. This observation is supported by the results of both the t and F-tests which indicated that substantial increases in portfolio size are required beyond size 8 for significant reduction to be recorded (at the .05 level) in either the mean portfolio standard deviation or in the dispersion about these mean values. For example, the t-test indicated that the addition of one security to a portfolio of size 2 caused significant reduction at the .05 level in the mean portfolio standard deviation; for portfolios of size 8, the necessary increase was 5 securities; for portfolios of size 16, the necessary increase was 19 securities; and for portfolio sizes greater than 19, no significant reduction was possible within the range of the analysis—40 securities. The F-test produced similar results with respect to the reduction of dispersion about the mean portfolio standard deviation.

It should be noted that due to the declining mean values, and the converging dispersion about these mean values, the absolute decrease in both the mean standard deviation and dispersion necessary for significant reduction to be indicated by the t and F-tests, tends to become smaller as portfolio size increases.[16] Therefore, significance of a given decrease may be a necessary, but not a sufficient condition to indicate that the corresponding increase in portfolio size is justifiable at the margin. Justification must be based upon the marginal analysis of the investor's cost function with respect to the *absolute* decrease in portfolio standard deviation.

14. The average systematic variation over the period was estimated by computing the standard deviation of the logarithms of the value relatives for a portfolio including the entire 470 securities examined. The mean standard deviation of a portfolio with 40 securities is .1197, while the standard deviation of the 470 security portfolios is .1166. It should be noted that the same general results were obtained in a similar study by the authors of 322 securities for the period 1936 to 1953 at yearly intervals.

15. Examination of the data, combined with the fact that F is a robust statistic, indicated that the necessary assumption of normality of the distribution of the actual observations about the mean was sufficiently met.

16. These conclusions hold for all levels of significance, although the level employed in this study was consistently .05. The computed covergence as well as other values derived for the function can be supplied to the interested reader by the authors.

The Conclusions

The results of the above analysis suggest that a relatively stable and predictable relationship does indeed exist between the number of securities included in a portfolio and the level of portfolio dispersion. Further, this relationship appears to take the form of a rapidly decreasing asymptotic function, with the asymptote approximating the level of systematic variation in the market.[17] The results also raise doubts concerning the economic justification of increasing portfolio sizes beyond 10 or so securities, and indicate the need for analysts and private investors alike to include some form of marginal analysis in their portfolio selection models.

17. This, of course, assumes that the relationship displayed in this analysis is homeogeneous over all ranges, and that the study data is representative.