Risk Reduction and Portfolio Size: An Analytical Solution

Edwin J. Elton; Martin J. Gruber

*Edwin J. Elton and Martin J. Gruber**

Risk Reduction and Portfolio Size: An Analytical Solution†

The relationship between the risk of a portfolio and the number of securities in that portfolio has been of interest to economists for a number of years. In fact, discussion of this problem dates back to the very inception of modern portfolio analysis.[1]

It is not surprising that this problem has received a great deal of attention. It has major implications for the structure and very existence of financial intermediaries, as well as for the behavior of all investors. When an investor decides on the size of the portfolio he will hold, he is making a trade-off between the decreased risk due to more effective diversification versus the increased transaction costs (decreased return) from adding more securities to his portfolio. If large portfolios are necessary to get most of the benefits from diversification, then financial intermediaries should exist just as a method for providing investors with the benefits of diversification. However, if one can obtain most of the benefits of diversification by holding quite small portfolios, then the individual can obtain effective diversification directly, and financial institutions must justify their existence by their ability to select security issues that will exhibit superior performance.[2]

Finally, an analytical expression for the relationship between risk and the number and type of securities is necessary to see the effects on risk of introducing new securities into the population of securities under study. For example, the opportunities and risk changes that occur when an American investor is allowed to diversify internationally can be examined via the expressions we will develop in this paper.

Before proceeding with our analysis it is worth while to briefly review some of the past studies which have examined the relationship between portfolio size and risk.

One of the earliest and perhaps the most often quoted study of risk and portfolio size was performed by Evans and Archer, who developed the methodology that was to be used by almost all following studies.[3] Evans and

1. See Harry Markowitz, *Portfolio Selection: Efficient Diversification of Investments* (New York: John Wiley & Sons, 1959), p. 111.

2. The theory of efficient markets casts doubt on the ability of financial institutions to do so. (Of course, there may be other sources of advantage: institutions may offer superior services of a custodial nature. They may also provide better and/or cheaper tax-oriented advice).

3. See J. Evans and S. Archer, "Diversification and the Reduction of Dispersion—an Empirical Analysis," *Journal of Finance* 23 (December 1968): 761–67. Subsequent

Archer measured the risk of a portfolio by the standard deviation of return from the average return for that portfolio. Then they ran a large-scale simulation and plotted the average standard deviation of return against the number of securities in the portfolio. Finally, they examined the confidence interval on the standard deviation of return. In doing so, they implicitly recognized the fact that the investor would hold one portfolio of N securities and that the risk on this portfolio might differ from the average risk of a portfolio of N securities.

The present study will differ from this standard literature in several ways. First we will derive an analytical expression for the relationship between portfolio size and risk. It is unnecessary to resort to simulation when well-developed statistical models allow you to arrive at a more exact expression. Analytical expressions not only allow the researcher to quickly determine the relationship between risk and return for new market conditions but also explicitly show what factors influence the effect of portfolio size on risk and the relative importance of each factor.

Next we will show that these earlier studies defined risk improperly. By measuring risk by the dispersion of a portfolio return around the mean return of that portfolio, they neglected the risk associated with the probability that the mean return on the portfolio held will be different from the return in the market. Put another way, the risk from holding a single security rather than the market is not just due to the variability of that security's return but is also due to the uncertainty of what the average return on that security will be. This can be illustrated with a simple example such as that presented here. Suppose that there are only three possible N security portfolios in the economy. Further, suppose that the investor is going to choose one of these portfolios at random (e.g., he knows the distribution of expected returns across the portfolios but not which expected return is identified with which portfolio). Examining only the variance around the portfolio mean would imply that they were riskless and that further diversification was senseless. However, all risk to the investor has not been eliminated. He is still unsure what return he will receive. Further diversification by reducing the uncertainty associated with the mean portfolio return may reduce his risk. If the investor bought all three portfolios in equal amounts, his risk from the mean portfolio return would in fact go to zero (see table 1).

Looking at the total risk from holding a portfolio rather than the dispersion of the portfolio return about its mean will result in higher levels of risk for any portfolio size.

There is one previous study which, while using simulation, in effect measures risk in the manner which we advocate. Fisher and Lorie simulate the distribution of returns for various portfolio sizes for all stocks listed on the New York Stock Exchange. Because they calculate the distribution of

Table 1
Example

Portfolio Number	Expected Return on Portfolio	Portfolio Variance
1	2	0
2	10	0
3	18	0

all possible portfolio outcomes their results are directly analogous to our measure of total risk.[4]

The paper discussed above, as well as our own paper, assumes equal investment in all securities in the portfolio. Equal investment is optimum if the investor has no information about future returns variances and covariances. To the extent that future returns variances and covariances can be forecast, it is possible that buying unequal amounts of each investment can lead to further reduction of risk for any size portfolio. In this case equal investment is an upper limit on the risk the investor faces.[5]

In the next section of this paper we will develop the analytical relationship between the average variance of a portfolio's return and the size of the portfolio. We will then develop an expression to measure the additional risk due to the probability that the mean return on the portfolio will differ from the mean return on the market. Two sets of expressions will be developed, one based on the full variance covariance matrix and one based on the Sharpe simplification.

In the last section of this paper the parameters of these models will be measured for one period of time and the effect of portfolio size on risk discussed.

I. RISK AND PORTFOLIO SIZE: THE ANALYTICAL RELATIONSHIP

In this section of the paper we will discuss the relationship between the number of securities in a portfolio and the investor's risk. The derivation of all the formulas discussed in the text is presented in the appendix, and all equation numbers in the text correspond to the equation numbers used in the appendix. We will discuss formulas that exactly measure the effect of the number of securities on investor's risk as well as approximations using the Sharpe single index model.

Before turning to our analytical solution let us discuss one concept that

4. Lawrence Fisher and James Lorie, "Some Studies of the Variability of Returns on Investment in Common Stock," *Journal of Business* 43, no. 2 (April 1970): 99–134. Their results are not directly analogous to the simulation studies described earlier because their study includes the risk due to mean differences as does our measure of total risk. They suggest that if one is willing to accept the existence of the second moment of the distribution, the problem could be solved analytically. Our total risk measure is an analytical solution to their problem based on the existence of the second moment.

5. For an initial exploration of this issue, see A. Johnson and B. Shannon, "A Note on the Diversification and the Reduction of Dispersion," *Journal of Financial Economics* 1 (1974): 365–72.

will be used throughout the rest of this paper. Risk will frequently be expressed as a function of the characteristics of the population of stocks from which portfolios are being selected. The analytical relationships hold however the population is defined. In the case where the investor is selecting stocks from the entire market then these characteristics are average values for the market. But if the investor is selecting stocks from a restricted population (e.g., Standard & Poor's 500), the expressions are equally valid when the characteristics (e.g., average stock variance) are measured for the restricted population.

A. Exact Formulas

The expected variance on a portfolio of N securities is (from eq. [B1] of Appendix B):

$$E(\sigma_p{}^2) = \frac{1}{N}\overline{\sigma^2} + \frac{N-1}{N}\overline{\text{cov}\,(i,j)}$$

where

$$
\begin{aligned}
E(\sigma_p{}^2) &= \text{expected variance of a portfolio;}\\
\overline{\sigma^2} &= \text{average variance for all stocks in the population;}\\
\overline{\text{cov}(i,j)} &= \text{average covariance between all stocks in the population;}\\
N &= \text{number of securities in the portfolio.}
\end{aligned}
$$

Since this formula was originally presented in Markowitz,[6] it is amazing to us how many subsequent studies have attempted to fit ad hoc equations to simulated data. These studies keep reappearing with still one more equation proposed as the appropriate one, and it is not surprising that the one that always fits best is the one that best approximates (B1).[7] One implication of this formula is that as N increases risk declines until it reaches the average covariance.

In estimating the expected variance on a portfolio of N securities a variate of equation (B1) is preferable. If $\sigma_I{}^2$ is the variance of an equally weighted portfolio of all securities in the population (EWPP), then the expected covariance can be replaced by

$$\overline{\text{cov}\,(i,j)} = \frac{M}{M-1}\left[\sigma_I{}^2 - \frac{1}{M}\overline{\sigma^2}\right], \tag{B2}$$

where M = number of securities in the population of securities under consideration.

Determining the covariance in this manner requires $2M$ rather than M^2 calculations. The expected variance on a portfolio of N securities becomes

---

6. Markowitz (see n. 1 above).

7. An example of the error that can arise from fitting an ad hoc model to the results of a simulation model can be seen comparing the results of Evans and Archer with Latané and Young. Evans and Archer fit the relationship $E(\sigma_p) = C_o + C_1(1/\sqrt{N})$, while Latané and Young fit $E(\sigma_p) = C_o + C_1(1/N)$ as well as the Evans and Archer equation to their simulated data. Latané and Young state that their model works better three out of four times. In fact, Markowitz had previously shown that the correct model was $[C_o + C_1(1/N)]^{1/2}$.

$$E(\sigma_p{}^2) = \frac{1}{N}\,\overline{\sigma^2}\left(1 - \frac{N-1}{M-1}\right) + \left(\frac{M}{M-1}\right)\left(\frac{N-1}{N}\right)\sigma_I{}^2.\quad\text{(B3)}$$

Note that equation (B3) leads to the logical conclusion that for portfolios of size M the risk is simply equal to the variance of an equally weighted population portfolio.

If the variance on a portfolio of N securities is the risk measure that is of interest to an investor, then he should be concerned not only with its expected value but with the dispersion of possible values it can take on.[8] Thus he should be concerned with the variance in the variance. The analytical expression for this is[9]

$$1/N^3\Bigg\{\left(1 - \frac{N-1}{M-1}\right)\sigma^2{}_{\sigma_i{}^2}$$

$$+\ 2(N-1)\left(1 - \frac{(N-2)(N-3)}{(M-2)(M-3)}\right)\overline{\text{cov}\,(i,j)^2}$$

$$-\ N(N-1)^2\left(1 - \frac{(N-2)(N-3)M(M-1)}{N(N-1)(M-2)(M-3)}\right)\overline{\text{cov}\,(i,j)}^2$$

$$+\ 4(N-1)(N-2)\qquad\qquad\qquad\qquad\qquad\text{(B17)}$$

$$\times\left(1 - \frac{N-3}{M-3}\right)E[\text{cov}\,(i,j)\,\text{cov}\,(i,k)]$$

$$+\ 4(N-1)\left(1 - \frac{N-2}{M-2}\right)E[\sigma_i{}^2\,\text{cov}\,(i,j)]$$

$$-\ 2(N-1)\Big[N - \frac{(N-2)M}{(M-2)}\Big]\overline{\sigma^2}\,\overline{\text{cov}\,(i,j)}\Bigg\}\,,$$

where

N = number of securities in the portfolio;

M = number of securities in the population under consideration for inclusion in the portfolio;

$\sigma^2{}_{\sigma_i{}^2} = E[\sigma_i{}^2 - \overline{\sigma^2}]^2$ = variance associated with the distribution of variances of individual securities;

$\overline{\text{cov}(i,j)^2}$ = average squared covariance for all stocks in the population;

$\overline{\text{cov}(i,j)}^2$ = the square of the average covariance for all stocks in the population;

$\sigma_i{}^2$ = the variance of security i;

$\text{cov}(i,j)$ = covariance between security i and security j);

$\overline{\sigma^2}$ and $\overline{\text{cov}(i,j)}$ as before.

8. The variance in variance was in effect simulated by the fourth condition above, N = number of securities in portfolio. As explained shortly, this is necessary in order to be able to make probability statements about the risk (variance) of any portfolio.

9. Terms like $(1 - N - 1/M - 1)$ arise because we are sampling without replacement. In other words, if the population is 500 securities and the portfolio size is 50, the first security is drawn from a sample of 500, the second from a sample of 499, the third from a sample of 498, etc.

The above expression can be used to determine confidence limits around the expected variance. This expression, along with equation (B3), allows an investor who used variance in the return on his portfolio as a measure of risk to determine the effect of portfolio size on the possible levels of risk. Hence an investor could decide on the portfolio size necessary so that there has been less than a 5% chance that his risk exceeded some predetermined level. An examination of equation (B17) shows that as N approaches M the variance in variance approaches zero. This is logical, since with M securities in the portfolio risk will be at its expected level which is the variance of an equally weighted population portfolio (hereafter called EWPP).

As discussed earlier a much more appropriate measure of risk is variation from the expected return on the population of stocks under consideration. This risk which we call total risk is composed of both the variance of the return on the portfolio of N securities from the portfolio's expected return and the variation caused by the difference between the expected return on the portfolio and the expected return for the population. The variance of the return on a portfolio of N securities from the market return is

$$\frac{1}{N}\,\overline{\sigma^2} + \frac{(N-1)}{N}\,\overline{\text{cov}\,(i,j)} + \frac{1}{N}\left(1 - \frac{N-1}{M-1}\right)\sigma_{\bar{r}}^2 \qquad \text{(B18)}$$

where

$\sigma_{\bar{r}}^2 = E(\bar{r}_i - \bar{I}_p)^2 = $ variance of the expected return on a stock
from the expected return on the population
of stocks under consideration;

All others as before.

This is identical to equation (B1) except for the last term. As discussed earlier, the expected covariance can be determined indirectly, vastly simplifying calculations. Substituting equation (B3) into equation (B18) yields

$$\frac{1}{N}\left(1 - \frac{N-1}{M-1}\right)(\overline{\sigma^2} + \overline{\sigma_{\bar{r}}^2}) + \left(\frac{M}{M-1}\right)\left(\frac{N-1}{N}\right)\sigma_I^2 \,.$$

Thus to determine the variance of the return on a portfolio of N securities requires $3\,M$ calculations. The values for the above equation for populations as large as the New York Stock Exchange can be easily calculated. As N approaches M the equation approaches the variance of EWPP. Thus, holding the EWPP leads to the maximum reduction in risk.

The expected return on a random portfolio of N securities is of course the expected return on the EWPP and is independent of the size of the portfolio held. Thus in the absence of transaction costs all investors selecting a random portfolio should hold the EWPP, since they would obtain the same expected return as they would obtain on a portfolio of any smaller size and they would be subject to less variability of the return. If transaction costs for a security were strictly proportional to the size of the transaction, then the total amount of transaction costs would be independent of the number of securities in the portfolio and their only effect would be to reduce expected

return. A fixed component to transaction costs would modify the conclusion. With a fixed component to transaction costs, transaction costs would increase as N increases. Therefore, expected return would decrease as N increases. Since expected return and variance are inversely related to N, it is likely that an investor should hold fewer securities than the market portfolio. The same consideration holds when one examines whether a small investor should hold a mutual fund to obtain diversification or randomly select his own securities.

Let us assume that neither the mutual fund nor the investor has the ability to select investments which offer an excess return. Then if transaction costs on a stock increase less than proportionally with the number of dollars invested, it may pay the investor to pool his wealth with others and buy a mutual fund. That is, the management fee charged by the fund plus the transaction costs incurred by the fund may be smaller than the cost of the investor engaging in "home-made" diversification. Economic justification for the existence of mutual funds may be found in the pooling of wealth to provide *cheap* diversification aside from any performance characteristics.

In addition to the exact formulas just derived, we also derived expressions assuming a single index model is appropriate. There are two reasons for deriving these formulas. First, the kinds of studies currently in vogue have already produced the parameters needed for these formulas as a natural by-product of other research. Second, the effect of the number of securities on risk has been analyzed ad hoc by a number of other authors in the context of the Sharpe model.[10] This analysis is an analytical derivation of the relationship between risk and the parameters of the Sharpe model.

B. Approximate Methods

Assume returns on securities can be expressed as

$$r_i = \alpha_i + \beta_i I + \epsilon_i,$$

where

r_i = return on security i;

β_i = measure of the change in the return on security i due to a change in the return on the EWPP;

I = the return on the EWPP;

α_i = the nonpopulation related return of security i;

ϵ_i = a variable with expected value of zero and variance equal to the nonpopulation related variance of security i.

In addition, make the following normal assumptions and definitions:

$$E(\epsilon_i \epsilon_j) = 0;$$
$$E(\epsilon_i)^2 = \sigma_{\epsilon}^2;$$
$$E(I - \bar{I})^2 = \sigma_I^2.$$

10. For example, see Solnick.

Then the expected variance on a portfolio of N securities is given by[11]

$$E(\sigma_p{}^2) = \bar{\beta}^2\sigma_I{}^2 + \frac{1}{N}\,\sigma_\beta{}^2\left(1 - \frac{N-1}{M-1}\right)\sigma_I{}^2 + \frac{1}{N}\,\overline{\sigma_\epsilon{}^2}. \qquad \text{(C3)}$$

This can be simplified by noting that $\bar{\beta} = 1$. As N increases, the effect of the residual risk and the variance of betas is reduced. As N approaches M the expected variance approaches the variance of the EWPP. Thus the variance of the population is the minimum risk an investor can be subject to.

The variance associated with the distribution of expected variance is[12]

$$\left(\frac{1}{N}\right)^3\left\{[\mu_4 + 4N\mu_1\mu_3 + 4N^2\mu_2\mu_1{}^2 + (2N-3)\mu_2{}^2]\sigma_I{}^4 + \{2\,\text{cov}\,(\beta_i{}^2\sigma_{\epsilon_i}{}^2)\right.$$

$$\left. + 4(N-1)\mu_1\,\text{cov}\,(\beta_i\sigma_{\epsilon_i}{}^2)]\sigma_I{}^2 + \sigma^2{}_{\sigma_\epsilon{}^2}\right\}$$

where

μ_i = the ith moment of beta around the mean beta;
$\sigma^2{}_{\sigma_\epsilon{}^2}$ = the variance of the residual variances;
all others as before.

As N gets large the above expression goes to zero.

The variance in the investor's return on a portfolio of N securities from the market return is[13]

$$\sigma_I{}^2 + \frac{1}{N}\,\sigma_\beta{}^2\left(1 - \frac{N-1}{M-1}\right)\sigma_I{}^2 + \frac{1}{N}\,\sigma_a{}^2\left(1 - \frac{N-1}{M-1}\right)$$

$$+ \frac{1}{N}\,\sigma_\beta{}^2\left(1 - \frac{N-1}{M-1}\right)I^2 + \frac{2}{N}\,E(\alpha_i\beta_i)\left(1 - \frac{N-1}{M-1}\right)\bar{I}$$

where

$\sigma_a{}^2$ = variance in α;
All else as before.

Once again as N approaches M the above expression approaches the variance in the market return. Having derived the equations necessary to determine the risk to an investor from holding a portfolio of N securities, it remains to estimate the variables in the equations. It is to this task that we now turn.

II. THE APPLICATION OF THE FORMULAS

In this section we examine the reasonableness of the Sharpe approximation and the effect of the number of securities on risk. The data we used to estimate all of the parameters consisted of weekly returns from 150 and 3,290 securities selected from the New York and American Stock Exchange over the period June 1971 to June 1974.[14]

11. Solnick derives a formula for this case, but it is incomplete.
12. In deriving this expression, we have ignored the fact that in constructing a portfolio of N securities you are sampling without replacement. Given that all formulas in this section are approximations, this seemed reasonable.
13. This equation is a simplified form of the sum of eq. (C16) and eq. (C1). The simplification is performed by recognizing that $\bar{\beta} = 1$, and $\bar{a} = 0$.
14. Our sample did not include the very few stocks that were delisted over this period.

To examine the reasonableness of the Sharpe approximations we estimated all parameters of the relevant equations using a 150 security sample. While the parameters of all equations except those necessary for the variance in variance are easy to compute, the calculators necessary for the variance in variance require a large amount of core and quickly exceeded that which we had available on a 370/50. Hence most of the analysis which follows is based on the estimate of parameters from a 150 security sample. However, in order to calculate expected portfolio variance and total variance of return, parameters were estimated for the 3,290 sample. The fact that the expansion of the sample from 150 to 3,290 will have only a minor effect on the results can be seen by comparing those parameters which were estimated on the bases of both a 150 and 3,290 security sample, as shown in table 2.

Tables 6 and 7 present comparisons of the exact method and the single index approximate method based on the values of the parameters listed in tables 3 and 4. Table 6 presents the results for the 150 stock population, and table 7 presents the results for a population of 1,000 securities which has the same characteristics (parameters) as the 150 stock population. An examination of these tables shows that the single index approximation is reasonably accurate when estimating the expected variance or total risk but is much less accurate in estimating variance in variance. This is not surprising in light of

Table 2

Parameters for Exact Method

| Variable Definition | Symbol | Parameter Value | |
		150 Security Sample	3,290 Security Sample
1. Average variance............	$\overline{\sigma^2}$	48.28	46.62
2. Average covariance	$\overline{cov(i,j)}$	6.76	7.05
3. Variance of mean security return*....................	$E(r_t - r_m)^2$	.176	.191

* From average return on market portfolio.

Table 3

Parameters Needed for Calculations Using Sharpe Approximation*

Variable Definition	Symbol	Value
1. Variance of beta......................	σ_β^2	.1675
2. Variance of alpha.................	σ_α^2	.1680
3. Mean market index........................	R_M	− .5134
4. Variance of market index....................	$\sigma_{R_M}^2$	7.0396
5. Mean variance of residuals..........	$\sigma_{\epsilon_i}^2$	40.0631
6. Third moment of beta.....................	$E(\beta_i - \beta)^3$	.0276
7. Fourth moment of beta...................	$E(\beta_i - \beta)^4$	.0954
8. Covariance of beta and residual variance...... .	$E[(\beta_i - \beta)(\epsilon_i)]$	4.0090
9. Covariance of squared beta and residual variance	$E[(\beta_i^2 - \beta^2)(\epsilon_i)]$	10.2660
10. Variance of residual variance...............	$\sigma_{\sigma_{\epsilon_i}^2}^2$	1,224.4587
11. Mean alpha beta.............................	$\alpha\beta$	.0356

* Values based on 150 security sample.

Table 4
Parameters to Compute Variance in Variance (Exact Method)

Variable Definition	Symbol	Value
Mean variance in variance.	$\overline{\sigma^2_{\sigma_i^2}}$	1,411.041
Mean of the squared covariance.	$\overline{cov(i, j)^2}$	76.384
Square of the mean covariance.	$\overline{cov(i, j)}^2$	45.675
$E[cov\ (i, j)\ cov\ (i, k)]$.	. . .	45.565
$E[\sigma_i^2\ cov\ (i, k)]$.	. . .	363.298
Mean variance times mean covariance.	$\overline{cov(i, j)}\,\overline{\sigma^2}$	326.307

Note.—Values based on 150 security sample.

Table 5
Parameters to Calculate Expected Portfolio Variance (Exact Method)[a]

Variable Definition	Symbol	Value
1. Mean variance.	$\overline{\sigma^2}$	46.619
2. Mean covariance.	$\overline{cov\ (i, j)}$	7.058
3. Variance of market portfolio.	$E(r_{im} - \bar{r}_m)^2$	7.070
4. Variance of mean security returns[b].	$E(\bar{r}_i - \bar{r}_m)^2$	.191

[a] Population estimates based on full 3,290 securities.
[b] From average return on market portfolio.

Table 6
Accuracy of Sharpe Approximation

Number of Securities in Portfolio	Expected Portfolio Variance		Variance in Variance		Variance of Outcomes	
	Exact	Approximate	Exact	Approximate	Exact	Approximate
1.	48.282	48.282	1,411.041	1,411.008	48.458	48.457
2.	27.520	27.657	201.369	203.970	27.607	27.744
4.	17.139	17.344	31.283	35.507	17.182	17.387
6.	13.679	13.907	11.027	14.717	13.707	13.935
8.	11.949	12.188	5.371	8.498	11.970	12.209
10.	10.911	11.157	3.107	5.788	10.927	11.173
20.	8.835	9.094	.590	2.120	8.842	9.102
40.	7.796	8.063	.111	.925	7.800	8.066
80.	7.277	7.547	.016	.436	7.278	7.548
150.	7.035	7.307	.000	.227	7.035	7.307

Note.—$m = 150$; parameters based on 150 security sample; values of parameters are given in table 3 for Sharpe approximation, table 4 for exact method, and $\overline{\sigma^2} = 48.2820\ \overline{cov(i, j)} = 6.7584$ and $E(\bar{r}_i - \bar{r}_m)^2 = .1758$.

the fact that the exact formula adjusted for selection without replacement while the approximate method ignored the replacement problem. An additional adjustment can be made in the approximation method so that it is even a better estimate of the true expected variance or total risk. As the number of securities in the portfolio approaches the number in the EWPP the expected portfolio variance should approach the variance of EWPP. The

Table 7

Accuracy of Sharpe Approximation

Number of Securities in Portfolio	Expected Portfolio Variance		Variance in Variance		Variance of Outcomes	
	Exact	Approximate	Exact	Approximate	Exact	Approximate
1..........	48.282	48.282	$1,411.008	1,411.008	48.458	48.458
5..........	15.603	15.287	17.887	21.539	15.098	15.322
10..........	10.911	11.163	3.241	5.788	10.928	11.180
20..........	8.835	9.101	.645	2.120	8.843	9.109
40..........	7.796	8.069	.135	.925	7.800	8.074
80..........	7.277	7.554	.028	.436	7.279	7.556
100..........	7.174	7.451	.017	.345	7.145	7.452
200..........	6.966	7.245	.003	.169	6.966	7.245
400..........	6.862	7.141	.000	.084	6.862	7.142
700..........	6.818	7.097	.000	.048	6.818	7.097

NOTE.—$m = 1,000$; parameters based on 150 security sample; values of parameters are given in table 3 for Sharpe approximation, $\sigma^2 = 48.2820$ cov $ij = 6.7584$ and table 5 for exact method, and $E(\bar{r}_i - \bar{r}_m) = .1758$.

Sharpe approximation approaches: $\sigma_I^2 + 1/m\,\sigma_e^2$. Adjusting the entries in tables 6 and 7 for large N improves the Sharpe approximation. This adjustment involves subtracting .27 from the entries in tables 4 and 5.[15]

Table 8 presents the effect of portfolio size on risk. The parameters used to calculate the expected portfolio variance and total risk are based on a sample of 3,290 securities which represents all securities that were contained in our data source. For reasons discussed previously the parameters used to calculate variance in variance were based on a sample of 150 out of the 3,290 securities.

Table 8

Effect of Diversification

Number of Securities	Expected Portfolio Variance	Variance in Variance	Total Risk
1..........	46.619	1,411.041	46.811
2..........	26.839	201.963	26.934
4..........	16.948	31.553	16.996
6..........	13.651	11.184	13.683
8..........	12.003	5.477	12.027
10..........	11.014	3.186	11.033
20..........	9.036	.623	9.045
50..........	7.849	.075	7.853
100..........	7.453	.013	7.455
200..........	7.255	.001	7.256
500..........	7.137	.000	7.137
1,000..........	7.097	.000	7.097
Minimum......	7.070	.000	7.070

NOTE.—Parameters based on 3,290 securities values shown in table 5.

15. Because the parameters were estimated from a 150 security sample, the approximate value for M in making the adjustment is 150.

Consider first the total risk. The minimum total risk is 7.07, the variance of the EWPP. The maximum total risk is 46.811, the variance of the outcomes of single security portfolios. As in other studies, the major decline in variance occurs at very low levels of N. For example, the variance of the return from the mean return or the EWPP for 10 security portfolios is 11.033, one fourth of what it was for a single security. Even though this represents a major decrease, the total risk for the 10 security portfolios is 156% of the minimum. For actual total risk to be only 20% higher than minimum total risk requires 28 securities; only 10% higher, 60 securities; and only 5% higher, 110 securities.

While total risk does go down at a slower and slower rate as more securities are added, table 6 and the analysis presented above make it clear that the decrease may still be of importance to management. For example, a 15 stock portfolio has 32% more risk than a 100 stock portfolio. Thus past studies which have shown that most of the advantages of diversification have been obtained when the portfolio contains 10–20 stocks, while correct, may have been misleading. The gains in decreased risk from adding stocks beyond 15 would appear to be significant.

For those who see risk in terms of expected variances the relevant numbers are in columns 1 and 2. The effect of the number of securities on risk follows the pattern discussed previously. As shown in column 2, our certainty about variance decreases even more rapidly than expected variance.

III. CONCLUSION

In this paper, we have presented formulas for determining the effect of diversification on risk. Exact formulas have been presented as well as approximate formulas that use data that are easily produced as a by-product from other analysis. We have shown that the approximations seem to be very good when estimating expected variance and total risk but much cruder in estimating variance in variance. Finally, we illustrated the use of the formulas on a sample of weekly data.

APPENDIX A

Throughout Appendices B and C we need to know the

$$E\left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)^2.$$

In this appendix we will derive it.

$$E\left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)^2 = \frac{1}{N}E\left(\sum_{i=1}^{N}\frac{X_i^2}{N} + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\frac{X_iX_j}{N}\right)$$

$$= \frac{1}{N}\overline{X^2} + \frac{N-1}{N}E(X_iX_j) \tag{A1}$$

But with M securities in the market

$$\overline{X} = \sum_{i=1}^{M} \frac{X_i}{M}$$

$$\overline{X^2} = \left(\sum_{i=1}^{M} \frac{X_i}{M} \right)^2 = \sum_{i=1}^{M} \frac{X_i^2}{M^2} + \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} \frac{X_i X_j}{M^2}$$

$$= \frac{M}{M^2} \overline{X^2} + \frac{M(M-1)}{M^2} E(X_i X_j)$$

rearranging

$$E(X_i X_j) = \frac{M}{M-1} \overline{X^2} - \frac{1}{M-1} \overline{X}^2 \tag{A2}$$

substituting (A2) into (A1) yields

$$\frac{1}{N} \overline{X^2} + \frac{N-1}{N} \frac{M}{M-1} \overline{X^2} - \frac{N-1}{N(M-1)} \overline{X}^2$$

or

$$\frac{1}{N} \left(1 - \frac{N-1}{M-1} \right) \overline{X^2} + \frac{1}{N} \left[N - \left(1 - \frac{N-1}{M-1} \right) \right] \overline{X}^2$$

or

$$\frac{1}{N} \left(1 - \frac{N-1}{M-1} \right) (\overline{X^2} - \overline{X}^2) + \overline{X}^2$$

or

$$\frac{1}{N} \left(1 - \frac{N-1}{M-1} \right) \sigma_X^2 + \overline{X}^2 .$$

APPENDIX B

ANALYSIS OF RISK EMPLOYING THE FULL VARIANCE-COVARIANCE STRUCTURE

I. THE EXPECTED LEVEL OF VARIANCE

The variance of any portfolio composed of an equal investment in N securities is

$$\sigma_p^2 = \sum_{i=1}^{N} \left(\frac{1}{N} \right)^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\frac{1}{N} \right)^2 \text{cov}\,(i,j) .$$

Taking expected values

$$E(\sigma_p^2) + N \left(\frac{1}{N} \right)^2 \overline{\sigma^2} + N(N-1) \left(\frac{1}{N} \right)^2 \overline{\text{cov}\,(i,j)} ,$$

$$E(\sigma^2) = \frac{1}{N} \overline{\sigma^2} + \overline{\text{cov}\,(i,j)} - \frac{1}{N} \overline{\text{cov}\,(i,j)} \tag{B1}$$

where $\overline{\sigma}^2$ is the average variance of all the securities in the market and $\overline{\text{cov}(i,j)}$ is the average covariance between any two securities in the market. We can simplify

the calculation of the above formula by developing a relationship between the variance of a market portfolio and $\overline{\text{cov}(i, j)}$.

Let M = the number of securities in the market;
$\sigma_{R_{m_i}}{}^2$ = the variance of an equally weighted market portfolio;
I = the return on an equally weighted market portfolio;
$\bar{I}$ = the expected value of I;
r_i = a return on security i.

$$\sigma_{R_m}{}^2 = E(I - \bar{I})^2 = E\left(\sum_{i=1}^{M} \frac{r_i}{M} - \sum_{i=1}^{M} \frac{\bar{r}_i}{M}\right)^2$$

$$= 1/M^2 E[(r_1 - \bar{r}_1) + (r_2 - \bar{r}_2) + \ldots + (\bar{r}_M - \bar{r}_M)]^2$$

$$= \sum_{i=1}^{M} \frac{E(r_i - \bar{r}_i)^2}{M^2} + \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} \frac{E(r_i - r_i)(r_j - \bar{r}_j)}{M^2}$$

$$\sigma_I{}^2 = 1/M\overline{\sigma^2} + \frac{M-1}{M} \overline{\text{cov}(i, j)} .$$

Rearranging,

$$\overline{\text{cov}(i, j)} = \frac{M}{M-1}[\sigma_I{}^2 - \frac{1}{M}\overline{\sigma_I{}^2}] . \tag{B2}$$

Substituting equation (B2) into equation (B1) yields

$$E(\sigma_p{}^2) = \frac{1}{N}\overline{\sigma^2}\left(1 - \frac{N-1}{M-1}\right) + \frac{M}{M-1}\frac{N-1}{N}\sigma_I{}^2 . \tag{B3}$$

II. VARIANCE IN VARIANCE

$$\sigma^2_{\sigma_p{}^2} = E(\sigma_p{}^2 - \overline{\sigma_p{}^2})^2 = E\left\{\left[\sum_{i=1}^{N}\left(\frac{1}{N}\right)^2\sigma_i{}^2 - \frac{1}{N}\overline{\sigma^2}\right]\right.$$
$$\left. + \left[\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j \neq i}}^{N}\left(\frac{1}{N}\right)^2 \text{cov}(i, j) - \frac{N^2 - N}{N^2}\overline{\text{cov}(i, j)}\right]\right\}^2 . \tag{B4}$$

Squaring this yields two squared terms and cross-product terms. Consider the first squared term. It can be written as

$$\left(\frac{1}{N}\right)^2 E\left[\sum_{i=1}^{N}\left(\frac{1}{N}\right)\sigma_i{}^2 - \overline{\sigma^2}\right]^2 . \tag{B5}$$

In Appendix A we demonstrate that

$$E\left(\frac{1}{N}\sum_{i=1}^{N} X_i\right)^2 = \overline{X_i{}^2} + \frac{1}{N}\sigma_X{}^2\left(1 - \frac{N-1}{M-1}\right) ,$$

where X is a characteristic of a stock. For the expression above

$$X_i = (\sigma_i{}^2 - \overline{\sigma^2}) .$$

Thus,

$$\overline{X_i{}^2} = 0 \quad \text{and} \quad \sigma_X{}^2 = E(\sigma_i{}^2 - \overline{\sigma^2})^2 = \sigma^2_{\sigma_i{}^2} .$$

Therefore expression (B5) is

$$\left(\frac{1}{N}\right)^2 E\left(\sum_{i=1}^{N} \frac{1}{N}\sigma_i^2 - \overline{\sigma^2}\right)^2 = \frac{1}{N^3}\left(1 - \frac{N-1}{M-1}\right)\sigma^2_{\sigma_i^2}. \qquad (B6)$$

Now consider the second squared term from equation (B4):

$$E\left[\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N}\left(\frac{1}{N}\right)^2 \text{cov}\,(i,j) - \frac{N^2-N}{N^2}\overline{\text{cov}\,(i,j)}\right]^2$$

$$= \frac{1}{N^4}E\left[\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N}(\text{cov}\,i,j)\right]^2 - 2\frac{N^2-N}{N^4}$$

$$- \overline{\text{cov}\,(i,j)}\, E\left[\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N}\text{cov}\,(i,j)\right] \qquad (B7)$$

$$+ \left(\frac{N^2-N}{N^2}\right)^2 \overline{\text{cov}\,(i,j)}^2$$

$$= \frac{1}{N^4}E\left(\sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N}\text{cov}\,(i,j)\right)^2 - \left(\frac{N-1}{N}\right)^2 \overline{\text{cov}\,(i,j)}^2.$$

The first term in this expression can be written as

$$\frac{4}{N^4}E\left[\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}(\text{cov}\,i,j)\right].$$

Expanding the expression inside the expected value operator yields three types of terms:

1. There are $N(N-1)/2$ terms of the form $\overline{\text{cov}\,(i,j)}^2$.
2. There are $N(N-1)(N-2)$ terms of the form $E[\text{cov}(i,j)\,\text{cov}(i,k)]$.
3. There are $N(N-1)(N-2)(N-3)/4$ terms of the form $E[\text{cov}(i,j)\,\text{cov}(k,p)]$.

Thus expression (B7) can be written as $1/N^3\{2(N-1)\,\overline{\text{cov}(i,j)}^2 + 4(N-1)(N-2)\,E[\text{cov}(i,j)\,\text{cov}(i,k)] + (N-1)(N-2)(N-3)\,E[\text{cov}(i,j)\,\text{cov}(k,p)] - N(N-1)^2\,\overline{\text{cov}(i,j)}^2\}$. \qquad (B8)

This expression can be simplified and the amount of calculations needed to compute it reduced by expanding the definition for $\overline{\text{cov}(i,j)}^2$:

$$\overline{\text{cov}\,(i,j)}^2 = \left[\frac{2\sum_{i=1}^{M-1}\sum_{j=i+1}^{M}\text{cov}\,(i,j)}{M(M-1)}\right]^2$$

Expanding this expression one gets the product of $4/M^2(M-1)^2$ and the following terms:

$$\frac{M(M-1)}{2} \text{ terms of the form } \overline{\text{cov}\,(i,j)}^2\,;$$

$$M(M-1)(M-2) \text{ terms of the form } E[\text{cov}\,(i,j)\,\text{cov}\,(i,k)]\,;$$

$$\frac{M(M-1)(M-2)(M-3)}{4} \text{ terms of the form } E[\text{cov}\,(i,j)\,\text{cov}\,(k,p)]\,.$$

Simplifying and rearranging

$$E[\text{cov}\,(i,j)\,\text{cov}\,(k,\,p)] = \frac{M(M-1)}{(M-2)(M-3)}\,\overline{\text{cov}\,(i,j)^2}$$

$$-\frac{2}{(M-2)(M-3)}\,\overline{\text{cov}\,(i,j)^2} - \frac{4}{(M-3)}\,E[\text{cov}\,(i,j)\,\text{cov}\,(i,k)]\,.$$

Substituting this expression into (B8) yields

$$\frac{1}{N^3}\left\{2(N-1)\,\overline{\text{cov}\,(i,j)^2}\left[1 - \frac{(N-2)(N-3)}{(M-2)(M-3)}\right]\right.$$

$$+ 4(N-1)(N-2)E[\text{cov}\,(i,j)\,\text{cov}\,(i,k)]\left(1 - \frac{N-3}{M-3}\right) \quad \text{(B9)}$$

$$\left. - N(N-1)^2\,\overline{\text{cov}\,(i,j)}^2\left[1 - \frac{(N-2)(N-3)M(M-1)}{N(N-1)(M-2)(M-3)}\right]\right\}\,.$$

There are four cross-product terms in (B4). They are

$$2\left(\frac{1}{N}\right)\frac{N^2-N}{N^2}\,\overline{\text{cov}\,(i,j)}\,\overline{\sigma^2} = \frac{2(N-1)}{N^2}\,\overline{\text{cov}\,(i,j)}\,\overline{\sigma^2} \quad \text{(B10a)}$$

$$\frac{-2}{N}\,\overline{\sigma^2}\,E\left[\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\left(\frac{1}{N^2}\right)\text{cov}\,(i,j)\right] \quad \text{(B10b)}$$

$$-2\,\frac{N^2-N}{N^2}\,\overline{\text{cov}\,(i,j)}\,E\left[\sum_{i=1}^{N}\left(\frac{1}{N}\right)^2\sigma_i{}^2\right] \quad \text{(B10c)}$$

$$2E\left(\sum_{i=1}^{N}\left(\frac{1}{N}\right)^2\sigma_i{}^2\right)\left[\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\frac{1}{N^2}\,\text{cov}\,(i,j)\right]\,. \quad \text{(B10d)}$$

Equation (B10b) can be written as

$$-2\left(\frac{1}{N}\right)\left(\frac{1}{N^2}\right)(N^2-N)\,\overline{\sigma^2}\,\overline{\text{cov}\,(i,j)}\,.$$

Simplifying yields

$$-2\left(\frac{N-1}{N^2}\right)\overline{\sigma^2}\,\overline{\text{cov}\,(i,j)}. \quad \text{(B11)}$$

Equation (B10c) can be written as

$$-2\left(\frac{N^2-N}{N^2}\right)\left(\frac{1}{N}\right)^2(N)\,\overline{\sigma^2}\,\overline{\text{cov}\,(i,j)}\,.$$

Simplifying yields

$$-2\left(\frac{N-1}{N^2}\right)\overline{\sigma^2}\,\overline{\text{cov}\,(i,j)}\,. \quad \text{(B12)}$$

Equation (B10d) can be written as

$$2\left(\frac{1}{N}\right)^2\left(\frac{1}{N}\right)^2 E\left(\sum_{i=1}^{N}\sigma_i{}^2\right)\left[\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\text{cov}\,(i,j)\right]\,.$$

There are two types of terms in the above, $\sigma_i^2 \text{cov}(i,j)$ and $\sigma_i^2 \text{cov}(k,j)$. The total number of terms is $N(N^2 - N)$. The number of the form $\sigma_i^2 \text{cov}(i,j)$ is $2N(N-1)$. The number not in common is $N(N^2 - N) - 2N(N-1)$. Thus, the above can be written as

$$2\left(\frac{1}{N}\right)^4 2N(N-1)E[\sigma_i^2 \text{cov}(i,j)]$$

plus

$$2\left(\frac{1}{N}\right)^4 [N(N^2 - N) - 2N(N-1)]E[\sigma_i^2 \text{cov}(k,j)] .$$

Simplifying the above two expressions we have

$$4\left(\frac{N-1}{N^3}\right)E[\sigma_i^2 \text{cov}(i,j)] \tag{B13}$$

and

$$\frac{2}{N^3}(N-1)(N-2)E[\sigma_i^2 \text{cov}(k,j)] . \tag{B14}$$

Now examine

$$\sum_{i=1}^{M} \frac{\sigma_i^2}{M} \sum_{i=1}^{M} \sum_{\substack{j=1 \\ j \neq i}}^{M} \frac{\text{cov}(i,j)}{M(M-1)} .$$

There are $2(M-1)$ terms where a particular σ_i^2 is multiplied by $\text{cov}(i,j)$. Since there are M terms of the form σ_i^2, there are $2M(M-1)$ terms of the form $\sigma_i^2 \text{cov}(i, j)$. In total there are $M^2(M-1)$ terms. Thus the above expression contains $M^2(M-1) - 2M(M-1) = M(M-1)(M-2)$ terms of the form $\sigma_i^2 \text{cov}(k,j)$. Therefore

$$\overline{\sigma^2}\,\overline{\text{cov}(i,j)} = \frac{(M-2)}{M}E[\sigma_i^2 \text{cov}(k,j)] + 2/ME[\sigma_i^2 \text{cov}(i,j)] .$$

Solving for $E[\sigma_i^2 \text{cov}(k,j)]$ and substituting into (B14) yields

$$\left(\frac{M}{M-2}\right)\left(\frac{2}{N^3}\right)(N-1)(N-2)\overline{\sigma^2}\,\overline{\text{cov}(i,j)}$$

$$-\left(\frac{2}{M-2}\right)\left(\frac{2}{N^3}\right)(N-1)(N-2)E[\sigma_i^2 \text{cov}(i,j)] . \tag{B15}$$

Thus the cross-product term is the sum of (B10), (B11), (B12), (B13), and (B15), or

$$1/N^3 \left\{ 4(N-1)\left(1 - \frac{N-2}{M-2}\right)E[\sigma_i^2 \text{cov}(i,j)] \right.$$

$$\left. - 2(N-1)\left(N - \frac{N-2}{M-2}M\right)\overline{\sigma^2}\,\overline{\text{cov}(i,j)} \right\} \tag{B16}$$

$E(\sigma_p{}^2 - \sigma_p{}^2)$ is the sum of (B6), (B9), and (B16), or

$$
\begin{aligned}
1/N^3 \Bigg\{ & \left(1 - \frac{N-1}{M-1}\right) \sigma^2{}_{\sigma_i{}^2} \\
& + 2(N-1)\left[1 - \frac{(N-2)(N-3)}{(M-2)(M-3)}\right]\overline{\text{cov}\,(i,j)^2} \\
& - N(N-1)^2\left[1 - \frac{(N-2)(N-3)M(M-1)}{N(N-1)(M-2)(M-3)}\right]\overline{\text{cov}\,(i,j)^2} \\
& + 4(N-1)(N-2) \qquad\qquad\qquad\qquad\qquad\qquad (\text{B17}) \\
& \times \left(1 - \frac{N-3}{M-3}\right)E[\text{cov}\,(i,j)\,\text{cov}\,(i,k)] \\
& + 4(N-1)\left(1 - \frac{N-2}{M-2}\right)E[\sigma_i{}^2\,\text{cov}\,(i,j)] \\
& - 2(N-1)\left(N - \frac{(N-2)M}{(M-2)}\right)\overline{\sigma^2\,\text{cov}\,(i,j)} \Bigg\} ,
\end{aligned}
$$

III. ANOTHER MEASURE OF RISK

Up to now we have followed the standard approach of assuming that risk is measured by the uncertainty (variability) of a portfolio's return over time. But total risk from holding a portfolio consists of the dispersion of returns on that portfolio plus the risk that we hold a portfolio with an average return different from the market.

$E(\sigma_{Tp}{}^2)$ = the total expected risk from holding a portfolio of securities.

$$E(\sigma_{Tp}{}^2) = E(\sigma_p{}^2) + E(\bar{R}_p - \bar{R}_M)^2$$

The last term in the above expression can be written as

$$E\left(\frac{\sum_{i=1}^{N}\tilde{R}_i}{N} - \bar{R}_M\right)^2 .$$

Once again using the fact that

$$E\left(\frac{1}{N}\sum_{i=1}^{N}X_i\right)^2 = \bar{X}_i{}^2 + \frac{1}{N}\sigma_X{}^2\left(1 - \frac{N-1}{M-1}\right),$$

we see that

$$E\left(\frac{\sum_{i=1}^{N}\tilde{R}_i}{N} - \bar{R}_M\right)^2 = \frac{1}{N}E(\bar{R}_i - \bar{R}_M)^2\left(1 - \frac{N-1}{M-1}\right).$$

Thus

$$E(\sigma_{Tp}{}^2) = E(\sigma_p{}^2) + \frac{1}{N}E(\bar{R}_i - \bar{R}_M)^2\left(1 - \frac{N-1}{M-1}\right) \qquad (\text{B18})$$

where $E(\sigma_p{}^2)$ is given by equation (B3).

APPENDIX C

ANALYSIS OF RISK USING SHARPE'S

SIMPLIFICATION METHOD

This appendix parallels Appendix B but makes a simplifying assumption. In this appendix we will make the Sharpe assumption that security returns are uncorrelated once the effect of the market is removed.

A. DERIVATION OF THE EXPECTED VARIANCE OF A PORT-FOLIO OF N SECURITIES

The return on a security can be written as $r_i = \alpha_i + \beta_i I + \epsilon_i$, where the terms are defined as earlier. If the residual risk is independent of the market risk (which is usually true because of the estimation procedure) then the total risk on a portfolio of N securities is equal to the expected portfolio β^2 times the market risk plus the expected residual risk, or

$$E(\sigma_p{}^2) = E(\beta_p{}^2) \sigma_I{}^2 + E(\sigma_{\epsilon_p}{}^2).$$

Employing the expression for

$$E \left(\sum_{i=1}^{N} \frac{X_i}{N} \right)^2$$

given in Appendix A,

$$E(\beta_p{}^2)\sigma_I{}^2 = \bar{\beta}^2\sigma_I{}^2 + \frac{1}{N} \sigma_\beta{}^2 \left(1 - \frac{N-1}{M-1} \right) \sigma_I{}^2. \tag{C1}$$

The expected residual risk is

$$E(\sigma_{\epsilon_p}{}^2) = E \left(\sum_{i=1}^{N} \frac{\epsilon_i}{N} \right)^2.$$

If the normal Sharpe assumption is made that $E(\epsilon_i\epsilon_j) = 0$, then

$$E(\sigma_{\epsilon_p}{}^2) = \frac{1}{N} E \left[\sum_{i=1}^{N} \frac{\epsilon_i{}^2}{N} \right] = \frac{1}{N} \sigma_{\epsilon_i}{}^2. \tag{C2}$$

Thus the expected variance of a portfolio is the sum of (C1) and (C2) or

$$E(\sigma_p{}^2) = \bar{\beta}^2\sigma_I{}^2 + \frac{1}{N} \sigma_\beta{}^2 \left(1 - \frac{N-1}{M-1} \right) \sigma_I{}^2 + \frac{1}{N} \overline{\sigma_\epsilon^2}. \tag{C3}$$

B. THE VARIANCE OF TOTAL RISK OF A PORTFOLIO OF N SECURITIES

In this section we will make one additional assumption. We will assume that the sample is sufficiently large so that the correction necessary to adjust for change in the population as additional securities are selected for the portfolio can be ignored.

The variance of total risk is

$$E \left\{ \left[\left(\sum_{i=1}^{N} \frac{\beta_i}{N} \right)^2 \sigma_I{}^2 + \sum_{i=1}^{N} \sigma_{\epsilon_i}{}^2 \right] \right.$$

$$\left. - \left(\bar{\beta}^2\sigma_I{}^2 + \frac{1}{N} \sigma_\beta{}^2 \left(1 - \frac{N-1}{M-1} \right) \sigma_I{}^2 + \frac{1}{N} \overline{\sigma_\epsilon^2} \right) \right\}^2.$$

Recalling that $\sigma_\beta^2 = \overline{\beta^2} - \bar{\beta}^2$, and ignoring the correction term, the above expression can be written as

$$E\left\{\left[\left(\sum_{i=1}^{N} \frac{\beta_i}{N}\right)^2 - \frac{1}{N}\,\overline{\beta^2} - \frac{N-1}{N}\,\bar{\beta}^2\right]\sigma_I^2 + \left(\sum_{i=1}^{N} \sigma_{\epsilon_i}^2 - \frac{1}{N}\,\overline{\sigma_\epsilon^2}\right)\right\}^2.$$

The terms involving β are

$$E\left[\left(\sum_{i=1}^{N} \frac{\beta_i}{N}\right)^2 - \frac{1}{N}\,\overline{\beta^2} - \left(1 - \frac{1}{N}\right)\bar{\beta}^2\right]^2, \tag{C4}$$

which can be written as

$$E\left[\sum_{i=1}^{N} \frac{\beta_i^2}{N^2} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{\beta_i\beta_j}{N^2} - \frac{1}{N}\,\overline{\beta^2} - \left(1 - \frac{1}{N}\right)\bar{\beta}^2\right]^2.$$

Five types of terms arise when the above expression is squared and expected values are taken. These are

$$\overline{\beta^4}, \ \overline{\beta^3}\,\bar{\beta}, \ \overline{\beta^2}\,\bar{\beta}^2, \ \bar{\beta}^4, \ \text{and} \ \overline{\beta^{22}}.$$

The coefficients of these terms will now be discussed. β^4 can only arise when one of the β_i^2 terms is squared. The coefficient of this term squared is $1/N^4$ and we have N terms of the form β_i^2. Thus the coefficient of β^4 is $(N)(1/N^4)$. $\overline{\beta^3}\bar{\beta}$ can only arise when one of the $\beta_i\beta_j$ terms is multiplied times one of the β_i^2 terms and the subscripts are the same. The coefficient of $\beta_i\beta_j$ is $1/N^2$ of β_i^2 is $1/N^2$. There are $(N^2 - N)$ terms of the form $\beta_i\beta_j$ and we can get a β^3 term twice, namely, when the subscript of β^2 is i or j. Multiplying this by the two which occurs because of the squaring process, we have as the coefficient of $\overline{\beta^3}\bar{\beta}(1/N^2)(1/N^2)(N^2 - N)(2)(2)$. A similar analysis can be performed for the other terms. $\overline{\beta^2}\bar{\beta}^2$ comes about in five ways. The ways this can occur are listed below along with the coefficients of the terms and the number of ways it can be obtained.

1. Associated with $\overline{\beta^2}$ times $\bar{\beta}^2$ is $\left[\dfrac{-(N-1)}{N}\right] - \left(\dfrac{-1}{N}\right)(2)$.

2. Associated with $\bar{\beta}^2$ times β_i^2 is $\left[\dfrac{-(N-1)}{N}\right]\left(\dfrac{1}{N^2}\right)(N)(2)$.

3. Associated with $\overline{\beta^2}$ times $\beta_i\beta_j$ is $\left(\dfrac{-1}{N}\right)\left(\dfrac{1}{N^2}\right)(N^2 - N)(2)$.

4. Associated with $\beta_i\beta_j$ times β_k^2 is $\left(\dfrac{1}{N^2}\right)\left(\dfrac{1}{N^2}\right)(N^2 - N)(N - 2)(2)$.

5. Associated with $\beta_i\beta_j$ times $\beta_i\beta_k$ is $\left(\dfrac{1}{N^2}\right)\left(\dfrac{1}{N^2}\right)N(N - 1)(N - 2)(4)$.

Simplifying, we have $[4(N - 1)(N - 3)/N^3]\overline{\beta^2}\bar{\beta}^2$. β^4 comes from three expressions:

1. Associated with $\bar{\beta}^2$ times $\beta_i\beta_j$ is $\left(\dfrac{-(N-1)}{N}\right)\left(\dfrac{1}{N^2}\right)(N^2 - N)(2)$.

2. Associated with $\bar{\beta}^2$ squared is $\left[\dfrac{-(N-1)}{N}\right]^2$.

3. Associated with $\beta_i\beta_j \times \beta_k\beta_L$ is

$$\left(\frac{1}{N^2}\right)\left(\frac{1}{N^2}\right)(N)(N - 1)(N - 2)(N - 3).$$

Simplifying, we have $[-2(2N - 3)(N - 1)\bar{\beta}^4]/N^3$.

Finally $\overline{\beta^2}^2$ comes from four terms:

1. Associated with β_i^2 times $\overline{\beta^2}$ is $\left(\dfrac{1}{N^2}\right)\left(\dfrac{-1}{N}\right)(N)(2)$

2. Associated with $\overline{\beta^2}$ squared is $\left(\dfrac{-1}{N}\right)^2$

3. Associated with β_i^2 times β_j^2 is $\left(\dfrac{1}{N^2}\right)\left(\dfrac{1}{N^2}\right)(N)(N - 1)$.

4. Associated with $\beta_i\beta_j$ squared is $\left(\dfrac{1}{N^2}\right)^2(N^2 - N)(2)$.

Simplifying yields $[(2N - 3)/N^3]\bar{\beta}^{22}$. Thus the expected value of (A4) is $1/N^3[\bar{\beta}^4 + 4(N - 1)\bar{\beta}^2\bar{\beta} + 4(N - 1)(N - 3)\bar{\beta}^2 \bar{\beta}^2 - 2(2N - 3)(N - 1)\bar{\beta}^4 + (2N - 3)\bar{\beta}^{22}]$. Defining μ_i as the ith moment around the mean the above is equal to

$$(1/N^3)[\mu_4 + 4N\mu_1\mu_3 + 4N^2\mu_2\mu_1^2 + (2N - 3)\mu_2^2] . \tag{C5}$$

Let us now turn to the cross-product term or

$$2\sigma_I^2 E\left\{\left[\frac{\left(\sum_{i=1}^{N}\beta_i\right)^2}{N^2} - \frac{1}{N}\overline{\beta^2} - \frac{N-1}{N}\bar{\beta}^2\right]\left(\frac{\sum_{i=1}^{N}\sigma_{\epsilon_i}^2}{N^2} - \frac{\overline{\sigma_\epsilon^2}}{N}\right)\right\} .$$

Recognizing this as the form $E(X - \overline{X})(Y - \overline{Y})$, which is equal to $E(XY) - \overline{X}\overline{Y}$, we have

$$\frac{2\sigma_I^2}{N^4}\left\{E\left[\left(\sum_{i=1}^{N}\beta_i\right)^2\left(\sum_{i=1}^{N}\sigma_{\epsilon_i}^2\right)\right] - (N^2\overline{\beta^2} + N^2(N - 1)\bar{\beta}^2)(\overline{\sigma_\epsilon^2})\right\} . \tag{C6}$$

Examine the term

$$E\left[\left(\sum_{i=1}^{N}\beta_i\right)^2\left(\sum_{i=1}^{N}\sigma_{\epsilon_i}^2\right)\right] .$$

Expanding the summation shows that terms of the form $\beta_i^2\sigma_{\epsilon_i}^2$ will appear N times, terms of the form $\beta_i^2\sigma_{\epsilon_j}^2$ will appear $N(N - 1)$ times, terms of the form $\beta_i\beta_j\sigma_{\epsilon_k}^2$ will appear $N(N - 1)(N - 2)$ times, and terms of the form $\beta_i\beta_j\sigma_{\epsilon_j}^2$ will appear $2(N)(N - 1)$ times.

Thus

$$\left(\sum_{i=1}^{N}\beta_i\right)^2\left(\sum_{i=1}^{N}\sigma_{\epsilon_i}^2\right) = N\beta_i^2\sigma_{\epsilon_i}^2 + N(N - 1)\beta_i^2\sigma_{\epsilon_j}^2$$

$$+ N(N - 1)(N - 2)\beta_i\beta_j\sigma_{\epsilon_k}^2 + 2N(N - 1)\beta_i\beta_j\sigma_{\epsilon_j}^2$$

$$E\left[\left(\sum_{i=1}^{N}\beta_i\right)\left(\sum_{i=1}^{N}\sigma_{\epsilon_i}^2\right)\right] = NE(\beta_i^2\sigma_{\epsilon_i}^2) + N(N - 1)\overline{\beta^2}\,\overline{\sigma_\epsilon^2}$$

$$+ N(N - 1)(N - 2)\overline{\beta^2\sigma_\epsilon^2} + 2N(N - 1)\bar{\beta}E(\beta_i\sigma_{\epsilon_i}^2)$$

Substituting this expression in equation (C6) and simplifying yields

$$\frac{2\sigma_I^2}{N^3}\left\{[E(\beta_i^2\sigma_{\epsilon_i}^2) - \overline{\beta^2}\,\overline{\sigma_\epsilon^2}] + 2(N-1)\bar{\beta}[E(\beta_i\sigma_{\epsilon_i}^2) - \bar{\beta}\,\overline{\sigma_\epsilon^2}]\right\}$$

or

$$\frac{2\sigma_I^2}{N^3}\left\{[\mathrm{cov}\,(\beta_i^2\sigma_{\epsilon_i}^2) + 2(N-1)\bar{\beta}\,\mathrm{cov}\,(\beta_i\sigma_{\epsilon_i}^2)]\right\}. \tag{C7}$$

Let us now turn to the risk due to the residual variance squared or

$$E\left(\frac{\sum_{i=1}^{N}\sigma_{\epsilon_i}^2}{N^2} - \frac{\overline{\sigma_\epsilon^2}}{N}\right)^2 = E\frac{\left(\sum_{i=1}^{N}\sigma_{\epsilon_i}^2\right)}{N^4} - \frac{\overline{\sigma_\epsilon^2}^2}{N^2}. \tag{C8}$$

Expanding

$$\sum_{i=1}^{N}\sigma_{\epsilon_i}^2$$

one gets N terms of the form $\sigma_{\epsilon_i}^2\sigma_{\epsilon_i}^2$ and $N(N-1)$ terms of the form $\sigma_{\epsilon_i}^2\sigma_{\epsilon_j}^2$.
Taking the expected values

$$E\left(\sum_{i=1}^{N}\sigma_{\epsilon_i}^2\right)^2 = NE[(\sigma_{\epsilon_i}^2)^2] + N(N-1)\overline{\sigma_\epsilon^2}^2$$

Substituting in equation (C8)

$$1/N^3\{E[(\sigma_{\epsilon_i}^2)^2] - \overline{\sigma_\epsilon^2}^2\} \tag{C9}$$

or the variance in the residual variance divided by N^3.

Combining (C5), (C7), and (C9) yields as the variance in the variance

$$1\,|\,N^3\,\sigma_I^2]^2[\mu_4 + 4N\mu_1\mu_3 + 4N^2\mu_2\mu_1^2 + (2N-3)\mu_2^2]$$

$$+ \sigma_I^2[2\,\mathrm{cov}\,(\beta_i^2\sigma_{\epsilon_i}^2) + 4(N-1)\mu_1\,\mathrm{cov}\,(\beta_i\sigma_{\epsilon_i}^2)] + [\sigma^2_{\sigma_\epsilon}].$$

C. ANOTHER MEASURE OF RISK

As discussed in Appendix B, the actual uncertainty in N security portfolio depends not only on the variance of the portfolio itself but also on the deviation of the expected return from the market mean. This risk associated with the fact that the average portfolio return can be different than the market return can be represented by

$$E[\bar{r}_p - \bar{I}]^2 = E\left[\sum_{i=1}^{N}\frac{\bar{r}_i}{N} - \bar{I}\right]^2 = E\left[\frac{1}{N}\sum_{i=1}^{N}r_i\right]^2 - I^2. \tag{C10}$$

$$E\left(\frac{1}{N}\sum_{i=1}^{N}r_i\right)^2 = E\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right) + E\left(\frac{1}{N}\sum_{i=1}^{N}\beta_i\right)^2 I^2$$

$$+ 2E\left(\frac{1}{N^2}\sum_{i=1}^{N}\alpha_i\sum_{i=1}^{N}\beta_i\right)\bar{I}. \tag{C11}$$

Employing the relationship derived in Appendix A

$$E\left(\frac{1}{N}\sum_{i=1}^{N}\alpha_i\right) = \bar{\alpha}^2 + \frac{1}{N}\sigma_\alpha^2\left(1 - \frac{N-1}{M-1}\right) \tag{C12}$$

$$E\left(\frac{1}{N}\sum_{i=1}^{N}\beta_i\right) = \bar{\beta}^2 + \frac{1}{N}\sigma_\beta^2\left(1 - \frac{N-1}{M-1}\right) \tag{C13}$$

The last past of expression (C11) can be simplified using

$$E\left(\frac{1}{N^2}\sum_{i=1}^{N}\alpha_i\sum_{i=1}^{N}\beta_i\right) = E\left(\frac{1}{N^2}\sum_{i=1}^{N}\alpha_i\beta_i\right) + \left(\frac{1}{N}\right)^2 E\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}(\alpha_i\beta_i)\right)$$
$$= \frac{1}{N}E(\alpha_i\beta_i) + \frac{N-1}{N}E(\alpha_i\beta_j) . \tag{C14}$$

But

$$\bar{\alpha}\,\bar{\beta} = \frac{1}{M}E(\alpha_i\beta_i) + \frac{M-1}{M}E(\alpha_i\beta_j) .$$

Solving for $E(\alpha_i\beta_i)$ and substituting into equation (C14) yields

$$E\left(\frac{1}{N^2}\sum_{i=1}^{N}\alpha_i\sum_{j=1}^{N}\beta_i\right) = \frac{(N-1)M}{N(M-1)}\bar{\alpha}\,\bar{\beta} + \frac{1}{N}\left(1 - \frac{N-1}{M-1}\right)E(\alpha_i\beta_i) . \tag{C15}$$

Combining equations (C12), (C13), (C15), and (C11) yields

$$E(\bar{r}_p - \bar{I})^2 = \bar{\alpha}^2 + \bar{\beta}^2\bar{I}^2 + 2\frac{M(N-1)}{N(M-1)}\bar{\alpha}\,\bar{\beta}\,\bar{I}$$

$$+ \frac{1}{N}\sigma_\alpha^2\left(1 - \frac{N-1}{M-1}\right) + \frac{1}{N}\sigma_\beta^2\left(1 - \frac{N-1}{M-1}\right)\bar{I}^2 \tag{C16}$$

$$+ \frac{2}{N}E(\alpha_i\beta_i)\left(1 - \frac{N-1}{M-1}\right)\bar{I} - \bar{I}^2 .$$

This expression plus equation (C3) represents the total risk from holding a portfolio of securities of size N rather than holding the market portfolio.