TABLE 1.9

Fractiles SR(p, T) of the Distribution of the Studentized Range in Samples of Size T from a Normal Population

SIZE OF SAMPLE	LOWER PERCENTAGE POINTS (p)					UPPER PERCENTAGE POINTS (p)					SIZE OF SAMPLE
T	.005	.01	.025	.050	.10	.90	.95	.975	.99	.995	T
						1.997	1.999	2.000	2.000	2.000	3
						2.409	2.429	2.439	2.445	2.447	4
						2.712	2.753	2.782	2.803	2.813	5
						2.949	3.012	3.056	3.095	3.115	6
						3.143	3.222	3.282	3.338	3.369	7
						3.308	3.399	3.471	3.543	3.585	8
						3.449	3.552	3.634	3.720	3.772	9
10	2.47	2.51	2.59	2.67	2.77	3.57	3.685	3.777	3.875	3.935	10
11	2.53	2.58	2.66	2.74	2.84	3.68	3.80	3.903	4.012	4.079	11
12	2.59	2.65	2.73	2.80	2.91	3.78	3.91	4.01	4.134	4.208	12
13	2.65	2.70	2.78	2.86	2.97	3.87	4.00	4.11	4.244	4.325	13
14	2.70	2.75	2.83	2.91	3.02	3.95	4.09	4.21	4.34	4.431	14
15	2.75	2.80	2.88	2.96	3.07	4.02	4.17	4.29	4.43	4.53	15
16	2.80	2.85	2.93	3.01	3.13	4.09	4.24	4.37	4.51	4.62	16
17	2.84	2.90	2.98	3.06	3.17	4.15	4.31	4.44	4.59	4.69	17
18	2.88	2.94	3.02	3.10	3.21	4.21	4.38	4.51	4.66	4.77	18
19	2.92	2.98	3.06	3.14	3.25	4.27	4.43	4.57	4.73	4.84	19
20	2.95	3.01	3.10	3.18	3.29	4.32	4.49	4.63	4.79	4.91	20
30	3.22	3.27	3.37	3.46	3.58	4.70	4.89	5.06	5.25	5.39	30
40	3.41	3.46	3.57	3.66	3.79	4.96	5.15	5.34	5.54	5.69	40
50	3.57	3.61	3.72	3.82	3.94	5.15	5.35	5.54	5.77	5.91	50
60	3.69	3.74	3.85	3.95	4.07	5.29	5.50	5.70	5.93	6.09	60
80	3.88	3.93	4.05	4.15	4.27	5.51	5.73	5.93	6.18	6.35	80
100	4.02	4.00	4.20	4.31	4.44	5.68	5.90	6.11	6.36	6.54	100
150	4.30	4.36	4.47	4.59	4.72	5.96	6.18	6.39	6.64	6.84	150
200	4.50	4.56	4.67	4.78	4.90	6.15	6.38	6.59	6.85	7.03	200
500	5.06	5.13	5.25	5.37	5.49	6.72	6.94	7.15	7.42	7.60	500
1000	5.50	5.57	5.68	5.79	5.92	7.11	7.33	7.54	7.80	7.99	1000

Source: H. A. David, H. O. Hartley, and E. S. Pearson, "The Distribution of the Ratio, in a Single Normal Sample, of Range to Standard Deviation," *Biometrika*, 61 (1954): 491. Reprinted by permission.

CHAPTER

2

The Distribution of the Return on a Portfolio

The next empirical question concerns the relationships between the returns on individual stocks and market returns. To what extent are returns on individual securities associated with or explained by market returns, as represented, for example, by the return R_{mt} on the equally weighted index or portfolio of NYSE common stocks?

Study of this topic requires two chapters of preliminary discussion of statistical concepts. Many of these concepts are also relevant for the model of portfolio selection pursued at length later in the book. Thus, to enliven the discussion of the new statistical tools and to set the stage for the later work in portfolio theory, this chapter introduces some concepts from portfolio theory and uses them as the framework for the discussion of new statistical tools.

The first step is to show how the return on a portfolio is related to the returns on the individual securities in the portfolio.

I. A Portfolio's Return as a Function of Returns on Securities

Consider a particular portfolio (call it p) and let h_{ip} be the number of dollars invested in security i at the end of month $t - 1$ (which, in a discrete time framework, is also the beginning of month t). Let $\tilde{R}_{it}$ be the simple return on the security from the end of month $t-1$ to the end of month t. The return is as defined by equation (13) of Chapter 1, so that $\tilde{R}_{it}$ is the return from the end of month $t-1$ to the end of month t per dollar invested in security i at the end of month $t-1$. As in Chapter 1, the tilde (˜) on $\tilde{R}_{it}$ indicates that the return is a random variable at $t-1$.

At the end of month t, the dollar value of the investment h_{ip} is

$$h_{ip} + h_{ip}\tilde{R}_{it} = h_{ip}(1 + \tilde{R}_{it});$$

that is, the end-of-month value is the initial investment h_{ip} plus the dollar return $h_{ip}\tilde{R}_{it}$. If n is the number of securities, the end-of-month dollar value of the portfolio is

$$\sum_{i=1}^{n} h_{ip} + \sum_{i=1}^{n} h_{ip}\tilde{R}_{it} = \sum_{i=1}^{n} h_{ip}(1 + \tilde{R}_{it}).$$

The end-of-month value of the portfolio can also be expressed as $h(1 + \tilde{R}_{pt})$, where $\tilde{R}_{pt}$ is the return on the portfolio p for month t and

$$h = \sum_{i=1}^{n} h_{ip}, \tag{1}$$

are the total funds invested at the beginning of the month. It follows that

$$h + h\tilde{R}_{pt} = \sum_{i=1}^{n} h_{ip} + \sum_{i=1}^{n} h_{ip}\tilde{R}_{it} = h + \sum_{i=1}^{n} h_{ip}\tilde{R}_{it},$$

so that

$$h\tilde{R}_{pt} = \sum_{i=1}^{n} h_{ip}\tilde{R}_{it}; \tag{2}$$

that is, the dollar return on the portfolio can be expressed either as the total investment times the return on the portfolio or as the sum of the dollar returns on the investments in each of the securities. If we let

$$x_{ip} = \frac{h_{ip}}{h}, \tag{3}$$

so that

$$\sum_{i=1}^{n} x_{ip} = 1, \tag{4}$$

then dividing through both sides of equation (2) by h, we have

$$\tilde{R}_{pt} = \sum_{i=1}^{n} x_{ip}\tilde{R}_{it}. \tag{5}$$

The quantity x_{ip} is the proportion of total portfolio funds h invested in security i to obtain portfolio p. Thus equation (5) says that the return on portfolio p is a weighted average of the returns on the individual securities in p, where the weight applied to a security's return is the proportion of portfolio funds invested in the security.

One example of a portfolio is the equally weighted index of NYSE stocks studied in Chapter 1. For this portfolio

$$\tilde{R}_{mt} = \frac{1}{n}\sum_{i=1}^{n} \tilde{R}_{it} = \sum_{i=1}^{n} x_{im}\tilde{R}_{it},$$

$$x_{im} = \frac{1}{n}, \quad i = 1, 2, \ldots, n,$$

where n is the number of securities on the NYSE at the end of month $t-1$.

In describing the collection or set of portfolios from which an investor can choose, it is convenient to let n be the total number of securities that are candidates for inclusion in portfolios. Then, given the returns on the n securities for month t, the only reason that different portfolios have different returns is that the weights or proportions of portfolio funds invested in securities vary from portfolio to portfolio. In this sense, the weights x_{ip}, $i = 1, 2, \ldots, n$, define or characterize the portfolio p. It is understood that some of the x_{ip} can be zero, which means that some securities do not appear in portfolio p.

II. The Mean and Variance of a Portfolio's Return

As indicated by the tilde notation, at the end of month $t-1$ the returns for month t on securities and portfolios are random variables; that is, the values of the returns that will be observed can be thought of as drawings from probability distributions. Since the return on a portfolio is a weighted sum of the returns on the securities in the portfolio, determining how the distribution of the return on a portfolio is related to the distributions of returns on securities involves, in statistical terms, determining how the distribution of a weighted sum of random variables is related to the distributions of the individual summands.

The problem is simplified by the fact that the portfolio models of this book are based on the assumption, supported by the empirical work of Officer (1971) and Blume (1968), and the data presented in Chapter 1, that, at least for monthly post-World War II data, distributions of portfolio returns, like distributions of returns for individual common stocks, are approximately normal. A normal distribution can be completely characterized from knowledge of its mean and standard deviation. Thus, the problem reduces to one of determining how means and standard deviations of portfolio returns are related to the parameters of distributions of returns on securities. In statistical terms, the problem is to develop expressions for the mean and standard deviation of a weighted sum of random variables.

Since the object of the book is to teach finance, not statistics, most of the relevant results are just stated in the text, with proofs left for the problems.

A. *The Mean or Expected Value of the Return on a Portfolio*

Since a portfolio's return is a weighted sum of returns on securities, to describe the mean and standard deviation of a portfolio's return we must first know something about the means and standard deviations of weighted random variables. There are two general results. First, the mean (or expected value, or expectation) of a constant times a random variable is the constant times the expected value of the random variable. Thus, for any constant α and any random variable $\tilde{y}$,

$$E(\alpha\tilde{y}) = \alpha E(\tilde{y}). \tag{6}$$

Second, the variance of a constant times a random variable is the constant squared times the variance of the random variable, so that the standard deviation of a constant times a random variable is the absolute value of the constant times the standard deviation of the random variable:

$$\sigma^2(\alpha\tilde{y}) = \alpha^2\sigma^2(\tilde{y}) \tag{7}$$

$$\sigma(\alpha\tilde{y}) = |\alpha|\sigma(\tilde{y}). \tag{8}$$

The absolute value sign is necessary in (8) since the constant α could be negative and the standard deviation of $\alpha\tilde{y}$, like any standard deviation, must be nonnegative.

PROBLEM II.A

1. Derive equations (6) and (7).

ANSWER

1. Let $f(y)$ be the density function for the random variable $\tilde{y}$, assumed to be continuous. Then

$$\begin{aligned} E(\alpha\tilde{y}) &= \int_y \alpha y f(y)dy \\ &= \alpha \int_y y f(y)dy \\ &= \alpha E(\tilde{y}), \end{aligned}$$

and

$$\begin{aligned} \sigma^2(\alpha\tilde{y}) &= E\{[\alpha\tilde{y} - E(\alpha\tilde{y})]^2\} \\ &= \int_y [\alpha y - E(\alpha\tilde{y})]^2 f(y)dy \\ &= \alpha^2 \int_y [y - E(\tilde{y})]^2 f(y)dy \\ &= \alpha^2\sigma^2(\tilde{y}). \end{aligned}$$

Although this interpretation is not rigorous, the nonmathematical reader can consider the integral notation $\int_y dy$ as calling for a "sum" over all possible values of $\tilde{y}$. Note that since we are summing over all possible specific values of $\tilde{y}$, in the above equations there are no tildes over the y's that follow an integral sign.

The reader will find it instructive to rewrite the expressions above for a discrete random variable $\tilde{y}$. This involves interpreting $f(y)$ as a probability function rather than as a density function and substituting the summation symbol Σ_y for the integral notation $\int_y dy$. The reader should always interpret what he or she does in words.

The return on a portfolio is a weighted sum of random variables. The mean or expected value of a random variable which is itself a weighted sum of random variables is the sum of the weighted means or expected values of the variables that make up the sum. Thus, if $\tilde{y}_1, \ldots, \tilde{y}_n$ are n arbitrary random variables and $\alpha_1, \ldots, \alpha_n$ are arbitrary weights, then

$$E\left(\sum_{i=1}^{n} \alpha_i\tilde{y}_i\right) = \sum_{i=1}^{n} \alpha_i E(\tilde{y}_i). \tag{9}$$

Expressed verbally, the expectation of a sum of weighted random variables is the sum of the weighted expectations.

Applying (9) to equation (5), the expected return on any portfolio p is

$$E(\tilde{R}_{pt}) = E\left(\sum_{i=1}^{n} x_{ip}\tilde{R}_{it}\right) = \sum_{i=1}^{n} x_{ip}E(\tilde{R}_{it}). \tag{10}$$

Thus, the mean or expected return on a portfolio of n securities is the weighted average of the means of the returns on individual securities, where the weight applied to the expected return on a given security is the proportion of portfolio funds invested in that security.

The results stated in equations (6) through (10) are used repeatedly in this and later chapters.

PROBLEM II.A

2. Establish (9) for the two-variable case; that is, show that for any two constants α_1 and α_2 and any random variables $\tilde{y}_1$ and $\tilde{y}_2$,

$$E(\alpha_1\tilde{y}_1 + \alpha_2\tilde{y}_2) = \alpha_1 E(\tilde{y}_1) + \alpha_2 E(\tilde{y}_2).$$

The answer requires some familiarity with the concepts of joint, conditional, and marginal probabilities, and some familiarity either with multiple integrals or multiple sums.

ANSWER

2. Let $f(y_1, y_2)$ be the joint density for the random variables $\tilde{y}_1$ and $\tilde{y}_2$; that is, $f(y_1, y_2)$ gives the likelihood that a joint drawing of $\tilde{y}_1$ and $\tilde{y}_2$ will yield the particular pair of values of the variables shown as arguments of the function. The expected value of $\alpha_1\tilde{y}_1 + \alpha_2\tilde{y}_2$ is then the weighted average of $\alpha_1 y_1 + \alpha_2 y_2$ over all possible combinations of y_1 and y_2, where the weight applied to any specific combination is its joint density $f(y_1, y_2)$.

$$E(\alpha_1\tilde{y}_1 + \alpha_2\tilde{y}_2) = \int_{y_1,y_2} (\alpha_1 y_1 + \alpha_2 y_2) f(y_1, y_2)\,dy_1 dy_2,$$

where $\int_{y_1,y_2} dy_1 dy_2$ is loosely read "sum over all possible combinations of y_1 and y_2."

Let $f(y_1|y_2)$ be the density function for $\tilde{y}_1$ conditional on some given value y_2 of $\tilde{y}_2$, and likewise let $f(y_2|y_1)$ be the conditional density function for $\tilde{y}_2$ given that y_1 is observed in the drawing of $\tilde{y}_1$. Let

$$f(y_1) = \int_{y_2} f(y_1, y_2)\,dy_2$$

be the marginal density function for $\tilde{y}_1$; that is, $f(y_1)$ shows the likelihood that y_1 is observed in the drawing of $\tilde{y}_1$ when no constraint is imposed on what is observed in the drawing of $\tilde{y}_2$. Thus, $f(y_1)$ is just the sum of $f(y_1, y_2)$ over all possible values of $\tilde{y}_2$. Likewise, the marginal density function for $\tilde{y}_2$ is

$$f(y_2) = \int_{y_1} f(y_1, y_2)\,dy_1.$$

Since the joint density $f(y_1, y_2)$ can always be expressed as

$$f(y_1, y_2) = f(y_1|y_2) f(y_2)$$

or as

$$f(y_1, y_2) = f(y_2|y_1) f(y_1),$$

the equation for $E(\alpha_1\tilde{y}_1 + \alpha_2\tilde{y}_2)$ given above can be developed as

(Step 1) $E(\alpha_1\tilde{y}_1 + \alpha_2\tilde{y}_2) = \int_{y_1,y_2} \alpha_1 y_1 f(y_1, y_2)\,dy_1 dy_2 + \int_{y_1,y_2} \alpha_2 y_2 f(y_1, y_2)\,dy_1 dy_2$

(Step 2) $= \alpha_1 \int_{y_1,y_2} y_1 f(y_2|y_1) f(y_1)\,dy_1 dy_2 + \alpha_2 \int_{y_1,y_2} y_2 f(y_1|y_2) f(y_2)\,dy_1 dy_2$

(Step 3) $= \alpha_1 \int_{y_1} y_1 f(y_1) \int_{y_2} f(y_2|y_1)\,dy_2 dy_1 + \alpha_2 \int_{y_2} y_2 f(y_2) \int_{y_1} f(y_1|y_2)\,dy_1 dy_2$

(Step 4) $= \alpha_1 \int_{y_1} y_1 f(y_1)\,dy_1 + \alpha_2 \int_{y_2} y_2 f(y_2)\,dy_2$

(Step 5) $= \alpha_1 E(\tilde{y}_1) + \alpha_2 E(\tilde{y}_2).$

Step 2 makes legitimate rearrangements of the terms in step 1. Step 4 takes account of the fact that the conditional probability distributions of step 3 are bona fide probability distributions; that is, for any given y_2 the sum of $f(y_1|y_2)$ over all possible values of y_1 is 1:

$$\int_{y_1} f(y_1|y_2)dy_1 = 1;$$

and likewise

$$\int_{y_2} f(y_2|y_1)dy_2 = 1.$$

Again, the reader may want to rework this problem for the case where the random variables $\tilde{y}_1$ and $\tilde{y}_2$ are discrete rather than continuous.

The expected value of a portfolio's return is the weighted sum of the expected values of returns on its constituent securities irrespective of the presence or absence of dependence among the security returns. This is not generally true for the variance of a portfolio's return. The variance of a portfolio's return is in part determined by the variances of security returns, but it is also determined in part and often primarily by the degree of dependence or comovement in the returns on different securities.

The notation used in the discussions that follow gets rather involved. To simplify things a little, we no longer explicitly include the subscript t on returns and on the parameters (e.g., means and variances) of distributions of returns. This should not cause confusion, since the specific period t to which the various quantities refer is of no particular importance. Thus, we now write equations (5) and (10) for the return and expected return on portfolio p as

$$\tilde{R}_p = \sum_{i=1}^{n} x_{ip}\tilde{R}_i \tag{11}$$

$$E(\tilde{R}_p) = E\left(\sum_{i=1}^{n} x_{ip}E(\tilde{R}_i)\right) = \sum_{i=1}^{n} x_{ip}E(\tilde{R}_i). \tag{12}$$

B. The Variance of the Return on a Portfolio

As for any random variable, the variance of the return on a portfolio is

$$\sigma^2(\tilde{R}_p) = E\left\{[\tilde{R}_p - E(\tilde{R}_p)]^2\right\}.$$

With equations (11) and (12), $\sigma^2(\tilde{R}_p)$ can be rewritten as

$$\sigma^2(\tilde{R}_p) = E\left(\left[\sum_{i=1}^{n} x_{ip}(\tilde{R}_i - E(\tilde{R}_i))\right]^2\right).$$

This expression calls for the expected value of a sum of weighted random variables. To see what is involved, it is best to begin with the simple case, $n = 2$. Then the preceding expression becomes

$$\begin{aligned}\sigma^2(\tilde{R}_p) &= E([x_{1p}(\tilde{R}_1 - E(\tilde{R}_1)) + x_{2p}(\tilde{R}_2 - E(\tilde{R}_2))]^2) \\ &= E(x_{1p}^2[\tilde{R}_1 - E(\tilde{R}_1)]^2 + x_{2p}^2[\tilde{R}_2 - E(\tilde{R}_2)]^2 \\ &\qquad + 2x_{1p}x_{2p}[\tilde{R}_1 - E(\tilde{R}_1)][\tilde{R}_2 - E(\tilde{R}_2)]).\end{aligned}$$

Since $\tilde{R}_1$ and $\tilde{R}_2$ are random variables, the cross-product $[\tilde{R}_1 - E(\tilde{R}_1)][\tilde{R}_2 - E(\tilde{R}_2)]$ is a random variable, as are the squared differences from means $[\tilde{R}_1 - E(\tilde{R}_1)]^2$ and $[\tilde{R}_2 - E(\tilde{R}_2)]^2$. In general, the value of any nonconstant function of one or more random variables is itself a random variable. Thus, the preceding equation says that $\sigma^2(\tilde{R}_p)$ is the expected value of a sum of weighted random variables. Since the expectation of a sum of weighted random variables is the sum of the weighted expectations of the component variables, we have

$$\begin{aligned}\sigma^2(\tilde{R}_p) &= x_{1p}^2E([\tilde{R}_1 - E(\tilde{R}_1)]^2) + x_{2p}^2E([\tilde{R}_2 - E(\tilde{R}_2)]^2) \\ &\qquad + 2x_{1p}x_{2p}E([\tilde{R}_1 - E(\tilde{R}_1)][\tilde{R}_2 - E(\tilde{R}_2)]).\end{aligned} \tag{13}$$

The expressions $E([\tilde{R}_1 - E(\tilde{R}_1)]^2)$ and $E([\tilde{R}_2 - E(\tilde{R}_2)]^2)$ are the return variances $\sigma^2(\tilde{R}_1)$ and $\sigma^2(\tilde{R}_2)$. To complete the interpretation of the preceding equation, we need only interpret the quantity $E([\tilde{R}_1 - E(\tilde{R}_1)][\tilde{R}_2 - E(\tilde{R}_2)])$ called the *covariance* between $\tilde{R}_1$ and $\tilde{R}_2$. The covariance $E([\tilde{R}_1 - E(\tilde{R}_1)][\tilde{R}_2 - E(\tilde{R}_2)])$ is an expected value which is evaluated by weighting each possible value of $[R_1 - E(\tilde{R}_1)][R_2 - E(\tilde{R}_2)]$ by $f(R_1, R_2)$, the joint density or likelihood of observing that combination of R_1 and R_2 in a joint drawing of $\tilde{R}_1$ and $\tilde{R}_2$, and then "summing" over all possible combinations of R_1 and R_2. In formal terms, the covariance between the returns on any two securities i and j is denoted either as cov $(\tilde{R}_i, \tilde{R}_j)$ or as σ_{ij}, and is defined as

$$\begin{aligned}\text{cov}(\tilde{R}_i, \tilde{R}_j) &= \sigma_{ij} = E([\tilde{R}_i - E(\tilde{R}_i)][\tilde{R}_j - E(\tilde{R}_j)]) \\ &= \int_{R_i,R_j} [R_i - E(\tilde{R}_i)][R_j - E(\tilde{R}_j)]f(R_i, R_j)dR_idR_j.\end{aligned} \tag{14}$$

As in Problem II.A.2, the integral notation $\int_{R_i,R_j} dR_idR_j$ calls for a "sum" over all possible combinations of R_i and R_j.

As its name implies, the covariance is a measure of the degree of covariation (or comovement or association) between the returns on securities i and j. In intuitive terms, the covariance is positive if deviations of $\tilde{R}_i$ and $\tilde{R}_j$ from their respective means tend to have the same sign, and it is negative if the deviations tend to have opposite signs. When the covariance is positive, we say that there is positive association or dependence between $\tilde{R}_i$ and $\tilde{R}_j$; roughly speaking, the returns on the two securities tend to move in the same direction. A negative covariance indicates negative association or dependence; the returns on the two securities tend to move in opposite directions. The covariance concept appears so frequently in future discussions that a thorough understanding evolves naturally.

PROBLEM II.B

1. The random variables $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n$ are statistically independent if

$$f(y_1, y_2, \ldots, y_n) = f(y_1)f(y_2)\ldots f(y_n);$$

that is, if their joint density is always the product of their marginal densities. Equivalently, statistical independence says that the likelihoods of different specific values of $\tilde{y}_i$ do not depend on the values observed for the other $n-1$ random variables. Show that if for all possible y_i and y_j

$$f(y_i, y_j) = f(y_i)f(y_j),$$

then

$$\text{cov}(\tilde{y}_i, \tilde{y}_j) = 0;$$

that is, independence implies zero covariance. Warning: The reverse is not true; zero covariance does not necessarily imply independence.

ANSWER

1. From the general definition of a covariance in equation (14),

$$\text{cov}(\tilde{y}_i, \tilde{y}_j) = \int_{y_i, y_j} [y_i - E(\tilde{y}_i)]\,[y_j - E(\tilde{y}_j)]\,f(y_i, y_j)dy_i dy_j.$$

Since $\tilde{y}_i$ and $\tilde{y}_j$ are assumed to be independent,

$$\begin{aligned}\text{cov}(\tilde{y}_i, \tilde{y}_j) &= \int_{y_i, y_j} [y_i - E(\tilde{y}_i)]\,[y_j - E(\tilde{y}_j)]\,f(y_i)f(y_j)dy_i dy_j \\ &= \int_{y_i} [y_i - E(\tilde{y}_i)]\,f(y_i)dy_i \int_{y_j} [y_j - E(\tilde{y}_j)]\,f(y_j)dy_j \\ &= [E(\tilde{y}_i) - E(\tilde{y}_i)]\,[E(\tilde{y}_j) - E(\tilde{y}_j)] \\ &= 0.\end{aligned}$$

With all the terms in equation (13) now interpreted, the variance of the return on a portfolio of two securities becomes

$$\sigma^2(\tilde{R}_p) = x_{1p}^2\sigma^2(\tilde{R}_1) + x_{2p}^2\sigma^2(\tilde{R}_2) + 2x_{1p}x_{2p}\sigma_{12}, \qquad n = 2.$$

Following precisely the same arguments for portfolios of three securities, we obtain

$$\begin{aligned}\sigma^2(\tilde{R}_p) = {} & x_{1p}^2\sigma^2(\tilde{R}_1) + x_{2p}^2\sigma^2(\tilde{R}_2) + x_{3p}^2\sigma^2(\tilde{R}_3) \\ & + 2x_{1p}x_{2p}\sigma_{12} + 2x_{1p}x_{3p}\sigma_{13} + 2x_{2p}x_{3p}\sigma_{23}.\end{aligned}$$

The new terms are the variance of the return on security 3 and the covariances between the returns on security 3 and the returns on securities 1 and 2.

PROBLEM II.B.

2. Derive the preceding equation for $\sigma^2(\tilde{R}_p)$ when $n = 3$.

ANSWER

2. Go back to the beginning of Section II.B and retrace the development of the equations, but for the case $n = 3$.

The same arguments also produce the general result that the variance of the return on a portfolio of n securities is the sum of the weighted variances of the returns on the individual securities in the portfolio plus twice the weighted sum of all the different possible pairwise covariances between the returns on individual securities. The weight applied to the variance of the return on security i is the square of the proportion of portfolio funds invested in security i, while the weight applied to the covariance between the returns on securities i and j is the product of the proportions of portfolio funds invested in these two securities. In formal terms, in the n security case, $\sigma^2(\tilde{R}_p)$ is

$$\begin{aligned}\sigma^2(\tilde{R}_p) = {} & x_{1p}^2\sigma^2(\tilde{R}_1) + x_{2p}^2\sigma^2(\tilde{R}_2) + \ldots + x_{np}^2\sigma^2(\tilde{R}_n) \\ & + 2x_{1p}x_{2p}\sigma_{12} + 2x_{1p}x_{3p}\sigma_{13} + \ldots + 2x_{1p}x_{np}\sigma_{1n} \\ & + 2x_{2p}x_{3p}\sigma_{23} + 2x_{2p}x_{4p}\sigma_{24} + \ldots + 2x_{2p}x_{np}\sigma_{2n} \\ & + 2x_{3p}x_{4p}\sigma_{34} + 2x_{3p}x_{5p}\sigma_{35} + \ldots + 2x_{3p}x_{np}\sigma_{3n} \\ & \vdots \\ & + 2x_{n-1,p}x_{np}\sigma_{n-1,n}; \end{aligned} \tag{15}$$

or equivalently,

$$\sigma^2(\tilde{R}_p) = \sum_{i=1}^{n} x_{ip}^2 \sigma^2(\tilde{R}_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} x_{ip}x_{jp}\sigma_{ij}, \tag{16}$$

where, as indicated by equation (15), the double sum

$$\sum_{i=1}^{n-1} \sum_{j=i+1}^{n}$$

is read "for each value of i from $i = 1$ to $i = n - 1$, sum over j from $j = i + 1$ to $j = n$; then sum the results over i from $i = 1$ to $i = n - 1$."

Equation (16) is not the only expression for the variance of the return on a portfolio. For example, from (14), it is clear that the order of the terms in the cross-product that defines a covariance is irrelevant:

$$\begin{aligned}\sigma_{ij} &= E([\tilde{R}_i - E(\tilde{R}_i)]\,[\tilde{R}_j - E(\tilde{R}_j)])\\ &= E([\tilde{R}_j - E(\tilde{R}_j)]\,[\tilde{R}_i - E(\tilde{R}_i)]) = \sigma_{ji}.\end{aligned}$$

It follows that in equation (16)

$$2x_{ip}x_{jp}\sigma_{ij} = x_{ip}x_{jp}\sigma_{ij} + x_{jp}x_{ip}\sigma_{ji},$$

so that an expression for $\sigma^2(\tilde{R}_p)$ equivalent to (16) is

$$\sigma^2(\tilde{R}_p) = \sum_{i=1}^{n} x_{ip}^2 \sigma^2(\tilde{R}_i) + \sum_{i=1}^{n} \sum_{\substack{j=1\\ j\neq i}}^{n} x_{ip}x_{jp}\sigma_{ij}. \tag{17}$$

Here the double sum notation

$$\sum_{i=1}^{n} \sum_{\substack{j=1\\ j\neq i}}^{n}$$

is read "for each value of i from $i = 1$ to $i = n$, sum over j from $j = 1$ to $j = n$, but omitting terms where $j = i$; then sum the results over i from $i = 1$ to $i = n$." Equivalently, the double sum can be read, "sum over all possible combinations of i and j except those where $j = i$."

Equations (16) and (17) still do not exhaust the possibilities. The variance of the return on a security can always be regarded as that return's covariance with itself:

$$\begin{aligned}\sigma^2(\tilde{R}_i) &= E([\tilde{R}_i - E(\tilde{R}_i)]^2)\\ &= E([\tilde{R}_i - E(\tilde{R}_i)]\,[\tilde{R}_i - E(\tilde{R}_i)])\\ &= \sigma_{ii}.\end{aligned}$$

With this notation, the security return variances in equation (17) can be included in the double sum, so that

$$\sigma^2(\tilde{R}_p) = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ip}x_{jp}\sigma_{ij}. \tag{18}$$

The double sum here is read "for each value of i from $i = 1$ to $i = n$, sum over j from $j = 1$ to $j = n$; then sum the results from $i = 1$ to $i = n$"; or equivalently, "sum over all possible combinations of i and j."

Since

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_{ip}x_{jp} = 1.0,$$

equation (18) expresses $\sigma^2(\tilde{R}_p)$ as a weighted average of the n^2 variances and covariances $\sigma_{ij}(i, j = 1, 2, \ldots, n)$. Equation (17) treats the n security return variances embedded in the double sum of (18) separately from the $n(n - 1)$ "true" covariances $\sigma_{ij}, j \neq i$, while equation (16) emphasizes that since $\sigma_{ij} = \sigma_{ji}$, only $n(n - 1)/2$ of the covariances in (17) are different.

Finally, at the moment we are concerned with the variance of the return on a portfolio, but the preceding analysis is general. That is, (11) can be regarded as a general expression for a sum of weighted random variables. Equations (16) to (18) are general expressions for the variance of such a sum, expressed in terms of the weights applied to the individual summands, the variances of the individual summands, and their pairwise covariances.

PROBLEMS II.B

3. Show that

$$\sum_{i=1}^{n} \sum_{j=1}^{n} x_{ip}x_{jp} = 1.0.$$

4. For the case $n = 4$, show that equations (16), (17), and (18) are equivalent expressions for $\sigma^2(\tilde{R}_p)$.

5. Let $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n$ be arbitrary random variables.

(a) What is the variance of their sum?

(b) What is the variance of their sample mean

$$\tilde{\bar{y}} = \frac{\tilde{y}_1 + \tilde{y}_2 + \ldots + \tilde{y}_n}{n}\,?$$

(c) What is $E(\tilde{\bar{y}})$ in terms of $E(\tilde{y}_i), i = 1, 2, \ldots, n$?

Note that the sample mean is itself a random variable. That is, the value of $\tilde{\bar{y}}$ varies from one sample to another, since each of the $\tilde{y}_i$, $i = 1, \ldots, n$, varies from one sample to another. Thus, this problem and those that follow are concerned in large part with determining the sampling distribution of the sample mean.

6. Suppose $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n$ are independent random variables. What is the variance of their sum? What is the variance of their sample mean? What is $E(\tilde{\bar{y}})$?

7. Suppose that $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n$ are independent and identically distributed. What is the variance of their sum? What is the variance of their sample mean? What is $E(\tilde{\bar{y}})$?

8. As an application of the results of Problem II.B.*7*, suppose successive monthly returns on security i are independent and identically distributed with mean $E(\tilde{R}_i)$ and variance $\sigma^2(\tilde{R}_i)$. What are the mean and standard deviation of the distribution of the average return on security i for T months?

9. As another application of the results of Problem II.B.*7*, suppose successive daily continuously compounded returns on security i are independent and identically distributed. What are the mean and standard deviation of the distribution of the continuously compounded monthly return on security i in terms of the mean and standard deviation of the continuously compounded daily return?

10. As an application of the results of Problem II.B.*9*, look again at Problem VI.C.*3* of Chapter 1.

ANSWERS

3.

$$\sum_{i=1}^{n} x_{ip} = 1 \quad \text{and} \quad \sum_{j=1}^{n} x_{jp} = 1.$$

Therefore

$$\left(\sum_{i=1}^{n} x_{ip}\right) \cdot \left(\sum_{j=1}^{n} x_{jp}\right) = 1.$$

But

$$\sum_{i=1}^{n} x_{ip} \cdot \sum_{j=1}^{n} x_{jp} = \sum_{i=1}^{n}\sum_{j=1}^{n} x_{ip}x_{jp}.$$

4. Do it.

5. With $\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_n$ as random variables,

(a)

$$\sigma^2\left(\sum_{i=1}^{n} \tilde{y}_i\right) = E\left[\left(\sum_{i=1}^{n} \tilde{y}_i - E\left(\sum_{i=1}^{n} \tilde{y}_i\right)\right)^2\right]$$

$$= E\left[\left(\sum_{i=1}^{n} (\tilde{y}_i - E(\tilde{y}_i))\right)^2\right]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} E[(\tilde{y}_i - E(\tilde{y}_i))(\tilde{y}_j - E(\tilde{y}_j))]$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \operatorname{cov}(\tilde{y}_i, \tilde{y}_j).$$

Thus the variance of a sum of random variables is just the sum of all the pairwise covariances, which also includes, of course, the n variances $\sigma^2(\tilde{y}_i)$, $i = 1, \ldots, n$.

(b) From equation (18), with each $x_{ip} = \dfrac{1}{n}$,

$$\sigma^2(\tilde{\bar{y}}) = \sigma^2\left(\frac{1}{n}\sum_{i=1}^{n} \tilde{y}_i\right)$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{n}\frac{1}{n}\sigma_{ij}$$

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij}.$$

Alternatively, this result follows from the answer to (a) and the fact that the sample mean is just the sum of random variables treated in (a) multiplied by the constant $1/n$.

(c)

$$E(\tilde{\bar{y}}) = E\left(\frac{1}{n}\sum_{i=1}^{n} \tilde{y}_i\right) = \frac{1}{n}\sum_{i=1}^{n} E(\tilde{y}_i).$$

6. If $\tilde{y}_1, \ldots, \tilde{y}_n$ are independent, $\sigma_{ij} = 0$, $i \neq j$. Therefore

$$\sigma^2\left(\sum_{i=1}^{n} \tilde{y}_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij} = \sum_{i=1}^{n} \sigma_{ii} = \sum_{i=1}^{n} \sigma^2(\tilde{y}_i)$$

$$\sigma^2\left(\frac{1}{n}\sum_{i=1}^{n} \tilde{y}_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} \sigma_{ij} = \frac{1}{n^2}\sum_{i=1}^{n} \sigma_{ii} = \frac{1}{n^2}\sum_{i=1}^{n} \sigma^2(\tilde{y}_i).$$

Thus, the variance of a sum of independent random variables is the sum of the variances of the component variables, while the variance of the sample mean is $(1/n)^2$ times the sum of the variances. Finally,

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i\right) = \frac{1}{n}\sum_{i=1}^{n}E(\tilde{y}_i).$$

7. If $\tilde{y}_1, \ldots, \tilde{y}_n$ are identically distributed, $\sigma^2(\tilde{y}_i) = \sigma^2(\tilde{y}_j)$ for all i and j; equivalently, $\sigma^2(\tilde{y}_i) = \sigma^2(\tilde{y})$, $i = 1, \ldots, n$. Moreover, $E(\tilde{y}_1) = E(\tilde{y})$, $i = 1, \ldots, n$. Then

$$\sigma^2\left(\sum_{i=1}^{n}\tilde{y}_i\right) = \sum_{i=1}^{n}\sigma^2(\tilde{y}_i) = n\sigma^2(\tilde{y})$$

$$\sigma^2\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i\right) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2(\tilde{y}_i) = \frac{n}{n^2}\sigma^2(\tilde{y}) = \frac{\sigma^2(\tilde{y})}{n}$$

$$E\left(\frac{1}{n}\sum_{i=1}^{n}\tilde{y}_i\right) = \frac{1}{n}\sum_{i=1}^{n}E(\tilde{y}_i) = \frac{n}{n}E(\tilde{y}) = E(\tilde{y}).$$

These are important results. Thus, suppose $\tilde{y}_i$, $i = 1, \ldots, n$ are n independent drawings of a random variable $\tilde{y}$, and we want to use the sample to estimate the population mean $E(\tilde{y})$. If we use the sample mean

$$\tilde{\tilde{y}} = \frac{\tilde{y}_1 + \ldots + \tilde{y}_n}{n} \tag{19}$$

as the estimator of the population mean $E(\tilde{y})$, then the preceding results tell us that the estimator is unbiased, which means that $E(\tilde{\tilde{y}})$, the mean of the sampling distribution of the sample mean $\tilde{\tilde{y}}$, is equal to $E(\tilde{y})$, the mean of the distribution of $\tilde{y}$. Moreover, since $\sigma^2(\tilde{\tilde{y}}) = \sigma^2(\tilde{y})/n$, the larger is the sample size n, the more tightly packed the sampling distribution of $\tilde{\tilde{y}}$ about its mean $E(\tilde{\tilde{y}}) = E(\tilde{y})$. In intuitive terms, the larger the sample size on which $\tilde{\tilde{y}}$ is based, the more reliable is the sample mean $\tilde{\tilde{y}}$ as an estimator of $E(\tilde{y})$. In the limit—that is, as n becomes arbitrarily large—$\sigma(\tilde{\tilde{y}})$ approaches zero, so that the sampling distribution of the sample mean becomes arbitrarily tightly packed about $E(\tilde{\tilde{y}}) = E(\tilde{y})$.

The preceding paragraph introduces some new statistical terms whose definitions should be emphasized. A procedure for estimating a parameter from a hypothetical sample is called an estimator. For example, the sample mean $\tilde{\tilde{y}}$ defined in equation (19) is an estimator of the population mean $E(\tilde{y})$. The value $\bar{y}$ of $\tilde{\tilde{y}}$ obtained from a specific sample $y_1, \ldots, y_n$ is called an estimate of the population mean. The properties of an estimator are described by its probability distribution, which is usually called its sampling distribution. The estimate obtained from a specific sample is a drawing from the distribution of the estimator.

One property that an estimator might have is unbiasedness. This means that the mean or expected value of the estimator is equal to the value of the parameter being estimated. Thus, the sample mean $\tilde{\tilde{y}}$ is an unbiased estimator of the population mean $E(\tilde{y})$, since $E(\tilde{\tilde{y}}) = E(\tilde{y})$.

Another example of an estimator is the sample variance

$$s^2(\tilde{y}) = \sum_{i=1}^{n}(\tilde{y}_i - \tilde{\tilde{y}})^2/(n-1),$$

which is an estimator of the population variance $\sigma^2(\tilde{y})$. Now that we know what unbiasedness means, we can state (without proof) that the purpose of dividing by $n - 1$ rather than n is to ensure that the sample variance is an unbiased estimator of the population variance; that is, dividing the sum of squares by $n - 1$ leads to the result that $E[s^2(\tilde{y})] = \sigma^2(\tilde{y})$. We might also note (without proof) that the sample variance has the desirable property that the larger the sample size, the more tightly the sampling distribution is packed about $\sigma^2(\tilde{y})$.

8.
$$E(\tilde{\tilde{R}}_i) = E(\tilde{R}_i)$$

$$\sigma^2(\tilde{\tilde{R}}_i) = \frac{1}{T}\sigma^2(\tilde{R}_i)$$

$$\sigma(\tilde{\tilde{R}}_i) = \sqrt{\frac{1}{T}}\,\sigma(\tilde{R}_i).$$

Note again that the distribution of the average return has a smaller standard deviation than the distribution of the return itself; the larger the sample size T, the smaller the standard deviation of the average return.

9. Suppose there are T days in the month. If $\tilde{r}_{it}$ is the simple return for day t, then the continuously compounded return for day t is $ln\,(1 + \tilde{r}_{it})$. From equation (17) of Chapter 1, the monthly continuously compounded return, $ln\,(1 + \tilde{R}_i)$ is related to the daily continuously compounded returns as

$$ln\,(1 + \tilde{R}_i) = \sum_{t=1}^{T} ln\,(1 + \tilde{r}_{it}).$$

Let

$$E(ln\,(1 + \tilde{r}_{it})) = \mu, \qquad t = 1, \ldots, T$$

$$\sigma^2(ln\,(1 + \tilde{r}_{it})) = \sigma^2, \qquad t = 1, \ldots, T.$$

From Problem II.B.7,

$$E(ln\,(1+\tilde{R}_i)) = T\mu$$
$$\sigma^2(ln\,(1+\tilde{R}_i)) = T\sigma^2$$
$$\sigma(ln\,(1+\tilde{R}_i)) = \sqrt{T}\,\sigma.$$

10. The answer to Problem II.B.*9* above tells us that if successive daily returns are independent and identically distributed, the standard deviation of the monthly returns is approximately the square root of the number of trading days times the standard deviation of the daily returns. Thus, the results of Problem VI.C.*3* of Chapter 1 are consistent with a world where daily returns are independent and identically distributed.

III. Portfolio Risk and Security Risk

The preceding results allow some simple insights into the measurement of risk when probability distributions of returns on portfolios are normal. In such a world, knowledge of its mean and variance is sufficient to describe completely the probability distribution of the return on a portfolio, and comparisons of portfolios can be made solely in terms of the means and variances of their returns. Thus, a portfolio model for a world where portfolio return distributions are normal is called a *two-parameter model.*

In this book, it is also assumed that investors like expected portfolio return but are risk-averse, which in a two-parameter world means that they are risk-averse with respect to variance of portfolio return; the most preferred portfolio among all those with the same level of expected return is the one with the lowest variance of return. In short, in portfolio models based on normal return distributions, the risk of a portfolio is measured by the variance of its return, and investors are assumed to dislike variance of portfolio return.

It is tempting to jump to the conclusion that the risk of a security is also measured by the variance of its return. In portfolio theory, however, the presumption is that the primary concern in the investment decision is the distribution of the return on the portfolio. Investors look at individual securities only in terms of their effects on distributions of portfolio returns. In a two-parameter world, an investor looks at an individual security in terms of its contributions to the mean and variance of the distribution of the return on his portfolio.

The mean or expected return on a portfolio is just the weighted average of

the expected returns on the securities in the portfolio. The contribution of a security to the expected return on a portfolio is $x_{ip}E(\tilde{R}_i)$, the expected return on the security weighted by the proportion of portfolio funds invested in the security.

From inspection of equations (16) to (18), it is clear that the contribution of a security to the variance of a portfolio's return is a somewhat more complicated matter. One important point, emphasized by writing $\sigma^2(\tilde{R}_p)$ as in equation (17), is that when the number of securities n in the portfolio is large, individual security return variances are much less numerous in $\sigma^2(\tilde{R}_p)$ than are covariances. In particular, $\sigma^2(\tilde{R}_p)$ contains only n terms for the security return variances, whereas there are $n(n-1)$ covariances. For example, with a portfolio of 50 securities, $\sigma^2(\tilde{R}_p)$ contains 50 variance terms and 2,450 covariance terms.

The large number of covariances relative to security return variances in $\sigma^2(\tilde{R}_p)$ does not in itself imply that the covariances dominate the variances in the determination of $\sigma^2(\tilde{R}_p)$. Relative magnitudes are also important. This question is studied empirically in Chapter 7, where the portfolio model is presented in detail. To foreshadow the results, at least for NYSE common stocks, pairwise covariances between individual security returns are nontrivial in magnitude relative to variances of individual security returns. In portfolios of 20 or more common stocks, $\sigma^2(\tilde{R}_p)$ is primarily determined by the pairwise covariances of security returns.

All this assumes that the portfolios are diversified in the sense that funds are spread fairly evenly across the securities in the portfolio, or at least that funds are not concentrated in a few securities. For example, if most of the portfolio is in one security, then that security's return variance is important in determining the variance of the return on the portfolio, regardless of how many other securities are also included in the portfolio.

We have strayed. What about the risk of a specific security? What is the contribution of an individual security to the variance of the return on a portfolio? To study this question, it is convenient to work with equation (18) and to rewrite it as

$$\sigma^2(\tilde{R}_p) = \sum_{i=1}^{n} x_{ip}\left(\sum_{j=1}^{n} x_{jp}\sigma_{ij}\right). \tag{20}$$

In equation (20), $\sigma^2(\tilde{R}_p)$ can be interpreted as the sum of n terms, one for each security in the portfolio. The term for security i is

$$x_{ip}\left(\sum_{j=1}^{n} x_{jp}\sigma_{ij}\right), \qquad i = 1, 2, \ldots, n.$$

This is the contribution of security i to the variance of the return on portfolio p. This contribution of security i to $\sigma^2(\tilde{R}_p)$ is itself made up of two parts: x_{ip}, the proportion of portfolio funds invested in security i, and

$$\sum_{j=1}^{n} x_{jp}\sigma_{ij}, \tag{21}$$

the weighted average of the pairwise covariances between the return on security i and the returns on each of the n securities (including security i) in the portfolio. If we call this weighted average of covariances the risk of security i in portfolio p, then equation (20) says that the risk of p, as measured by the variance of its return, is the weighted average of the risks of the securities in the portfolio where the risk of security i in portfolio p is weighted by the proportion of portfolio funds invested in this security.

There are two points in this analysis that should be emphasized. First, to be precise we must always say "the risk of security i in portfolio p" since the risk of a given security is different for different portfolios. That is, the pairwise covariances σ_{ij} in (21) are parameters of the joint distribution of security returns and thus are the same for all portfolios. The weights $x_{jp}, j = 1, 2, \ldots, n$, vary from portfolio to portfolio, however, and this is why the risk of security i, as measured by the weighted average of pairwise covariances in (21), is different for different portfolios.

Second, one of the terms in the risk of security i in portfolio p is the variance of the return on that security, $\sigma^2(\tilde{R}_i) = \sigma_{ii}$, which is weighted by x_{ip}. There are, however, $n - 1$ covariance terms in (21). If the covariances are not trivial in magnitude relative to $\sigma^2(\tilde{R}_i)$, then in a diversified portfolio the risk of security i is determined primarily by the covariances of its return with the returns on each of the other $n - 1$ securities in the portfolio.

Finally, expression (21) can be put into a form that provides a natural introduction to the next chapter. In particular,

$$\sum_{j=1}^{n} x_{jp}\sigma_{ij} = \operatorname{cov}(\tilde{R}_i, \tilde{R}_p). \tag{22}$$

That is, the risk of security i in portfolio p, as described by (21), is also the covariance between the return on the security and the return on the portfolio.

PROBLEMS III

1. Derive equation (22).

2. Show that, in general, the covariance of a random variable $\tilde{y}$ with a random variable $\tilde{z} = \sum_{i=1}^{n} a_i\tilde{z}_i$ which is itself a sum of weighted random variables is the weighted sum of the pairwise covariances:

$$\operatorname{cov}(\tilde{y}, \tilde{z}) = \operatorname{cov}\left(\tilde{y}, \sum_{i=1}^{n} a_i\tilde{z}_i\right)$$

$$= \sum_{i=1}^{n} a_i \operatorname{cov}(\tilde{y}, \tilde{z}_i).$$

ANSWERS

1. The steps are as follows:

$$\operatorname{cov}(\tilde{R}_i, \tilde{R}_p) = \operatorname{cov}\left(\tilde{R}_i, \sum_{j=1}^{n} x_{jp}\tilde{R}_j\right) \tag{22a}$$

$$= E\left([\tilde{R}_i - E(\tilde{R}_i)]\left[\sum_{j=1}^{n} x_{jp}\tilde{R}_j - E\left(\sum_{j=1}^{n} x_{jp}\tilde{R}_j\right)\right]\right) \tag{22b}$$

$$= E\left([\tilde{R}_i - E(\tilde{R}_i)]\left[\sum_{j=1}^{n} x_{jp}\tilde{R}_j - \sum_{j=1}^{n} x_{jp}E(\tilde{R}_j)\right]\right) \tag{22c}$$

$$= E\left([\tilde{R}_i - E(\tilde{R}_i)]\left[\sum_{j=1}^{n} x_{jp}(\tilde{R}_j - E(\tilde{R}_j))\right]\right) \tag{22d}$$

$$= E\left(\sum_{j=1}^{n} x_{jp}[\tilde{R}_i - E(\tilde{R}_i)]\,[\tilde{R}_j - E(\tilde{R}_j)]\right) \tag{22e}$$

$$= \sum_{j=1}^{n} x_{jp}E([\tilde{R}_i - E(\tilde{R}_i)]\,[\tilde{R}_j - E(\tilde{R}_j)]) \tag{22f}$$

$$= \sum_{j=1}^{n} x_{jp}\sigma_{ij}. \tag{22g}$$

In going from (22a) to (22b), we make use of the definition of a covariance as an expected value. The step from (22b) to (22c) makes use of the result that the expectation of a sum of weighted random variables is the sum of the weighted expectations, which is also used to go from (22e) to (22f). The final step from (22f) to (22g) then makes use of the definition of σ_{ij} as an expected value.

2. Except for a trivial change in notation, the steps are (22a) to (22g). The only point of this problem is to get you to recognize the generality of the development of equations (22a) to (22g).

It is also convenient to define

$$\beta_{ip} = \frac{\operatorname{cov}(\tilde{R}_i, \tilde{R}_p)}{\sigma^2(\tilde{R}_p)}, \tag{23}$$