I. INTRODUCTION

In this chapter we begin to develop a model for consumption-investment decisions by individuals under conditions of uncertainty. The general problem can be described as follows: Consider an individual who must make a consumption-investment decision at each of \( r \) discrete points in his lifetime. At the first decision point he has a quantity of wealth \( w_1 \) that represents the maximum possible level of consumption during time period 1. At the beginning of period 1, \( w_1 \) must be split between current consumption \( c_1 \) and investment \( h_1 = w_1 - c_1 \). At the beginning of period 2, the individual's wealth level is

\[
\omega_2 = h_1 (1 + R_2) = (w_1 - c_1) (1 + R_2),
\]

It is assumed here that the reader is familiar with elementary statistical concepts, such as expected value, probability distribution, and variance. A brief review of these concepts is provided in the Appendix to this chapter.
where $R_t$ is the one-period or percentage return at the beginning of period 2 per dollar of investment at the beginning of period 1. The return $R_t$ is assumed to be a random variable, that is, the observed value of $R_t$ is drawn from some probability distribution. Thus the wealth level $\bar{w}_t$ is also a random variable.

At the beginning of period 2, $\bar{w}_2$ must in turn be allocated to consumption and investment, and the consumption-investment decision problem is faced at the beginning of each subsequent period until period $\tau$, the last period of the individual's life, at which time the entire available wealth $\bar{w}_\tau$ is consumed and a bequest is considered consumption. The individual is assumed to derive satisfaction only from consumption, and his problem is to map out a consumption-investment strategy that maximizes the level of satisfaction provided by anticipated consumption over his lifetime.

Under uncertainty the decision problem is of course complicated by the fact that the actual lifetime consumption sequence is to some extent unpredictable, because, as indicated above, the wealth levels produced through time by any given investment strategy are usually random variables. Thus in order to solve the individual’s sequential consumption-investment problem, we need a theory of choice under uncertainty that defines the criteria that the individual uses in choosing among different probability distributions of lifetime consumption. Developing such a theory of choice is the purpose of this chapter.

II. THE EXPECTED UTILITY MODEL: GENERAL AXIOMATIC TREATMENT

The theory of choice under uncertainty that we apply to the consumption-investment problem is the “expected utility hypothesis.” In general terms, the expected utility hypothesis states that when faced with a set of mutually exclusive actions, each involving its own probability distribution of “outcomes,” the individual behaves as if he attaches numbers called, purely for convenience, utilities to each outcome and then chooses that action whose associated probability distribution of outcomes provides maximum expected utility.

In general, the one-period or percentage return at $t + 1$ on the investment $h_t$ undertaken at $t$ is

$$R_{t+1} = \frac{\bar{w}_{t+1} - h_t}{h_t}.$$  

Tildes (~) are used throughout this and following chapters to denote random variables. When we talk about a specific observed value of such a variable, however, the tilde is dropped.

The expected or average value of utility of any probability distribution of outcomes is computed like any other expected value; that is, the utility of each possible outcome
In the $\tau$ period consumption-investment problem, an outcome is a complete sequence of lifetime consumptions $C = (c_1, c_2, \ldots, c_\tau)$, and an action is a $\tau$ period consumption-investment strategy that produces a probability distribution for different possible lifetime consumption sequences. But because the consumption-investment problem is just one possible application of the expected utility model, we initially present the model in the most general terms and then turn to its applications that are of major interest in finance.\footnote{The reader who finds the general treatment in this section overly abstract is nevertheless encouraged to continue. Later sections provide concrete applications of the expected utility model, and these give some perspective on the initial development of the model presented here. (Indeed the reader may find it helpful to reconsider the material in this section after reading through the chapter.) Our mode of presentation is designed to provide a more general understanding of the expected utility model than would be obtained only from its specific application to the consumption-investment problem.}

Thus we envisage a decision maker faced with a set $S$ of prospects, whose characteristics are at this point purposely left unspecified, from which a choice must be made. $S$ includes all the prospects that are relevant for the decision at hand. Moreover, $S$ can include both "elementary prospects," for each of which there is only one possible outcome, and "random prospects," which are probability distributions of mutually exclusive elementary prospects.

At this point we could simply assume that the decision maker's behavior conforms to the expected utility model; that is, he behaves as if he assigned utilities to elementary prospects and then ranked random prospects on the basis of expected utility. Alternatively, we can show that behavior in conformity with the expected utility model is implied by a more basic set of axioms concerning how the individual ranks outcomes and probability distributions of outcomes, just as behavior in conformity with the ordinary utility model in Chapter 1 can be shown to follow from a more basic set of axioms concerning consumer choice under conditions of certainty. Because an axiomatic treatment can help to produce a better understanding of the model, this approach is taken here.

is weighted (multiplied) by the probability of the outcome, and the sum of these products over all possible outcomes is the expected or average value of utility for this probability distribution of outcomes.

Note that we say that the individual behaves as if he were an expected utility maximizer. As always, we do not presume that he formally goes through the optimization process prescribed by the theory. Rather his observable behavior is assumed to be as if his decision process conformed to the model (compare Chap. 1, Sec. I). As usual, however, we use words a little loosely and talk about an individual maximizing his utility. But such statements are always meant to be interpreted in an "as if" sense.
II.A. The Axiom System

The set of axioms we use is as follows:

Axiom 1 (Comparability). The individual can define a complete preference ordering over the set of prospects in \( S \); that is, for any two prospects \( x \) and \( y \) in \( S \), he can say that \( x > y \) or \( y > x \) or \( x \sim y \).¹

Axiom 2 (Transitivity). The ordering of prospects assumed in Axiom 1 is also completely transitive. For example, \( x > y \) and \( y > z \) imply \( x > z \); or \( x \sim y \) and \( y \sim z \) imply \( x \sim z \); or \( x \sim y \) and \( y > z \) imply \( x > z \); and so on.

Axiom 3 (Strong Independence). If \( x \sim y \), then for any third prospect \( z \) in \( S \), \( G(x,z;\alpha) \sim G(y,z;\alpha) \). Here \( G(x,z;\alpha) \) represents a gamble, that is, a random prospect, in which the individual gets either \( x \), with probability \( \alpha \), or \( z \), with probability \( 1 - \alpha \), and \( G(y,z;\alpha) \) likewise represents a gamble that produces either \( y \) or \( z \), with probabilities \( \alpha \) and \( 1 - \alpha \). We also assume that if \( x > y \), then \( G(x,z;\alpha) > G(y,z;\alpha) \); or if \( x \geq y \), then \( G(x,z;\alpha) \geq G(y,z;\alpha) \). In short, the rankings of two prospects are not changed when each is combined in the same way into a gamble or probability distribution involving a common third prospect.

Axiom 4. If the prospects \( x, y, \) and \( z \) are such that either \( x > y \geq z \) or \( x \geq y > z \), there is a unique \( \alpha \) such that

\[
y \sim G(x,z;\alpha).
\]

Axiom 5. If \( x \geq y \geq z \) and \( x \geq u \geq z \), and \( y \sim G(x,z;\alpha_1) \) and \( u \sim G(x,z;\alpha) \), then \( \alpha_1 > \alpha \) implies \( y > u \) and \( \alpha_1 = \alpha \) implies \( y \sim u \).

Axioms 1 and 2 are analogous to the axioms of comparability and transitivity assumed in the theory of choice under certainty in Chapter 1. And as in the certainty model, these two axioms are sufficient to define a consistent preference ordering for all prospects in \( S \). But if in addition we wish to say that the ordering of random prospects in \( S \) is according to expected utility, additional behavioral postulates are required. Axioms 3 to 5 provide one possible set of such additional behavioral restrictions.

¹The notation \( x > y \) is read "\( x \) is strictly preferred to \( y \)," and \( x \sim y \) is read "\( x \) and \( y \) are regarded as equivalent." Likewise \( x \geq y \) is read "\( x \) is at least equivalent to \( y \)."

Note that in the statements of the axioms there is no restriction on whether the prospects are elementary or random. For example, in Axiom 1, \( x \) and \( y \) can be elementary or random prospects.
Axiom 3 is the "strong independence" axiom. Intuitively, the importance of the axiom to the expected utility model is easy to see. If the expected utility rule is to be applicable—that is, if the utility of a random prospect is to be just the weighted sum of the utilities of its component elementary prospects, with weights equal to the probabilities of obtaining each of the elementary prospects—then necessarily the decision maker's attitudes toward particular (mutually exclusive) prospects cannot be affected when these are combined in various ways into random prospects. And this is the direct assumption of Axiom 3.

In judging the reasonableness of the strong independence axiom as a description of behavior, one must keep in mind that a random prospect is a probability distribution of mutually exclusive elementary prospects; ultimately one obtains only one of the elementary prospects in the probability mixture. Thus although in general one may not be willing to say that an ordering of objects is unchanged when the objects are combined into some mixture, in the case of a probability mixture of mutually exclusive outcomes this assumption may seem more reasonable.

For example, the elementary prospects under consideration may be bundles of consumption goods with all the goods in a given bundle to be consumed by the decision maker. The decision maker's rankings of different elementary prospects (bundles) is usually affected by the degrees of complementarity and substitutability among the goods in a particular bundle, so that the separate ranks of each good may not be a good indication of the rank of the bundle. (Or in utility terms, it is not usually possible to rank a consumption bundle by assigning utilities to each good, without regard to the quantities of other goods in the bundle, and then obtain the utility of the bundle as the sum of the separate utilities of each good.) But in ranking probability distributions of such bundles, because a particular distribution ultimately yields only one of its component bundles, one may well find it reasonable to assume the strong independence axiom for such probability mixtures. Thus, for example, one may be willing to say that if $x$ and $y$ are two commodity bundles such that $x \sim y$, then for any third bundle $z$, the individual is indifferent between (1) a gamble in which either $x$ is obtained with probability $\alpha$ or $z$ with probability $1 - \alpha$ and (2) the corresponding gamble in which either $y$ is obtained with probability $\alpha$ or $z$ with probability $1 - \alpha$.

Intuitively it is clear that ranking random prospects, which are just probability distributions of elementary prospects, according to expected utility requires a utility function in which the differences between the utility levels assigned to different elementary prospects have some meaning; that is, if the utility of a random prospect is to be just the expected or average value of the separate utilities of each of its component elementary
prospects, differences in utility levels must have some meaning. In later discussions we shall see that Axioms 4 and 5 play a critical role in defining utility functions for which this is the case.

Before moving on, however, we should note that the axiom system that we have chosen is far from the least restrictive set of behavioral postulates that could be shown to lead to the expected utility rule. We chose this particular set of axioms to simplify the derivation of the expected utility rule and to make this derivation contribute as much as possible to a fuller understanding of the model. The reader who is interested in a derivation from less restrictive assumptions is encouraged to seek out the references at the end of the chapter, especially Herstein and Milnor [2].

II.B. Derivation of the Expected Utility Rule

To show that the expected utility rule follows from the axioms, we must show that the axioms imply two things:

1. There exists an order-preserving utility function; that is, if \( U(\cdot) \) is the function, \( U(x) > U(y) \) implies \( x > y \) and \( U(x) = U(y) \) implies \( x \sim y \).

2. The ordering of random prospects given by the function is according to expected utility; that is, \( U(G(x,y; \alpha)) = \alpha U(x) + (1 - \alpha) U(y) \).

For simplicity let us suppose that the set \( S \) of prospects is bounded by two extreme prospects \( a \) and \( b \), such that \( a > b \) and for any prospect \( x \) in \( S \) either

\[
\begin{align*}
\text{a} & \geq x \geq b \\
\text{or} \\
\text{a} & \geq x > b.
\end{align*}
\]

Axioms 4 and 5 can then be used to rank all prospects in \( S \) in terms of the two extreme prospects \( a \) and \( b \). Thus let us define the function \( \alpha(x) \) as the probability such that

\[
x \sim G(a,b; \alpha(x));
\]

that is, \( \alpha(x) \) is the probability value for which the individual is indifferent between (1) obtaining the prospect \( x \) for certain and (2) engaging in a gamble, or equivalently, obtaining a probability distribution, that yields

\[\text{This is of course in contrast with a purely ordinal function, such as those in Chap. 1, or those which would be implied by the first two axioms of the expected utility model, in which only the ordering of prospects provided by the function is meaningful and, except for sign, differences in assigned levels of utility are completely arbitrary.}\]

\[\text{We use the notation } U(\cdot) \text{ when we talk about the function in general terms, that is, without reference to any specific value of its argument.}\]
either prospect \( a \) with probability \( \alpha(x) \) or prospect \( b \) with probability \( 1 - \alpha(x) \). The existence and uniqueness of \( \alpha(x) \) for any \( x \) in \( S \) is guaranteed by Axiom 4, and existence and uniqueness mean that \( \alpha(x) \) is a function defined for all prospects in \( S \). Thus for the prospect \( y \), \( \alpha(y) \) is the probability value such that

\[
y \sim G(a,b; \alpha(y)).
\]

Indeed from Axiom 5 we can see immediately that \( \alpha(x) \) is an order-preserving utility function; that is, \( \alpha(x) > \alpha(y) \) implies \( x > y \), and \( \alpha(x) = \alpha(y) \) implies \( x \sim y \). Thus to show that the expected utility rule follows from the axioms, it remains only to show that the function \( \alpha(x) \) ranks random prospects according to expected utility; that is, consider a random prospect \( G(x,y;\beta) \) in which the individual obtains either \( x \), with probability \( \beta \), or \( y \), with probability \( 1 - \beta \). From Axiom 4 we know that there is always a unique probability \( \alpha(G(x,y;\beta)) \) such that

\[
G(x,y;\beta) \sim G(a,b; \alpha(G(x,y;\beta)));
\]

that is, there is always a probability \( \alpha(G(x,y;\beta)) \) such that the individual is indifferent between (1) engaging in the gamble \( G(x,y;\beta) \) and (2) engaging in the gamble \( G(a,b; \alpha(G(x,y;\beta))) \) in which he obtains either \( a \), with probability \( \alpha(G(x,y;\beta)) \), or \( b \), with probability \( 1 - \alpha(G(x,y;\beta)) \). We already know that \( \alpha(G(x,y;\beta)) \) ranks \( G(x,y;\beta) \) relative to other prospects in \( S \). It remains only to show that

\[
\alpha(G(x,y;\beta)) = \beta \alpha(x) + (1 - \beta) \alpha(y);
\]

that is, the ranking is according to expected utility, where \( \alpha(\cdot) \) is the utility function. And we see that at this stage in the analysis, Axiom 3 (strong independence) begins to play a critical role.

Because

\[ x \sim G(a,b; \alpha(x)), \]

from Axiom 3 we can conclude that

\[
G(x,y;\beta) \sim G(G(a,b; \alpha(x));y;\beta)\quad (5.1)
\]

Here \( G(G(a,b; \alpha(x));y;\beta) \) represents a gamble in which (1) with probability \( \beta \) the individual obtains the prospect \( G(a,b; \alpha(x)) \), which is of course itself a random prospect, or (2) with probability \( 1 - \beta \) he obtains the prospect \( y \). Likewise, because

\[ y \sim G(a,b; \alpha(y)), \]

from Axiom 3 we can again conclude that

\[
G(G(a,b; \alpha(x));y;\beta) \sim G(G(a,b; \alpha(x));G(a,b; \alpha(y));\beta)\quad (5.2)
\]
Applying Axiom 2 (transitivity) to expressions (5.1) and (5.2), we obtain

\[ G(x, y; \beta) \sim G(G(a, b; \alpha(x)), G(a, b; \alpha(y)); \beta). \]  

(5.3)

Now \( G(G(a, b; \alpha(x)), G(a, b; \alpha(y)); \beta) \) is a double gamble in which either (1) with probability \( \beta \) the individual engages in the gamble \( G(a, b; \alpha(x)) \) or (2) with probability \( 1 - \beta \) he engages in the gamble \( G(a, b; \alpha(y)) \). But both of these component gambles involve only the extreme prospects \( a \) and \( b \). Indeed the reader can easily determine that \( G(G(a, b; \alpha(x)), G(a, b; \alpha(y)); \beta) \) is identical with \( G(a, b; \beta\alpha(x) + (1 - \beta)\alpha(y)) \), a gamble in which the individual obtains either the prospect \( a \), with probability \( \beta\alpha(x) + (1 - \beta)\alpha(y) \), or the prospect \( b \), with probability \( 1 - [\beta\alpha(x) + (1 - \beta)\alpha(y)] \). Thus from Equation (5.3) and Axiom 2

\[ G(x, y; \beta) \sim G(a, b; \beta\alpha(x) + (1 - \beta)\alpha(y)). \]

But from Axiom 4

\[ G(x, y; \beta) \sim G(a, b; \alpha(G(x, y; \beta))). \]

And because Axiom 4 also tells us that \( \alpha(G(x, y; \beta)) \) is unique, we must have

\[ \alpha(G(x, y; \beta)) = \beta\alpha(x) + (1 - \beta)\alpha(y). \]

Thus the ranking of random prospects provided by the function \( \alpha(\cdot) \) is indeed according to expected utility.

To review, Axioms 4 and 5 allowed us to define a utility function that ranked prospects in \( S \) in terms of the two extreme prospects \( a \) and \( b \). The strong independence axiom then played a critical role in showing that this utility function \( \alpha(\cdot) \) ranks random prospects according to expected utility.

II.C. Some Properties of the Utility Functions Implied by the Expected Utility Model

It is useful to examine the utility function \( \alpha(\cdot) \) in a little more detail. Note first that we must have

\[ a \sim G(a, b; 1) \quad \text{and} \quad b \sim G(a, b; 0), \]

so that \( \alpha(a) = 1 \) and \( \alpha(b) = 0 \). Thus the function ranges from 0 to 1, and the utility value assigned to any particular prospect depends, from Axiom 4, on how the probabilities of obtaining a gamble involving \( a \) and \( b \) must be balanced to make the individual indifferent between the prospect and the gamble.

We must emphasize, though, that \( \alpha(\cdot) \) is not the only utility function consistent with the expected utility rule. Once we have shown that the axioms imply the expected utility rule, any function that provides the same rankings of prospect as \( \alpha(\cdot) \) and also ranks random prospects according to expected utility is equivalent to \( \alpha(\cdot) \) as a representation of the individual's
tastes. In fact, we now see that any positive linear transformation of \( \alpha(\cdot) \) is equivalent to \( \alpha(\cdot) \).

Thus consider the function

\[
U(\cdot) = \gamma_1 + \gamma_2 \alpha(\cdot), \quad \gamma_2 > 0.
\]

We want to examine the rankings of two arbitrary gambles, \( G(x, y: \beta) \) and \( G(u, z: \phi) \), provided by the functions \( \alpha(\cdot) \) and \( U(\cdot) \). We know that the rankings provided by \( \alpha(\cdot) \) are according to expected utility; that is,

\[
\begin{align*}
\alpha(G(x, y: \beta)) &= \beta \alpha(x) + (1 - \beta) \alpha(y), \quad (5.4) \\
\alpha(G(u, z: \phi)) &= \phi \alpha(u) + (1 - \phi) \alpha(z). \quad (5.5)
\end{align*}
\]

On the other hand, if we use the function \( U(\cdot) \) to compute the expected utilities \( E(U) \) of the two gambles, we get.*

For \( G(x, y: \beta) \):

\[
E(U) = \beta U(x) + (1 - \beta) U(y)
\]

\[
= \beta[\gamma_1 + \gamma_2 \alpha(x)] + (1 - \beta)[\gamma_1 + \gamma_2 \alpha(y)]
\]

\[
= \gamma_1 + \gamma_2[\beta \alpha(x) + (1 - \beta) \alpha(y)]; \quad (5.6)
\]

For \( G(u, z: \phi) \):

\[
E(U) = \gamma_1 + \gamma_2[\phi \alpha(u) + (1 - \phi) \alpha(z)]. \quad (5.7)
\]

Comparing Equations (5.6) and (5.7) with Equations (5.4) and (5.5), we see that the expected utility rankings of the two gambles given by the function \( U(\cdot) \) is the same as that given by the function \( \alpha(\cdot) \).

This result is easy to explain. An expected utility is just a linear combination of the utilities of elementary prospects, that is, the expected utility of a given probability distribution of elementary prospects is obtained by first multiplying the probability of each elementary prospect by the utility of the prospect and then summing the resulting products. Thus when we take a positive linear transformation of the function \( \alpha(\cdot) \), the result is the same positive linear transformation of the expected utilities of all random prospects, that is, utilities and expected utilities are all first multiplied by a positive constant \( \gamma_2 \), and then another constant \( \gamma_1 \) is added to each. But these operations leave the rankings of the prospects completely unchanged. And the same kind of reasoning can be used to conclude that in general a nonlinear transformation of the utility function \( \alpha(\cdot) \) does not provide the same expected utility rankings of random prospects as the function \( \alpha(\cdot) \). We summarize these results with the statement that the utility functions implied by the expected utility hypothesis are unique up to positive linear transformations.

* Note that the level of utility to be obtained from a gamble is indeed a random variable: thus the tilde over the \( U \) in \( E(U) \).
In sum, Axioms 1 and 2 would provide an ordinal utility function, say, \( V(\cdot) \), for prospects in the set \( S \). As always, the function would be ordinal in the sense that it would provide only an ordering of prospects in \( S \); any transformation of \( V(\cdot) \), say, \( Z(V(\cdot)) \), that is an increasing function of values of \( V \) would provide a representation of the individual's tastes that is equivalent to \( V(\cdot) \). Adding Axioms 3 to 5, however, has allowed us to imply a finer or more exact calibration of utilities: finer in the sense that random prospects can be ranked on the basis of expected utilities. But the utility functions implied by the expanded axiom set are themselves unique only up to positive linear transformations, and so in a mathematical sense they are not strictly cardinal, that is, unique. Following the common usage of the utility literature, however, we henceforth refer to these utility functions as cardinal.

It is clear that with utility functions that are unique only up to positive linear transformations, levels of utility do not have hedonic meaning, so that, for example, interpersonal comparisons of utility are not possible. The latter require utility functions that are strictly cardinal, that is, unique.

But with the utility functions of the expected utility model, the change in the level of utility from one prospect to another is unique up to a proportionality factor; that is, if \( U(\cdot) \) and \( \alpha(\cdot) \) are two utility functions that provide the same expected utility rankings of all prospects, there must be a positive constant \( \gamma \) such that for any two prospects \( x \) and \( y \) in \( S \), \( U(x) - U(y) = \gamma[\alpha(x) - \alpha(y)] \). This means that with the utility functions implied by the expected utility rule, one can sensibly talk about increasing and decreasing marginal utility, which, as we soon see, is important in developing meaningful notions of risk aversion and risk preference. By way of contrast, recall that with a purely ordinal utility function only the signs of marginal utilities have meaning.

III. THE TIMELESS EXPECTED UTILITY OF WEALTH MODEL

The initial application of the expected utility model is to the problem of choice from among various available "timeless" gambles. Specifically, we assume that at the beginning of period 1 the individual has the opportunity to use his initial wealth \( w \) to engage in gambles whose outcomes, in this case, levels of wealth, are known before the consumption-investment decision for period 1 is made. The gambles are timeless in the sense that no consumption takes place between the time when a gamble is undertaken and the time when its outcome is realized. We assume that the individual's behavior in this decision is in conformity with the axioms of the expected utility model, so that he chooses the gamble or the "portfolio" of gambles that maximizes his expected utility. We want to show first how a utility of wealth function
can be obtained from the axioms and then introduce the notions of risk aversion, risk preference, and risk neutrality.

III.A. Obtaining the Utility of Wealth Function from Axiom 4

In the timeless expected utility of wealth model, elementary prospects are just levels of wealth; random prospects are probability distributions of wealth levels. Moreover, a nonsatiation axiom is usually added to the five axioms presented in the preceding section; that is, it is assumed that more wealth for certain is preferred to less.

To say that the utility of wealth functions implied by the axioms of the expected utility model are unique only up to positive linear transformations is equivalent to saying that two points on a utility function can be assigned arbitrarily, as long as they are assigned in accordance with the nonsatiation axiom. Then, as we now see, the remaining points on the utility function can be determined from Axiom 4.

Thus suppose that we assign

\[ U(\$0) = 0 \quad \text{and} \quad U(\$100) = 100. \]

According to Axiom 4, for any specific wealth level \( w \) between \$0 and \$100, there is a unique probability \( \alpha \) such that

\[ w \sim G(100,0: \alpha). \]

Thus, given that we now know that the axiom system implies ranking of random prospects according to expected utility, the utility of \( w \) is just the expected utility of the gamble

\[ U(w) = \alpha U(100) + (1 - \alpha) U(0) = 100\alpha. \]

Thus one can determine \( U(w) \) for a particular individual by asking the following question: What is the probability \( \alpha \) that would make you indifferent between \( w \) dollars of wealth for certain and a gamble that could result in \$100 of wealth with probability \( \alpha \) and \$0 with probability \( 1 - \alpha \)? Because we know \( U(0) \) and \( U(100) \), the answer to this question defines \( U(w) \).

The utility of a wealth level \( w \) greater than \$100 can be determined by noting that according to Axiom 4 the individual can always define a probability \( \alpha \) such that

\[ 100 \sim G(w,0: \alpha), \]

from which, applying the expected utility rule to \( G \), we can infer that

\[ U'(100) = \alpha U'(w) + (1 - \alpha) U(0), \]

for a specific \( w > \$100. \)

\[ ^8 \text{And the reader may well find it useful to restate these five axioms in terms of the current problem.} \]
Thus one asks the individual the following question: What is the probability \( \alpha \) that would make you indifferent between a wealth level of $100 for certain and a gamble involving a probability \( \alpha \) that wealth will be $w$ and \( (1 - \alpha) \) that it will be $0$? The answer to this question defines \( U(w) \) as

\[
U(w) = \frac{U(100) - (1 - \alpha)U(0)}{\alpha} = \frac{100}{\alpha}.
\]

Thus once two points on an individual's utility of wealth function have been assigned arbitrarily, the remaining points on the function can, in principle, be determined by a series of questions of the type presented above.

III.B. Usual Types of Utility of Wealth Functions: Risk Aversion, Risk Preference, and Risk Neutrality

The nonsatisfaction axiom says that all utility functions must be monotone-increasing functions of wealth. Thus to indicate that marginal utility is always positive, a graph of utility against wealth must have a positive slope at all levels of wealth. Given this basic restriction, there are three general types of utility functions: (1) linear, (2) concave, and (3) convex, which apply respectively to individuals who (1) have neither risk aversion nor risk preference, (2) have risk aversion, and (3) have risk preference.

We can best elaborate on the implications of the individual's attitudes toward risk for the shape of his utility of wealth function by reference to Figure 5.1, which shows the three general types. Let us begin by noting some of the general properties of the three types of functions. A linear utility function (Figure 5.1a) implies constant marginal utility of wealth at all levels of wealth. Mathematically, the first derivative of the function is a constant. A loss in wealth of \( \varepsilon \) decreases utility by exactly the same amount that an equivalent gain of \( \varepsilon \) would increase it. On the other hand, if the individual's utility function is strictly concave (Figure 5.1b), marginal utility is a decreasing function of wealth. The graph of utility against wealth is monotone-increasing, but utility increases with wealth at a slower and slower rate. Mathematically, the first derivative of utility with respect to wealth is positive, but the second derivative is negative. A loss in wealth of \( \varepsilon \) decreases the individual's level of utility more than an equivalent gain of \( \varepsilon \) would increase it. Finally, if the individual's utility of wealth function is strictly convex (Figure 5.1c), marginal utility is an increasing function of wealth; utility increases with wealth at a faster and faster rate. Mathematically, the second derivative of utility with respect to wealth is positive, so that the first derivative is an increasing function of wealth. An increase
in wealth of $e$ increases the individual's level of utility more than an equivalent loss of $e$ would decrease it.\footnote{The statements concerning first and second derivatives hold only if these derivatives exist. It is easy to show that the axioms of the expected utility hypothesis imply the existence of a continuous utility of wealth function, but the function need not be differentiable.}

Suppose now that at the beginning of period 1 the individual's wealth level is $w'$. Suppose that if gamble $A$ is chosen, his wealth $w$ is $2w'$ with probability 0.5 and 0 with probability 0.5. The expected utility of $A$ is

$$E(U) = 0.5U(0) + 0.5U(2w'),$$

which is just halfway between 0 and $U(2w')$ along the vertical axes of each of the three graphs in Figure 5.1.

If we assume that the three graphs in Figure 5.1 represent the utility of wealth functions of three different individuals, what can we say about

Figure 5.1 Utility of Wealth Functions

\[ (a) \]

\[ (b) \]

\[ (c) \]
the attitudes of these three individuals toward $A$? For individual $i$, $i = 1, 2, 3$, define the certainty equivalent level of wealth corresponding to the probability distribution of wealth provided by the gamble $A$ as the level of wealth $B_i$, $i = 1, 2, 3$, such that

$$U_i(B_i) = E(U_i)$$

that is, as far as individual $i$ with utility function $U_i$ is concerned, a probability distribution of wealth with expected utility $E(U_i)$ is exactly equivalent to $B_i$ of wealth obtained for certain. The individual would be indifferent in a choice between $B_i$ for certain and the probability distribution of wealth associated with action $A$.

The relationship between the expected value of wealth provided by a given probability distribution and the certainty equivalent level of wealth for this distribution gives us a way of defining whether the individual is a risk averter, has risk preference, or is risk-neutral. For example, for the individual in Figure 5.1b, the certainty equivalent level of wealth $B_i$ for the probability distribution of $A$ is less than $w' = E(\delta)$, the expected value of wealth provided by $A$. Or in other words, for this individual the money value of $A$ is less than the expected wealth provided by $A$. (Or reading along the vertical axis of the figure, the expected utility of $A$ is less than the utility of the expected wealth provided by $A$.) The individual regards any level of wealth obtained for certain that is greater than $B_i$ as superior to the probability distribution of wealth provided by the gamble $A$. Thus if $A$ were the only gamble available, he would be willing to pay an insurance premium of $w' - K$, ($K \geq B_i$), in order to avoid $A$ and obtain a level of wealth of $K$ for certain.\footnote{Thus this analysis explains in part why people buy insurance, although insurance is in general an unfair gamble; that is, the expected value of the payoff is less than the premium that must be paid. In essence they are willing to pay to avoid the probability distributions of losses that must be faced in the absence of insurance.}

In general, for an individual with a strictly concave utility of wealth function, the certainty equivalent level of wealth associated with a given probability distribution is always less than the expected value of wealth associated with the distribution. Or in other words, the money value of the distribution to the individual is less than the expected value of its payoff. Thus it is convenient to classify such a person as a risk averter.

By contrast, for the individual in Figure 5.1c, whose utility of wealth function is strictly convex, the certainty equivalent level of wealth $B_i$ for the probability distribution of the gamble $A$ is greater than $w' = E(\delta)$, the expected level of wealth provided by $A$. For this individual the money value of $A$ is greater than the expected payoff from $A$. In essence for him the chance of a large gain is more than sufficient to compensate for an equal chance of an equivalent large loss. Moreover, given a strictly convex
utility function, these results hold with respect to all probability distributions of terminal wealth; the individual would always prefer to have the distribution rather than its expected value for certain. Thus it is meaningful to classify such a person as having risk preference.

For the individual in Figure 5.1a, whose utility of wealth function is linear, the certainty equivalent level of wealth $B_1$ for the probability distribution of $A$ is exactly equal to $\bar{w} = E(\bar{w})$; the money value of the distribution is exactly equal to its expected money payoff. Or in other words, the expected utility of the distribution and the utility of its expected value are equal. Moreover, these results hold with respect to all probability distributions of wealth; if his utility function is linear, the individual chooses among probability distributions of wealth solely in terms of their expected values, always choosing the distribution with maximum expected value. The “dispersion” or “riskiness” of the distributions has no effect on his choice. Thus such a person is neutral with respect to risk; he has neither risk aversion nor risk preference.

Finally, for convenience of exposition we have assumed that an individual is a consistent risk averter, risk preferrer, or neutral with respect to risk at all levels of wealth. This is, of course, not necessarily the case, and mixtures of the three types of utility functions are possible. The only real restrictions on the shape of the function are that it must be continuous and its slope must be positive at all levels of wealth. Given these conditions, there is no reason why the function cannot be concave over some regions of wealth, convex over others, and linear over still others.

IV. EXPECTED UTILITY AND THE THEORY OF FINANCE

IV.A. The Multiperiod Expected Utility of Consumption Model

The timeless expected utility of wealth model is useful for introducing important concepts like risk aversion and risk preference. But the basic problem in finance is the allocation of resources through time. Thus in the $T$ period consumption-investment problem, the individual is concerned with his lifetime consumption sequence

$$C_t = (c_1, c_2, \ldots, c_T).$$

In this model, the elementary prospects are the different possible lifetime consumption sequences, that is, the different possible values of $C_t$, and random prospects are probability distributions of lifetime consumption sequences, that is, probability distributions of $C_t$. If we assume that the individual's behavior in solving this problem conforms to the axioms of the expected utility model, when these axioms are, of course, restated in terms
of \( C_r \), then we can infer that the individual's tastes can be represented by a utility function

\[
U(C_r) = U(c_1, c_2, \ldots, c_r),
\]

and the rankings of random prospects are according to expected utility.

If we assume the nonstaisation axiom—that is, holding consumption in other periods constant, more consumption in any given period is preferred to less—then the marginal utility of consumption in any period is positive. Moreover, as in the timeless expected utility of wealth model, concavity, convexity, and linearity of the utility function \( U(C_r) \) respectively imply risk aversion, risk preference, and risk neutrality.

Thus, by definition, strict concavity of the function \( U(C_r) \) says that for any two consumption sequences \( C_r = (c_1, c_2, \ldots, c_r) \) and \( \hat{C}_r = (\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_r) \), and any \( \alpha \) such that \( 0 < \alpha < 1 \),

\[
U(\alpha c_1 + (1 - \alpha) \hat{c}_1, \alpha c_2 + (1 - \alpha) \hat{c}_2, \ldots, \alpha c_r + (1 - \alpha) \hat{c}_r) > \alpha U(c_1, c_2, \ldots, c_r) + (1 - \alpha) U(\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_r); \quad (5.8)
\]

or equivalently

\[
U(\alpha C_r + (1 - \alpha) \hat{C}_r) > \alpha U(C_r) + (1 - \alpha) U(\hat{C}_r), \quad 0 < \alpha < 1.\]

(5.9)

In words, strict concavity of the function \( U \) implies that the utility of a weighted average of two consumption sequences is greater than the weighted average of the utilities of the two sequences. Geometrically, a straight line between any two points on \( U \) lies below the function.

But suppose now that we consider a gamble in which the individual obtains the consumption sequence \( C_r \) with probability \( \alpha \) or the sequence \( \hat{C}_r \) with probability \( 1 - \alpha \). Thus the expected payoff from the gamble is \( \alpha C_r + (1 - \alpha) \hat{C}_r \); its expected utility is \( \alpha U(C_r) + (1 - \alpha) U(\hat{C}_r) \). Then expression (5.9) tells us that when the individual's utility function is concave, the expected utility of the gamble is less than the utility of its expected payoff: the individual is risk-averse in the sense that if given the choice, he would prefer to have the expected payoff for certain rather than engage in the gamble.

---

\( ^{12} \) The notation \( \alpha C_r + (1 - \alpha) \hat{C}_r \) is just a convenient way of summarizing the operations involved in obtaining a weighted average of the two consumption sequences, as written out explicitly in Equation (5.8); that is, for \( i = 1, 2, \ldots, r \) we simply compute \( \alpha c_i + (1 - \alpha) \hat{c}_i \). These are, of course, just the usual conventions for multiplication of a vector by a scalar and for addition of vectors.
On the other hand, when the individual’s utility function is strictly convex,

\[ U(\alpha C_r + (1 - \alpha) \hat{C}_r) < \alpha U(C_r) + (1 - \alpha) U(\hat{C}_r), \quad 0 < \alpha < 1. \]

Thus the expected utility of a gamble that pays \( C_r \) with probability \( \alpha \) and \( \hat{C}_r \) with probability \( 1 - \alpha \) is greater than the utility of the expected payoff: the individual has risk preference in the sense that he would prefer to engage in the gamble rather than obtain its expected payoff for certain.

Finally, when the individual’s utility function is linear,

\[ U(\alpha C_r + (1 - \alpha) \hat{C}_r) = \alpha U(C_r) + (1 - \alpha) U(\hat{C}_r). \]

The individual is risk-neutral in the sense that he is indifferent between (1) engaging in a gamble that pays \( C_r \) with probability \( \alpha \) or \( \hat{C}_r \) with probability \( 1 - \alpha \) or (2) receiving for certain the expected payoff \( \alpha C_r + (1 - \alpha) \hat{C}_r \) from the gamble.

It is well to note that risk aversion (concavity) or risk preference (convexity) in the utility function \( U(C_r) = U(c_1, c_2, \ldots, c_r) \) implies risk aversion or risk preference with respect to consumption in any given period. Thus, for example, if \( U(C_r) \) is concave in \( C_r \), it is also concave with respect to any component \( c_t \) of \( C_r \); that is, holding consumption in other periods constant, \( U(C_r) \) is a concave function of \( c_t \) for any \( t = 1, 2, \ldots, r \). Indeed as a function of \( c_t \), \( U(C_r) \) looks in this case like the utility of wealth function shown in Figure 5.1b. Likewise if \( U(C_r) \) is convex, as a function of consumption in any given period \( U(C_r) \) looks like the utility of wealth function shown in Figure 5.1c.

And these remarks are of more than passing interest. The next two chapters concentrate on two-period consumption-investment models, that is, the case \( r = 2 \). In these two-period models, given some level of consumption at period 1, the individual’s period 1 portfolio decision depends on the shape of his utility function \( U(c_1, c_2) \) as a function of consumption in period 2. Thus, for example, we know now that if the individual is risk-averse, that is, \( U(c_1, c_2) \) is concave in \( (c_1, c_2) \), then \( U(c_1, c_2) \) is a concave function of \( c_1 \), and the individual behaves like a risk averter in choosing among different probability distributions of period 2 consumption associated with different investment decisions at period 1.

Finally, although we say that utility functions that are concave, convex, and linear respectively imply risk aversion, risk preference, and risk neutrality, the direction of causation is, of course, the other way around: The individual’s tastes determine the shape of his utility function. Thus, for example, his utility function is concave because he is risk-averse. It is always well to keep in mind that a utility function is just a convenient way of representing tastes. The individual’s behavior is guided by his tastes,
which we, as outside observers, find convenient to summarize in terms of a utility function.

IV.B. Utilities for Consumption Dollars from Utilities for Consumption Goods

In the problems involving intertemporal allocation of resources that are of most interest in finance it is convenient to deal with utility functions for aggregate consumption, that is, utility functions for dollars of consumption. We must nevertheless recognize, as in Chapter 1, that the individual's tastes for aggregate consumption ultimately derive from his tastes for consumption goods. More specifically, utilities for consumption dollars are just the utilities obtained from optimal allocation of these consumption dollars to goods and services. For example, in the two-period case the utility function \( U(c_1, c_2) \) for aggregate consumption is obtained from the utility function \( V(q_1, q_2) \) for consumption goods as

\[
U(c_1, c_2) = \max_{q_1, q_2} V(q_1, q_2) \tag{5.10a}
\]

subject to the constraints

\[
K^{(1)} \sum_{k=1}^{K(1)} p_{1k}q_{1k} = c_1 \quad \text{and} \quad K^{(2)} \sum_{k=1}^{K(2)} p_{2k}q_{2k} = c_2, \tag{5.10b}
\]

where \( K^{(1)} \) and \( K^{(2)} \) are the numbers of consumption goods available in periods 1 and 2, \( q_{1k} \) and \( q_{2k} \) are quantities of good \( k \) consumed in periods 1 and 2, \( p_{1k} \) and \( p_{2k} \) are the per unit prices of these goods, and \( Q_1 = (q_{11}, q_{12}, \ldots, q_{1K(1)}) \), \( Q_2 = (q_{21}, q_{22}, \ldots, q_{2K(2)}) \) are the vectors of quantities of commodities consumed.

Although the subject is briefly considered again in Chapter 8, for our purposes the detailed relationships between utility functions for consumption goods and utility functions for consumption dollars are not a topic of major concern. We wish, however, to make a few points in passing, and without proof.\(^{11}\)

First, as stated in Equation (5.10), the individual's tastes for aggregate consumption derive from his tastes for consumption goods. Thus if we assume that he can choose among prospects stated in terms of consumption dollars on the basis of expected utility, his tastes with respect to consumption goods must also satisfy the axioms of the expected utility model. In

\(^{14}\) The notation \( \max_{q_1, q_2} V(q_1, q_2) \)

is read "choose \( q_1 \) and \( q_2 \) in such a way as to maximize \( V(q_1, q_2) \)."

\(^{15}\) The reader with a stronger interest in this area is referred to Fama [6].
short, cardinal utilities for aggregate consumption derive from cardinal utilities for consumption goods.

Second, in deriving the utility function \( U(c_1,c_2) \) as in Equation (5.10), we have implicitly assumed that the prices of consumption goods in both periods are known; that is, a utility function for consumption dollars is conditional on some set of prices of consumption goods. Thus rather than writing \( U(c_1,c_2) \), for completeness, we should write \( U(c_1,c_2 | P_1,P_2) \), where \( P_1 = (p_{11},p_{12},\ldots,p_{1,K}) \) and \( P_2 = (p_{21},p_{22},\ldots,p_{2,K}) \) are the vectors of prices of consumption goods in the two periods. Assuming that tastes with respect to consumption goods conform to the axioms of the expected utility model, the function \( U(c_1,c_2 | P_1,P_2) \) could then be used to order random prospects involving uncertainty with respect to both dollar levels of consumption and prices of consumption goods on the basis of expected utility. This possibility is only mentioned here, however, because subsequent chapters bypass almost completely questions concerning the effects of uncertainty in the prices of consumption goods on optimal consumption-investment decisions. But this omission is consistent with the current state of the literature: A most pressing field for future research is accounting for the effects of uncertain prices for consumption goods first on the nature of optimal consumption-investment decisions by individuals and then on the process of price formation in the capital market.

V. CONCLUSION

Our discussion of the expected utility model as a way of representing tastes, that is, as a theory of choice, in a world of uncertainty is now completed. We turn in the next chapter to the other side of the consumption-investment problem, representing the market opportunities available to the individual. In the usual manner, tastes and opportunities are then combined into a theory of consumer-investor equilibrium, that is, a theory of optimal consumption-investment decisions by individuals. The model of consumer-investor equilibrium presented in the next chapter is then the basic building block for the model of capital market equilibrium to be developed in Chapter 7.

REFERENCES

Although the expected utility model dates back at least several centuries to the work of Daniel Bernoulli, the first axiomatic development of the model is due to

An elegant derivation of the model from, apparently, the least restrictive set of axioms is in

First recognition of the importance of the strong independence axiom is due to


An important study of the concept of risk aversion is


The relationships between utility functions for money and utility functions for commodities are treated in


A detailed and, more or less, elementary discussion of the expected utility model is in

APPENDIX

Statistical Review

I. INTRODUCTION

The purpose of this appendix is to present a brief review of some of the statistical concepts that are used in the analysis of uncertainty.

II. EXPECTED VALUES OF WEIGHTED SUMS OF RANDOM VARIABLES

A "random variable" is simply a variable that is subject to a probability distribution; that is, an observed value of the variable represents a drawing from some probability distribution.

The "expected value" of a random variable is just the mean of its probability distribution. For example, consider a bounded and discrete random variable $W$, where boundedness plus discreteness implies that the variable can only take on a finite number $N$ of values $W_i$, $i = 1, 2, \ldots, N$, with associated probabilities $p(W_i)$, $i = 1, 2, \ldots, N$. The expected value of $W$ is then just

$$E(W) = \sum_{i=1}^{N} W_i p(W_i).$$
where \( p(W) \) is the probability that the random variable takes the value \( W \). Thus the expected value is just the weighted average of the different possible values of \( W \), where the probabilities are used as weights. Another way of writing \( E(\bar{W}) \) is

\[
E(\bar{W}) = \sum \bar{W} p(W),
\]

where \( \sum \) is read the sum over all possible values of \( W \).\(^1\)

Consider now two different random variables \( \bar{W}_1 \) and \( \bar{W}_2 \). The expected value of the sum \( (\bar{W}_1 + \bar{W}_2) \) is just

\[
E(\bar{W}_1 + \bar{W}_2) = \sum \sum (W_1 + W_2) p(W_1, W_2),
\]

where \( p(W_1, W_2) \) is the joint probability of \( W_1 \) and \( W_2 \); that is, it provides the probability that a joint drawing from the \( W_1 \) and \( W_2 \) distributions results in a particular pair \( (W_1, W_2) \).

Now the joint probability can always be expressed as the product of an unconditional and a conditional probability in either of two ways:

\[
p(W_1, W_2) = p(W_1)p(W_2 | W_1) = p(W_2)p(W_1 | W_2),
\]

where \( p(W_2 | W_1) \) is the conditional probability of \( W_2 \) given \( W_1 \); that is, it is the probability that a particular \( W_2 \) is observed in the drawing from the \( W_2 \) distribution, given that a particular \( W_1 \) is observed in the drawing from the \( W_1 \) distribution. Similarly, \( p(W_1 | W_2) \) is the conditional probability distribution of \( W_1 \) given that a particular \( W_2 \) is observed in the drawing from the distribution of \( W_2 \). Conditional probability distributions have the usual properties of probability distributions; that is, \( p(W_1 | W_2) \geq 0 \) and \( \sum_{W_1} p(W_1 | W_2) = 1 \).

\(^1\) If the random variable \( W \) is continuous and unbounded, its expected value is defined as

\[
E(\bar{W}) = \int_{-\infty}^{\infty} \bar{W} f(W) \, dW,
\]

where \( f(W) \) is the density function of \( W \).

In general the reader should note that all the concepts presented in this appendix apply equally well to both discrete and continuous random variables. The change from discrete to continuous variables simply involves replacing summations with the appropriate integrals.
Thus the expected value of the sum \( \bar{W}_1 + \bar{W}_2 \) can always be written as
\[
E(\bar{W}_1 + \bar{W}_2) = \sum_{W_1} \sum_{W_2} (W_1 + W_2) p(W_1, W_2)
\]
\[
= \sum_{W_1} \sum_{W_2} W_1 p(W_1) p(W_2 | W_1) + \sum_{W_1} \sum_{W_2} W_2 p(W_2) p(W_1 | W_2)
\]
\[
= \sum_{W_1} W_1 p(W_1) \sum_{W_2} p(W_2 | W_1) + \sum_{W_1} \sum_{W_2} W_2 p(W_2) p(W_1 | W_2)
\]
\[
= \sum_{W_1} W_1 p(W_1) + \sum_{W_2} W_2 p(W_2)
\]
\[
= E(\bar{W}_1) + E(\bar{W}_2).
\]

Thus the expectation of a sum is just the sum of the expectations. And it is important to note that this result holds regardless of the degree of dependence between the two variables.

Using similar arguments, it can also be shown that the expectation of a sum of any number of random variables is just the sum of the individual expectations; that is,
\[
E(\bar{W}_1 + \bar{W}_2 + \cdots + \bar{W}_m) = E(\bar{W}_1) + E(\bar{W}_2) + \cdots + E(\bar{W}_m).
\]

Suppose now that the variable \( \bar{W}_1 \) is weighted by the constant \( A \) and \( \bar{W}_2 \) is weighted by the constant \( B \). The expected value of the weighted sum is then
\[
E(A\bar{W}_1 + B\bar{W}_2) = \sum_{W_1} \sum_{W_2} (AW_1 + BW_2) p(W_1, W_2)
\]
\[
= AE(\bar{W}_1) + BE(\bar{W}_2),
\]
where the reader can fill in the missing steps from the derivation above. Thus the expectation of a weighted sum is the weighted sum of the expectations of the individual variables.

III. THE VARIANCE OF A WEIGHTED SUM OF RANDOM VARIABLES

The variance of a random variable is just the expectation of the squared deviation of the variable from its mean or expected value. Thus
\[
\sigma^2(\bar{W}) = \text{var}(\bar{W}) = E[(\bar{W} - E(\bar{W}))^2] = \sum_{W} (W - E(\bar{W}))^2 p(W).
\]

The variance provides a measure of the degree of dispersion in the probability distribution of a random variable. The more common measure of variability, however, is the standard deviation \( \sigma \), which is just the square root of the variance
\[
\sigma(\bar{W}) = \sqrt{\sigma^2(\bar{W})}.
\]
It is interesting to examine the relationship between the variance and another measure of dispersion, the second moment. Expanding the expression for the variance, we get
\[
\sigma^2(\tilde{W}) = E[(\tilde{W} - E(\tilde{W}))^2] = E[\tilde{W}^2 - 2\tilde{W}E(\tilde{W}) + E(\tilde{W})^2] \\
= E(\tilde{W}^2) - E(\tilde{W})^2.
\]

$E(\tilde{W}^2)$ is called the second moment of the random variable $\tilde{W}$. Thus the variance is just the second moment minus the square of the mean.\(^3\)

If the variable $\tilde{W}$ is weighted by the constant $A$, the variance of the weighted variable $A\tilde{W}$ is just
\[
\sigma^2(A\tilde{W}) = E[(A\tilde{W} - E(A\tilde{W}))^2] \\
= \sum_w A^2[w - E(\tilde{W})]^2 p(w) \\
= A^2 \sigma^2(\tilde{W}).
\]

Thus the variance of a weighted random variable is just the weight squared times the variance of the unweighted random variable.

We now consider the variance of a sum of many random variables; that is, we want
\[
\sigma^2(\tilde{W}_1 + \tilde{W}_2 + \cdots + \tilde{W}_N) \\
= E[(\tilde{W}_1 + \tilde{W}_2 + \cdots + \tilde{W}_N) - E(\tilde{W}_1 + \tilde{W}_2 + \cdots + \tilde{W}_N)^2] \\
= E[(\tilde{W}_1 - E(\tilde{W}_1)) + [\tilde{W}_2 - E(\tilde{W}_2)] + \cdots + [\tilde{W}_N - E(\tilde{W}_N)]^2] \\
= E[\tilde{w}_1^2 + \tilde{w}_2 + \cdots + \tilde{w}_N],
\]

where $\tilde{w}_i = \tilde{W}_i - E(\tilde{W}_i)$. Expanding the squared sum, we get
\[
\sigma^2(\tilde{W}_1 + \tilde{W}_2 + \cdots + \tilde{W}_N) = E \left[ \tilde{w}_1^2 + \tilde{w}_1\tilde{w}_2 + \tilde{w}_1\tilde{w}_3 + \cdots + \tilde{w}_1\tilde{w}_N \right] \\
+ \tilde{w}_2^2 + \tilde{w}_2\tilde{w}_3 + \tilde{w}_2\tilde{w}_4 + \cdots + \tilde{w}_2\tilde{w}_N \\
+ \tilde{w}_3^2 + \tilde{w}_3\tilde{w}_4 + \tilde{w}_3\tilde{w}_5 + \cdots + \tilde{w}_3\tilde{w}_N \\
+ \cdots \\
+ \tilde{w}_N^2 + \tilde{w}_N\tilde{w}_2 + \tilde{w}_N\tilde{w}_3 + \cdots + \tilde{w}_N^2
\]

\(^3\)In general the $k$th moment of $\tilde{W}$ is $E(\tilde{W}^k)$. Thus the mean is the first moment, and the variance is the second moment minus the square of the first.
This, however, is just the expectation of a sum, which is just the sum of the expectations, so that

$$\sigma^2(\bar{W}_1 + \bar{W}_2 + \cdots + \bar{W}_N) = \sum_{k=1}^{N} \bar{W}_k^2 + \sum_{k=1}^{N} \sum_{j \neq k} \bar{W}_k \bar{W}_j.$$

The first term on the right of the equality is just the sum of the individual variances of the \( \overline{W} \)'s. The second term is a sum of terms of the form

$$E(\bar{W}_k \bar{W}_j) = E(\{[\overline{W}_k - E(\overline{W}_k)][\overline{W}_j - E(\overline{W}_j)]\}), \quad j \neq k.$$

This expression is called the “covariance” between \( \overline{W}_k \) and \( \overline{W}_j \) for which we henceforth use the notation \( \sigma_{kj} \) or \( \text{cov} (\overline{W}_k, \overline{W}_j) \). The covariance is a measure of the relationship between the variables \( \overline{W}_k \) and \( \overline{W}_j \). For example, if values of \( \overline{W}_k \) above \( E(\overline{W}_k) \) tend to be associated with values of \( \overline{W}_j \) above \( E(\overline{W}_j) \) and values of \( \overline{W}_k \) below \( E(\overline{W}_k) \) tend to be associated with values of \( \overline{W}_j \) below \( E(\overline{W}_j) \), then \( \text{cov} (\overline{W}_k, \overline{W}_j) \) is positive. In this case we say that there is “positive dependence” in the relationship between the two variables. Similarly a negative covariance implies “negative dependence”; that is, on the average the deviations of \( \overline{W}_k \) and \( \overline{W}_j \) from their respective means tend to have opposite signs.\(^3\)

It is also interesting to note that the order in which the terms in the covariance are written is irrelevant; that is,

$$E(\bar{W}_k \bar{W}_j) = E(\bar{W}_j \bar{W}_k) = \text{cov} (\overline{W}_k, \overline{W}_j) = \text{cov} (\overline{W}_j, \overline{W}_k) = \sigma_{kj} = \sigma_{jk}.$$

Finally, the variance of a random variable can also be regarded as a covariance; it is the covariance of the variable with itself:

$$\sigma^2(\overline{W}_j) = E(\bar{W}_j \bar{W}_j) = \sigma_{jj}.$$

With these comments in mind, it is easy to see that the variance of a sum can be written in many equivalent ways. In particular,

$$\sigma^2 \left( \sum_{i=1}^{N} \overline{W}_i \right) = \sum_{k=1}^{N} \sigma^2(\overline{W}_k) + \sum_{k=1}^{N} \sum_{j \neq k} \sigma_{kj}$$

$$= \sum_{k=1}^{N} \sigma^2(\overline{W}_k) + 2 \sum_{k=1}^{N} \sum_{j=k+1}^{N} \sigma_{kj}$$

$$= \sum_{k=1}^{N} \sum_{j=1}^{N} \sigma_{kj}.$$

\(^3\) The covariance is closely related to a more familiar measure of association, the correlation coefficient. The correlation coefficient is just

$$\text{corr} (\overline{W}_k, \overline{W}_j) = \frac{\text{cov} (\overline{W}_k, \overline{W}_j)}{\sqrt{\sigma^2(\overline{W}_k) \sigma^2(\overline{W}_j)}}.$$
It is also easy to show, and the reader should convince himself, that if each variable $\bar{W}_k$ is weighted by the constant $A_k$, the variance of the weighted sum is

$$\sigma^2 \left( \sum_{k=1}^{N} A_k \bar{W}_k \right) = \sum_{k=1}^{N} A_k^2 \sigma^2(\bar{W}_k) + \sum_{k=1}^{N} \sum_{j=k+1}^{N} A_k A_j \sigma_{kj}$$

$$= \sum_{k=1}^{N} A_k^2 \sigma^2(\bar{W}_k) + 2 \sum_{k=1}^{N} \sum_{j=k+1}^{N} A_k A_j \sigma_{kj}$$

$$= \sum_{k=1}^{N} \sum_{j=k}^{N} A_k A_j \sigma_{kj}.$$ 

Means and variances of weighted sums appear repeatedly in the models in following chapters. The reader should be sure to have a thorough familiarity with these concepts before moving on from here.