CHAPTER 1

Basic Concepts

Sets, Functions, and Variables

In mathematics, sets, functions, and variables are three fundamental concepts. First, a **set** is a well-defined collection of objects. A set is a gathering together into a whole of definite, distinct objects of our perception, which are called elements of the set. Sets are one of the most fundamental concepts in mathematics. Set theory is seen as the foundation from which virtually all of mathematics can be derived. For example, structures in abstract algebra, such as groups, fields, and rings, are sets closed under one or more operations. One of the main applications of set theory is constructing relations. Second, a **function** is a relation between a set of inputs and a set of permissible outputs with the property that each input is related to exactly one output. Functions are the central objects of investigation in most fields of modern mathematics. There are many ways to describe or represent a function. Some functions may be defined by a formula or algorithm that tells how to compute the output for a given input. Others are given by a picture, called the **graph of the function**. A function can be described through its relationship with other functions, for example, as an inverse function or as a solution of a differential equation. Finally, a **variable** is a value that may change within the scope of a given problem or set of operations. In contrast, a **constant** is a value that remains unchanged, though often unknown or undetermined. Variables are further distinguished as being either a dependent variable or an independent variable. Independent variables are regarded as inputs to a system and may take on different values freely. Dependent variables are

JWBT1003-c01 JWBT1003-Foccardi Printer: Courier Westford August 13, 2013 15:56 Trim: 6in × 9in

those values that change as a consequence of changes in other values in the system.

The concepts of sets, functions, and variables are fundamental to many areas of finance and its applications. Starting with the mean-variance portfolio theory of Harry Markowitz in 1952, then the capital asset pricing model of William Sharpe in 1964, the option pricing model of Fischer Black and Myron Scholes in 1973, and the more recent developments in financial econometrics, financial risk management and asset pricing, financial economists constantly use the concepts of sets, functions, and variables. In this chapter we discuss these concepts.

What you will learn after reading this chapter:

- The notion of sets and set operations
- How to define empty sets, union of sets, and intersection of sets.
- The elementary properties of sets.
- How to describe the dynamics of quantitative phenomena.
- The concepts of distance and density of points.
- How to define and use functions and variables.

INTRODUCTION

In this chapter we discuss three basic concepts used throughout this book: sets, functions, and variables. These concepts are used in financial economics, financial modeling, and financial econometrics.

SETS AND SET OPERATIONS

The basic concept in calculus and in probability theory is that of a **set**. A set is a collection of objects called **elements**. The notions of both elements and set should be considered primitive. Following a common convention, let's denote sets with capital Latin or Greek letters: A,B,C, Ω . . . and elements with small Latin or Greek letters: a,b, ω. Let's then consider collections

JWBT1003-c01 JWBT1003-Foccardi Printer: Courier Westford August 13, 2013 15:56 Trim: 6in × 9in

of sets. In this context, a set is regarded as an element at a higher level of aggregation. In some instances, it might be useful to use different alphabets to distinguish between sets and collections of sets.[1]

Proper Subsets

An element a of a set A is said to belong to the set A written as $a \in A$. If every element that belongs to a set A also belongs to a set B, we say that A is contained in B and write: $A \subset B$. We will distinguish whether A is a **proper subset** of B (i.e., whether there is at least one element that belongs to B but not to A) or if the two sets might eventually coincide. In the latter case we write $A \subseteq B$.

In the United States there are indexes that are constructed based on the price of a subset of common stocks from the universe of all common stock in the country. There are three types of common stock (equity) indexes:

1. Produced by stock exchanges based on all stocks traded on the particular exchanges (the most well known being the New York Stock Exchange Composite Index).
2. Produced by organizations that subjectively select the stocks included in the index (the most popular being the Standard & Poor's 500).
3. Produced by organizations where the selection process is based on an objective measure such as market capitalization.

The Russell equity indexes, produced by Frank Russell Company, are examples of the third type of index. The Russell 3000 Index includes the 3,000 largest U.S. companies based on total market capitalization. It represents approximately 98% of the investable U.S. equity market. The Russell 1000 Index includes 1,000 of the largest companies in the Russell 3000 Index while the Russell 2000 Index includes the 2,000 smallest companies in the Russell 3000 Index. The Russell Top 200 Index includes the 200 largest companies in the Russell 1000 Index and the Russell Midcap Index includes the 800 smallest companies in the Russell 1000 Index. None of the indexes include non-U.S. common stocks.

[1]In this book we consider only the elementary parts of set theory which is generally referred to as naive set theory. This is what is needed to understand the mathematics of calculus. However, set theory has evolved into a separate mathematical discipline which deals with the logical foundations of mathematics.

Let us introduce the notation:

A = all companies in the United States that have issued common stock
I_{3000} = companies included in the Russell 3000 Index
I_{1000} = companies included in the Russell 1000 Index
I_{2000} = companies included in the Russell 2000 Index
I_{Top200} = companies included in the Russell Top 200 Index
I_{Midcap} = companies included in the Russell Midcap 200 Index

We can then write the following:

$I_{3000} \subset A$ (every company that is contained in the Russell 3000 Index is contained in the set of all companies in the United States that have issued common stock)
$I_{1000} \subset I_{3000}$ (the largest 1,000 companies contained in the Russell 1000 Index are contained in the Russell 3000 Index)
$I_{Midcap} \subset I_{1000}$ (the 800 smallest companies in the Russell Midcap Index are contained in the Russell 1000 Index)

$I_{Top200} \subset I_{1000} \subset I_{3000} \subset A$
$I_{Midcap} \subset I_{1000} \subset I_{3000} \subset A$

Throughout this book we will make use of the convenient logic symbols ∀ and ∃ that mean respectively, "for any element" and "an element exists such that." We will also use the symbol ⇒ that means "implies." For instance, if A is a set of real numbers and $a \in A$, the notation $\forall a: a < x$ means "for any number a smaller than x" and $\exists a: a < x$ means "there exists a number a smaller than x."

Empty Sets

Given a subset B of a set A, the complement of B with respect to A written as B^C is formed by all elements of A that do not belong to B. It is useful to consider sets that do not contain any elements called **empty sets**. The empty set is usually denoted by $\emptyset$. For example, stocks with negative prices form an empty set.

JWBT1003-c01 JWBT1003-Foccardi Printer: Courier Westford August 13, 2013 15:56 Trim: 6in × 9in

Union of Sets

Given two sets A and B, their **union** is formed by all elements that belong to either A or B. This is written as $C = A \cup B$. For example,

$I_{1000} \cup I_{2000} = I_{3000}$ (the union of the companies contained in the Russell 1000 Index and the Russell 2000 Index is the set of all companies contained in the Russell 3000 Index)

$I_{\text{Midcap}} \cup I_{\text{Top200}} = I_{1000}$ (the union of the companies contained in the Russell Midcap Index and the Russell Top 200 Index is the set of all companies contained in the Russell 1000 Index)

Let $I_{\text{Long lived}}$ be those stocks that existed in the last 30 years.

Intersection of Sets

Given two sets A and B, their **intersection** is formed by all elements that belong to both A and B. This is written as $C = A \cap B$. For example, let

$$I_{\text{S\&P}} = \text{companies included in the S\&P 500 Index}$$

The S&P 500 is a stock market index that includes 500 widely held common stocks representing about 77% of the New York Stock Exchange market capitalization. (**Market capitalization** for a company is the product of the market value of a share and the number of shares outstanding.) Call $I_{\text{Long lived}}$ those stocks that existed in the last 30 years. Then

$I_{\text{S\&P}} \cap I_{\text{Long lived}} = \mathbf{C}$ (the stocks contained in the S&P 500 Index that existed for the last 30 years)

We can also write:

$I_{1000} \cap I_{2000} = \emptyset$ (companies included in both the Russell 2000 and the Russell 1000 Index is the empty set since there are no companies that are in both indexes)

Elementary Properties of Sets

Suppose that the set Ω includes all elements that we are presently considering (i.e., that it is the total set). Three elementary properties of sets are given below:

Property 1. The complement of the total set is the empty set and the complement of the empty set is the total set:

$$\Omega^{C} = \emptyset, \emptyset^{C} = \Omega$$

Property 2. If A,B,C are subsets of Ω, then the distribution properties of union and intersection hold:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$
$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

Property 3. The complement of the union is the intersection of the complements and the complement of the intersection is the union of the complements:

$$(B \cup C)^C = B^C \cap C^C$$
$$(B \cap C)^C = B^C \cup C^C$$

DISTANCES AND QUANTITIES

Calculus describes the dynamics of quantitative phenomena. This calls for equipping sets with a metric that defines distances between elements. Though many results of calculus can be derived in abstract metric spaces, standard calculus deals with sets of ***n*-tuples** of real numbers. In a quantitative framework, real numbers represent the result of observations (or measurements) in a simple and natural way.

n-tuples

An n-tuple, also called an ***n*-dimensional vector**, includes n components: $(a_1, a_2, \ldots, a_n)$. The set of all n-tuples of real numbers is denoted by R^n. The R stands for real numbers.

For example, suppose the monthly rates of return on a hedge fund portfolio in 2011 are as shown in Table 1.1 with the actual return for the S&P 500 (the benchmark index for the hedge fund portfolio manager).[2]

Then the monthly returns, r_{port}, for the hedge fund portfolio can be written as a 12-tuple and has the following 12 components:

$$r_{\text{port}} = \begin{bmatrix} 0.41\%, 1.23\%, 0.06\%, 1.48\%, -1.20\%, -1.18\% \\ 0.23\%, -3.21\%, -3.89\%, 2.67\%, -1.29\%, -0.43\% \end{bmatrix}$$

[2]The monthly rate of return on the S&P 500 is computed as follows:

$$\frac{\begin{matrix}\text{Dividends paid on all the} \\ \text{stock in the index}\end{matrix} + \begin{matrix}\text{Change in the index value} \\ \text{for the month}\end{matrix}}{\text{Value of the index at the beginning of the period}} - 1$$

TABLE 1.1 Monthly Returns for the Hedge Fund Composite and S&P 500 Indexes

Month	Hedge Fund Portfolio	S&P 500
January	0.41%	2.26%
February	1.23%	3.20%
March	0.06%	−0.10%
April	1.48%	2.85%
May	−1.20%	−1.35%
June	−1.18%	−1.83%
July	0.23%	−2.15%
August	−3.21%	−5.68%
September	−3.89%	−7.18%
October	2.67%	10.77%
November	−1.29%	−0.51%
December	−0.43%	0.85%

Similarly, the return $r_{S\&P}$ on the S&P 500 can be expressed as a 12-tuple as follows:

$$r_{S\&P} = \begin{bmatrix} 2.26\%, 3.20\%, -0.10\%, 2.85\%, -1.35\%, -1.83\% \\ -2.15\%, -5.68\%, -7.18\%, 10.77\%, -0.51\%, 0.85\% \end{bmatrix}$$

One can perform standard operations on n-tuples. For example, consider the hedge fund portfolio returns in the two 12-tuples. The 12-tuple that expresses the deviation of the hedge fund portfolio's performance from the benchmark S&P 500 index is computed by subtracting from each component of the return 12-tuple from the corresponding return on the S&P 500. That is,

$$\begin{aligned} r_{port} - r_{S\&P} &= \begin{bmatrix} 0.41\%, 1.23\%, 0.06\%, 1.48\%, -1.20\%, -1.18\% \\ 0.23\%, -3.21\%, -3.89\%, 2.67\%, -1.29\%, -0.43\% \end{bmatrix} \\ &\quad - \begin{bmatrix} 2.26\%, 3.20\%, -0.10\%, 2.85\%, -1.35\%, -1.83\% \\ -2.15\%, -5.68\%, -7.18\%, 10.77\%, -0.51\%, 0.85\% \end{bmatrix} \\ &= \begin{bmatrix} -1.86\%, -1.96\%, 0.17\%, -1.37\%, 0.15\%, 0.65\% \\ 2.37\%, 2.46\%, 3.29\%, -8.10\%, -0.78\%, -1.29\% \end{bmatrix} \end{aligned}$$

It is the resulting 12-tuple that is used to compute the **tracking error** of a portfolio—the standard deviation of the variation of the portfolio's return from its benchmark index's return.

JWBT1003-c01 JWBT1003-Foccardi Printer: Courier Westford August 13, 2013 15:56 Trim: 6in × 9in

Coming back to the portfolio return, one can compute a logarithmic return for each month by adding 1 to each component of the 12-tuple and then taking the natural logarithm of each component. One can then obtain a geometric average, called the **geometric return**, by multiplying each component of the resulting vector and taking the 12th root.

Distance

Consider the real line R^1 (i.e., the set of real numbers). Real numbers include rational numbers and irrational numbers. A **rational number** is one that can be expressed as a fraction, c/d, where c and d are integers and $d \neq 0$. An **irrational number** is one that cannot be expressed as a fraction. Three examples of irrational numbers are

$$\sqrt{2} \cong 1.4142136$$
$$\text{Ratio between diameter and circumference}$$
$$= \pi \cong 3.14159265358979323846 26$$
$$\text{Natural logarithm} = e \cong 2.71828182845904523536028747135 26$$

On the real line, distance is simply the absolute value of the difference between two numbers $|a - b|$ which also can be written as

$$\sqrt{(a - b)^2}$$

R^n is equipped with a natural metric provided by the Euclidean distance between any two points

$$d[(a_1, a_2, \ldots, a_n), (b_1, b_2, \ldots, b_n)] = \sqrt{\sum (a_i - b_i)^2}$$

Given a set of numbers A, we can define the least upper bound of the set. This is the smallest number s such that no number contained in the set exceeds s. The quantity s is called the **supremum** and written as $s = \sup A$. More formally, the supremum is that number, if it exists, that satisfies the following properties:

$$\forall a : a \in A, s \geq a$$
$$\forall \varepsilon > 0, \exists a : s - a \leq \varepsilon$$

where ε is any real positive number. The supremum need not belong to the set A. If it does, it is called the **maximum.**

JWBT1003-c01 JWBT1003-Foccardi Printer: Courier Westford August 13, 2013 15:56 Trim: 6in × 9in

Similarly, **infimum** is the greatest lower bound of a set *A*, defined as the greatest number *s* such that no number contained in the set is less than *s*. If infimum belongs to the set it is called the **minimum.**

Density of Points

A key concept of set theory with a fundamental bearing on calculus is that of **density of points**. In fact, in financial economics we distinguish between discrete and continuous quantities. **Discrete quantities** have the property that admissible values are separated by finite distances. **Continuous quantities** are such that one might go from one to any of two possible values passing through every possible intermediate value. For instance, the passing of time between two dates is considered to occupy every possible instant without any gap.

The fundamental continuum is the set of real numbers. A **continuum** can be defined as any set that can be placed in a one-to-one relationship with the set of real numbers. Any continuum is an **infinite non-countable set**; a proper subset of a continuum can be a continuum. It can be demonstrated that a finite interval is a continuum as it can be placed in a one-to-one relationship with the set of all real numbers.

The intuition of a continuum can be misleading. To appreciate this, consider that the set of all rational numbers (i.e., the set of all fractions with integer numerator and denominator) has a dense ordering, that is, has the property that given any two different rational numbers *a*,*b* with $a < b$, there are infinite other rational numbers in between. However, rational numbers have the cardinality of natural numbers. That is to say rational numbers can be put into a one-to-one relationship with natural numbers. This can be seen using a clever construction that we owe to the seventeenth-century Swiss mathematician Jacob Bernoulli.

Using Bernoulli's construction, we can represent rational numbers as fractions of natural numbers arranged in an infinite two-dimensional table in which columns grow with the denominators and rows grow with the numerators. A one-to-one relationship with the natural numbers can be established following the path: (1,1) (1,2) (2,1) (3,1) (2,2) (1,3) (1,4) (2,3) (3,2) (4,1) and so on (see Table 1.2).

TABLE 1.2 Bernoulli's Construction to Enumerate Rational Numbers

1/1	1/2	1/3	1/4
2/1	2/2	2/3	2/4
3/1	3/2	3/3	3/4
4/1	4/2	4/3	4/4

Bernoulli thus demonstrated that there are as many rational numbers as there are natural numbers. Though the set of rational numbers has a dense ordering, rational numbers do not form a continuum as they cannot be put in a one-to-one correspondence with real numbers.

Given a subset A of R^n, a point $a \in A$ is said to be an **accumulation point** if any sphere centered in a contains an infinite number of points that belong to A. A set is said to be "closed" if it contains all of its own accumulation points and "open" if it does not.

FUNCTIONS

The mathematical notion of a function translates the intuitive notion of a relationship between two quantities. For example, the price of a security is a function of time: to each instant of time corresponds a price of that security.

Formally, a **function** f is a mapping of the elements of a set A into the elements of a set B. The set A is called the **domain** of the function. The subset $R = f(A) \subseteq B$ of all elements of B that are the mapping of some element in A is called the **range** R of the function f. R might be a proper subset of B or coincide with B.

The concept of function is general: the sets A and B might be any two sets, not necessarily sets of numbers. When the range of a function is a set of real numbers, the function is said to be a **real function** or a **real-valued function**.

Two or more elements of A might be mapped into the same element of B. Should this situation never occur, that is, if distinct elements of A are mapped into distinct elements of B, the function is called an **injection**. If a function is an injection and $R = f(A) = B$, then f represents a one-to-one relationship between A and B. In this case the function f is invertible and we can define the **inverse function** $g = f^{-1}$ such that $f(g(a)) = a$.

Suppose that a function f assigns to each element x of set A some element y of set B. Suppose further that a function g assigns an element z of set C to each element y of set B. Combining functions f and g, an element z in set C corresponds to an element x in set A. This process results in a new function, function h, and that function takes an element in set A and assigns it to set C. The function h is called the composite of functions g and f, or simply a **composite function**, and is denoted by $h(x) = g[f(x)]$.

VARIABLES

In applications in finance, one usually deals with functions of numerical variables. Some distinctions are in order. A **variable** is a symbol that represents any element in a given set. For example, if we denote time with a variable *t*, the letter *t* represents any possible moment of time. **Numerical variables** are symbols that represent numbers. These numbers might, in turn, represent the elements of another set. They might be thought of as numerical indexes which are in a one-to-one relationship with the elements of a set. For example, if we represent time over a given interval with a variable *t*, the letter *t* represents any of the numbers in the given interval. Each of these numbers in turn represents an instant of time. These distinctions might look pedantic but they are important for the following two reasons.

First, we need to consider **numeraire** or units of measure. Suppose, for instance, that we represent the price *P* of a security as a function of time *t*: $P = f(t)$. The function *f* links two sets of numbers that represent the physical quantities price and time. If we change the time scale or the currency, the numerical function *f* will change accordingly though the abstract function that links time and price will remain unchanged.

Variables can be classified as qualitative or quantitative. Qualitative (or categorical) variables take on values that are names or labels. Examples of qualitative variables would include the color of a ball (e.g., red, green, blue) or a dummy variable (also known as an indicator variable) taking the values 0 or 1. Quantitative variables are numerical. They represent a measurable quantity. For example, when we speak of the population of a city, we are talking about the number of people in the city, which is a measurable attribute of the city. Therefore, population would be a quantitative variable.

Variables can also be classified as deterministic or random. In probability and statistics, a random variable, or stochastic variable, is a variable that can take on a set of possible different values, each with an associated probability. For example, when a coin is tossed 10 times, the random variable is the number of tails (or heads) that are noted. *X* can only take the values 0, 1, . . . , 10, so in this example *X* is a discrete random variable. Variables might represent phenomena that evolve over time. A deterministic variable evolves according to fixed rules, for example an investment that earns a fixed compound interest rate that grows as an exponential function of time. A random variable might evolve according to chance.

One important type of function is a sequence. A **sequence** is a mapping of the set of natural numbers into real numbers.

KEY POINTS

- A set is a collection of objects called elements.
- Empty sets are sets that do not contain any elements.
- The union of two sets is formed by all elements that belong to either of the two sets.

JWBT1003-c01 JWBT1003-Foccardi Printer: Courier Westford August 13, 2013 15:56 Trim: 6in × 9in

- The intersection of two sets is formed by all elements that belong to both of the sets.
- Calculus describes the dynamics of quantitative phenomena.
- Real numbers represent the result of observations (or measurements) in a simple and natural way.
- Discrete quantities have the property that admissible values are separated by finite distances.
- Continuous quantities are such that one might go from one to any of two possible values passing through every possible intermediate value.
- A function is a relation between a set of inputs and a set of permissible outputs with the property that each input is related to exactly one output.
- A variable is a value that may change within the scope of a given problem or set of operations.
- Numerical variables are symbols that represent numbers.
- A deterministic variable is a variable whose value is not subject to variations due to chance.
- A random variable or stochastic variable is a variable whose value is subject to variations due to chance or randomness.