CHAPTER 3

Statistical Analysis of Financial Data

Chapters 1 and 2 presented various financial instruments in the form of market data familiar to Wall Street traders—namely, Bloomberg screens. Chapter 3 lays the mathematical foundation for the valuation of financial instruments, which depends in the first place on an analysis of the likelihood of future events using the tools of statistics and probability. This chapter shows you how to perform a statistical analysis of a given financial instrument by first identifying a suitable probability distribution and then calibrating it appropriately. Finally, this chapter discusses Risk measures such as *value at risk*, *conditional value at risk,* the *term structure of statistics, temporal autocorrelations,* and *volatility convexity*.

Tools in Probability Theory

Probability theory deals with mathematical models of processes whose outcomes are uncertain (random). One must define what these outcomes are and assign appropriate probabilities to every outcome. Certain formal mathematical definitions using set-theoretic notation follow, which equip the reader to study the extensive literature on probability theory.[1] Figure 3-1 tabulates the translation between set theory and probability theory.

[1]See, for example, Patrick Billingsley, *Probability and Measure*, Anniversary Ed., John Wiley and Sons, 2012; S. R. S. Varadhan, *Probability Theory*, American Mathematical Society, 2001.

Notation	Set Theory	Probability Concept
Ω	Collection of elements	Sample Space of all possible outcomes
ω	$\omega \in \Omega$	Elementary outcome or event
A	$A \subset \Omega$	Event that the outcomes in the set A occurs
A^c	Complement of A	Event that *no* outcome in the set A occurs
$A \bigcup B$	Union of A and B	Events in either A or B occur
$A \bigcap B$	Intersection of A and B	Events in both A and B occur
$\emptyset$	Empty set	Impossible Event

Figure 3-1. *Set-theoretic notation and probability*

- *Sample space* Ω: The set of all possible outcomes or elementary events w of some process is called the *sample space*. This sample space forms a mathematical *set*. For example,
    - The process of coin tossing has a sample space: $\Omega = \{H,T\}$.
    - The process of dice throwing has a sample space: $\Omega=\{1,2,3,4,5,6\}$.
- *Events* $A \subset \Omega$: Not every interesting observable of a random process is an elementary outcome. One may be interested in a combination of outcomes. Therefore, one defines an event A as a subset of the sample space. When throwing dice, possible events could be
    - an even number $A = \{2,4,6\}$
    - not an even number $A = \{2,4,6\}^c$
    - an even number less than 4, $A = \{2,4,6\} \cap \{1,2,3\}$

    Obviously, the basic sample elements of the sample space can be events, such that $A = \omega = \{2\}$.
- *σ-Algebra*: Simply choosing an interesting set of events is not sufficient for defining probabilities. The set of events needs to form a closed system under the set-theoretic operations of union and intersection. A *σ-algebra* is the set of all possible *countable* unions and intersections of a chosen set of events. This collection of events $\Im$ satisfies the following four conditions:

    a. $\Omega \in \Im$

    b. If $A_1, A_2, A_3, \ldots \in \Im$ then $\bigcup_{i=1}^{A_i} \in \Im$

c. If $A_1, A_2, A_3, \ldots \in \Im$ then $\bigcap_{i=1}^{A_i} \in \Im$

d. If $A \in \Im$ then $A^c \in \Im$

■ **Note** Condition (c) can be derived from (d) and (b).[2]

For example, if A is any subset of Ω, then $\Im = \{\varnothing, A, A^c, \Omega\}$ is a σ-algebra.

- *Probability Measure*: Now that a mathematically consistent set of events has been given (the σ-algebra $\Im$), one can assign probabilities to these events. There are a few basic restrictions on probabilities that are intuitively obvious but must be stated precisely in a mathematical way. First, probabilities must lie between 0 and 1. There are no negative probabilities, and there is no such thing as being more than 100% certain of some event. Also, the impossible event $\varnothing$ has a probability of 0. The probability of any event happening (that is, Ω) is clearly 100%. Finally, the probability of two *disjoint* or mutually exclusive events occurring must equal the sum of the probabilities of the individual events. A probability measure P on $(\Omega, \Im)$ is a function $P{:}\,\Im \to [0,1]$ such that

  a. $P(\varnothing) = 0$, $P(\Omega) = 1$

  b. If $A_1, A_2, A_3, \ldots \in \Im$ and $A_i \cap A_j = \varnothing$, then

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i) \quad . \tag{3.1}$$

  The triplet $(\Omega,\Im,P)$ is called a *probability space*.
  Assigning probabilities to the σ-algebra $\Im$ is often straightforward. For example, for coin tossing one has $\Omega = \{H,T\}$, $\Im = \{\varnothing,H,T,\Omega\}$ and

$$P(\varnothing) = 0, P(H) = p, P(T) = 1 - p, P(\Omega) = 1 \quad . \tag{3.2}$$

  If the coin is unbiased, $p = 0.5$.

- *Random Variables*: To create a mathematical model of probability, one must deal with numerical constructs and not symbols such as H and T. A *random variable* is a variable whose possible values are *numerical* outcomes of a random phenomenon (such as tossing a coin).

---

[2]For discussion of countability conditions, see, for example, Patrick Billingsley, *Probability and Measure*, Anniversary Ed. John Wiley and Sons, 2012.

A random variable is a function $X: \Omega \to \Re$ with the property that

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \Im, \forall x \in \Re \quad . \tag{3.3}$$

Basically, a random variable assigns a real valued number to an event from the sample space Ω.
Example: Coin tossing

$$\Omega = \{H, T\} \quad X(\omega) = \begin{cases} 1 & \omega = H \\ 0 & \omega = T \end{cases} \quad . \tag{3.4}$$

Example: Marks on an exam

$$\Omega = \{A, B, C, F\} \quad X(\omega) = \begin{cases} 1 & \omega = A \\ 2 & \omega = B \\ 3 & \omega = C \\ 4 & \omega = F \end{cases} \quad . \tag{3.5}$$

- *Distribution Function*: Now that a random variable has been defined, it is clear that it is more likely to be found in certain subsets of $\Re$ depending on the probability space (Ω, $\Im$, P) and the mapping X. What one really wants is the distribution $F(x)$ of the likelihood of X taking on certain values. One could define $F(x)$ as being the probability that the random number X equals a particular number x. This does not work in general, and therefore the more appropriate definition is

$$F_X(x) = \text{probability that } X \text{ does not exceed } x \quad . \tag{3.6}$$

Therefore, the distribution function of a random variable X is the function $F: \Re \to [0,1]$ given by

$$F_X(x) = P(X \leq x) \quad , \tag{3.7}$$

such that

i. $P(X > x) = 1 - F(x)$

ii. $P(x < X \leq y) = F(y) - F(x)$

iii. If $x < y$ then $F(x) \leq F(y)$

iv. $\lim_{x \to -\infty} F(x) \to 0, \ \lim_{x \to \infty} F(x) \to 1 \quad .$

Two types of random variables will be discussed in the remainder of this chapter: *discrete* and *continuous*.

- *Discrete Random Variable*: A random variable X is called discrete if it takes values in some countable subset $\{x_1, x_2, \ldots, x_n\}$ of $\Re$. It has a *density function* $f: \Re \to [0,1]$ given by $f(x) = P(X = x)$. The distribution function for a discrete variable is

$$F(x) = \sum_{x_i \le x} f(x_i) \quad , \tag{3.8}$$

with the *normalization* condition (the probability of all discrete events)

$$\sum_{\forall i} f(x_i) = 1 \quad . \tag{3.9}$$

- *Continuous Random Variable*: A random X is called continuous if its distribution function can be expressed as

$$F(x) = \int_{-\infty}^{x} f(u)\,du \quad x \in \Re \quad , \tag{3.10}$$

with a density function $f: \Re \to [0, \infty)$ with the properties

  i. $\int_{-\infty}^{\infty} f(x)\,dx = 1$ normalization

  ii. $P(X = x) = 0, \ \forall x \in \Re$ (probability of getting precisely a specific real number is zero)

  iii. $P(a \le X \le b) = \int_a^b f(x)\,dx$

One can think of $f(x)dx$ as a differential element of probability, such that

$$P(x < X < x + dx) = F(x + dx) - F(x) \cong f(x)\,dx \tag{3.11}$$

Several popular discrete and continuous distribution functions are described as follows:

- *Binomial Distribution*: Let a random variable X take values 1 and 0 with probabilities p and $1-p$ (known as a Bernoulli trial). The discrete density function is $f(0) = 1-p, f(1) = p$. Perform n independent trials $\{X_1 \ldots X_n\}$ and count the total numbers of 1's, such that $Y = X_1 + X_2 + \ldots X_n$. The density function of Y is

$$f(k) = P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n \tag{3.12}$$

with the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \quad . \tag{3.13}$$

One can check that this probability mass function has the proper normalization

$$\sum_{k=0}^{n} f(k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (1-p)^n \left(\frac{p}{1-p}\right)^k \quad . \tag{3.14}$$

Using the identity

$$\sum_{k=0}^{n} \binom{n}{k} x^k = (1+x)^n \quad , \tag{3.15}$$

the desired result is obtained

$$(1-p)^n \sum_{k=0}^{n} \binom{n}{k} \left(\frac{p}{1-p}\right)^k = (1-p)^n \left[1 + \frac{p}{1-p}\right]^n = (1-p)^n \left[\frac{1}{1-p}\right]^n = 1 \quad . \tag{3.16}$$

In finance, the binomial distribution is used in the *binomial tree model* (Chapter 4).

- *Poisson Distribution*: This is another counting-type discrete probability distribution that considers the random count of a given number of *independent* events occurring in a fixed interval (space or time) with a known average occurrence rate. Poisson distributions are found in financial processes with sudden large moves in price (*jumps*) or in credit default situations (Chapter 6). A discrete random variable Y with a Poisson distribution with parameter λ has a probability mass function given by

$$f(k) = P(Y = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0,1,\ldots,n, \quad \lambda > 0 \quad . \tag{3.17}$$

- *Normal (Gaussian) Distribution*: As the ubiquitous distribution of probability theory, the normal distribution has been overused on account of its simplicity. Its use in finance is also fraught with peril. A large portion of this chapter is dedicated to describing why *not* to use the following distribution and to providing alternative distributions with better properties for describing financial data:

$$f(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) dx, \quad x \in \Re, \sigma > 0 \quad . \tag{3.18}$$

This distribution is often denoted as $N(\mu, \sigma^2)$. Note that by substituting $y=\frac{x-\mu}{\sigma}$, $dy=\frac{dx}{\sigma}$, the foregoing reduces to $N(0,1)$ for the variable y (because $-\infty<x<\infty$). Therefore, if one has an $N(0,1)$ variable y, one can create an $N(\mu, \sigma^2)$ variable by the transformation $x = \mu + \sigma y$.

- *Log-normal Distribution*: This is another popular but overused distribution seen in finance, suffering from many of the same problems as the normal distribution.

$$f(x)dx = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{[\ln(x)-\mu]^2}{2\sigma^2}\right)\frac{dx}{x}, \quad x \in \Re, x>0 \quad . \tag{3.19}$$

The reason this is called the *log-normal distribution* is that the natural logarithm of the variable x is normally distributed. This is easy to see by substituting $y = \ln(x)$, $dy = \frac{dx}{x}$ into (3.19).

- *Gamma Distribution*: This has become a popular distribution in finance in conjunction with the *variance-gamma* model [Glasserman, 2003]. It is a two-parameter continuous distribution given by

$$f(x) = \frac{\lambda}{\Gamma(\alpha)}(\lambda x)^{\alpha-1} e^{-\lambda x}, \quad x \in \Re, x>0 \quad \lambda>0, \quad \alpha>0 \quad , \tag{3.20}$$

where the gamma function is

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t}\, dt \quad . \tag{3.21}$$

- *Beta Distribution*: This continuous distribution differs from the ones above in that the range of the underlying random variable is between zero and one. A financial variable with this property is the recovery rates of defaulted bonds (0% to 100%). This will be explained later in this chapter. The density function has two parameters and is given by

$$f(x) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, \quad 0<x<1, \quad p,q>0 \quad , \tag{3.22}$$

or, when written in terms of the beta function,

$$f(x) = \frac{1}{B(p,q)} x^{p-1}(1-x)^{q-1}, \quad 0<x<1, \quad p,q>0 \quad , \tag{3.23}$$

where the beta function is

$$B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\,dt \quad . \tag{3.24}$$

One can check the normalization condition as follows:

$$\int_0^1 f(x)\,dx = \frac{1}{B(p,q)}\int_0^1 x^{p-1}(1-x)^{q-1}\,dx = \frac{B(p,q)}{B(p,q)} = 1 \quad . \tag{3.25}$$

- *Generalized Student's*-t *Distribution*: This is a three-parameter (μ, ν, λ) continuous distribution that will be used later in this chapter. Historically, the random variable is denoted as t rather than the more traditional x. The probability density function is

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\sqrt{\frac{\lambda}{\nu\pi}}\left(1+\frac{\lambda(t-\mu)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad , \tag{3.26}$$

where the parameter ν is often referred to as the *degrees of freedom*. (Note that this is still a one-dimensional probability distribution.)

Moments of a Distribution

What does one want to measure in a probabilistic system? One can either measure a specific value of a random variable (for example, the number of heads that come up with multiple coin tosses), or one can construct some sort of average of the random variable itself. If X_n is the value of a random variable from the n_{th} trial of that variable, the (obvious) average of this variable after N trials is

$$\bar{X}_N = \frac{1}{N}\sum_{n=1}^{N} X_n \tag{3.27}$$

In the limit of a very large number of trials, by the law of large numbers $\bar{X}_N$ approaches the *mean* or *expected value* $E[X]$ of the random variable X,

$$\lim_{N\to\infty} \bar{X}_N = \lim_{N\to\infty}\frac{1}{N}\sum_{n=1}^{N} X_n = E[X] \quad . \tag{3.28}$$

The mean or expected value of a discrete random variable X is given by

$$E[X] = \sum_{x} x\, f(x) \quad . \tag{3.29}$$

For continuous random variables, one has

$$E[X] = \int_{x \in \Re} x\, f(x)\, dx \quad . \tag{3.30}$$

The *expectation operator* has the following *linearity property*. If X and Y are random variables and a,b and c are constants, then

$$E[aX + bY + c] = a\,E[X] + b\,E[Y] + c \quad . \tag{3.31}$$

The expectation is often called the *first moment* of a distribution. If n is a positive integer, the n^{th} moment m_n of a random variable X is

$$m_n = E[X^n] \quad . \tag{3.32}$$

Moments can describe the *shape* of a density function. Consider the symmetric Gaussian density function. The first moment $\mu = E[X]$ of this density function is where the shape of the density function is centered. The second moment $\sigma^2 = E[(X - \mu)^2]$ is a measure of the *width* of the density function, describing the probability of how far the random numbers can deviate from the mean μ. For $n > 2$, it is common to calculate the *normalized central moments*

$$\frac{E[(X-\mu)^n]}{\sigma^n} \quad , \tag{3.33}$$

where $\mu = E[X]$ and $\sigma^2 = E[(X-\mu)^2]$.

σ^2 is called the *variance*, and σ is the *standard deviation*. These moments are dimensionless quantities and are invariant to any linear change of scale of the random variable X. The next two moments are the most critical to risk management: the third moment, *skewness*, and the fourth moment, *kurtosis*. Skewness describes the asymmetry of the shape of the density function; kurtosis describes the "thickness" or "fatness" of the tails of the density function. Financial return data are often described by their empirical moments. The standard deviation of financial return data is called the *volatility* of returns. Moments are the primary method for determining the suitability of using a distribution to represent empirical facts. Matching the moments of empirical data to the moments of a theoretical distribution is one of the key calibration methods discussed below.

What are the moments of the commonly used Gaussian distribution

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \sim N(0, \sigma^2) \quad ? \tag{3.34}$$

To calculate the moments, the following integral relation proves useful

$$\int_{-\infty}^{\infty} \exp(-ax^2)\,dx = \sqrt{\frac{\pi}{a}} \quad . \tag{3.35}$$

Its proof is as follows. First, take a square of the left-hand side and make a transformation to polar coordinates:

$$I = \int_{-\infty}^{\infty} \exp(-ax^2)\,dx = \sqrt{\frac{\pi}{a}}$$

$$I^2 = \int_{-\infty}^{\infty} \exp(-ax^2)\,dx \int_{-\infty}^{\infty} \exp(-ay^2)\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-a(x^2+y^2)\right] dx\,dy \tag{3.36}$$

$$x = r\cos(\theta)$$
$$y = r\sin(\theta)$$

Recall that making a transformation requires calculating the Jacobian of the transformation:

$$dx\,dy = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} dr\,d\theta$$

$$\begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos(\theta) & \sin(\theta) \\ -r\sin(\theta) & r\cos(\theta) \end{vmatrix} = r\cos^2(\theta) + r\sin^2(\theta) = r \tag{3.37}$$

Calculating the integral in polar coordinates is straightforward:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[-a(x^2+y^2)\right] dx\,dy = \int_0^{2\pi}\int_0^{\infty} \exp\left[-ar^2\right] r\,dr\,d\theta$$

$$I^2 = \int_0^{2\pi}\int_0^{\infty} \exp\left[-ar^2\right] r\,dr\,d\theta = -\frac{1}{2a} 2\pi \exp\left[-ar^2\right]_0^{\infty} = \frac{\pi}{a} \tag{3.38}$$

$$I = \int_{-\infty}^{\infty} \exp(-ax^2)\,dx = \sqrt{\frac{\pi}{a}}$$

From this relation, one has, using $a = 1/(2\sigma^2)$, the normalization constraint of the density function

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \frac{\sqrt{2\pi\sigma^2}}{\sqrt{2\pi\sigma^2}} = 1 \quad . \tag{3.39}$$

How can one use the integral relation to calculate moments as powers of x are needed in the integrand? One can use the following trick to calculate all the moments.

Take a derivative of the integral relation (both left- and right-hand sides) with respect to *a* (not *x*)

$$\int_{-\infty}^{\infty} x^2 \exp(-ax^2)\,dx = \frac{1}{2}\sqrt{\frac{\pi}{a^3}} \quad . \tag{3.40}$$

A power of x^2 has been obtained. This relation will be useful when calculating the second moment. One can repeat this process on the above result to get all the even powers of X

$$\int_{-\infty}^{\infty} x^4 \exp(-ax^2)\,dx = \frac{3\sqrt{\pi}}{4} a^{-\frac{5}{2}} \quad . \tag{3.41}$$

Odd powers of the integral are zero because the integrand is an even function and the range of the integration is from -∞ to ∞. The positive and negative parts of the integral cancel each other out.

The variance of the Gaussian distribution using (3.37) is

$$E[X^2] = \int_{-\infty}^{\infty} \frac{x^2}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx = \frac{1}{\sqrt{2\pi\,\sigma^2}} \frac{1}{2}\sqrt{\pi(2\sigma^2)^3} = \sigma^2 \quad . \tag{3.42}$$

The third moment, *skewness*, of the Gaussian distribution is zero because all odd powers of the required integral are zero.

The fourth moment, kurtosis, of the Gaussian distribution is

$$\frac{E[(X-\mu)^4]}{\{E[(X-\mu)^2]\}^2} = \frac{E[X^4]}{\sigma^4} = \frac{1}{\sigma^4} \frac{1}{\sqrt{2\pi\,\sigma^2}} \frac{3\sqrt{\pi}}{4} (2\sigma^2)^{\frac{5}{2}} \rightarrow \frac{E[X^4]}{\sigma^4} = 3 \quad . \tag{3.43}$$

■ **Note** All normal distributions have a kurtosis of 3.

Kurtosis is a measure of how peaked a distribution is and how heavy (*fat*) its tails are. A high kurtosis distribution has a sharper peak and longer, fatter tails, whereas a low kurtosis distribution has a more rounded peak and shorter, thinner tails (Figure 3-2). A higher kurtosis distribution means more of the variance is the result of rare extreme deviations as opposed to frequent modestly sized deviations. Kurtosis risk is commonly referred to as *fat-tail risk*.

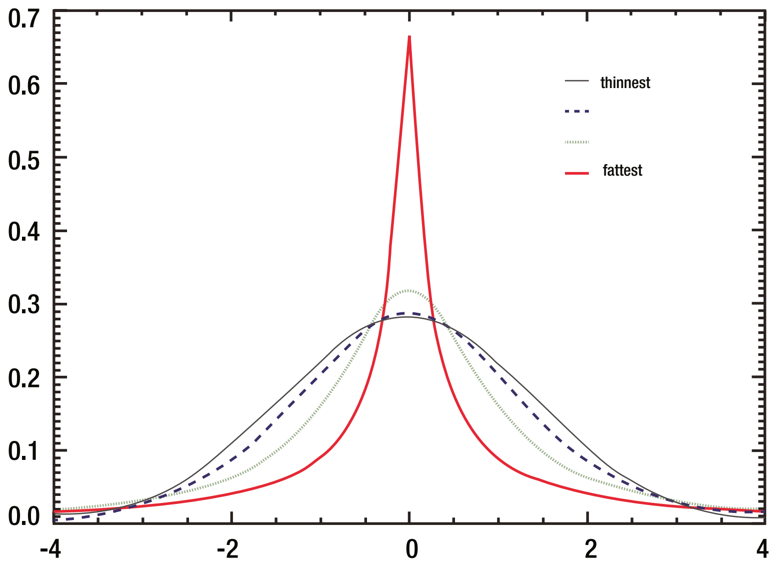

Figure 3-2. *Fat-tailed distributions*

■ **Caution** Not recognizing the importance of kurtosis risk has precipitated crises at many Wall Street firms, such as Long-Term Capital Management, Lehman Brothers, Bear Stearns, and Merrill Lynch.

How peaked a distribution and how fat the tails of a distribution are with respect to a normal distribution, which has a kurtosis of 3, is measured by the *excess kurtosis*

$$\frac{E[(X-\mu)^4]}{\{E[(X-\mu)^2]\}^2}-3 \quad . \tag{3.44}$$

(In many statistical systems, such as the one in Microsoft Excel, the kurtosis calculation is in fact the excess kurtosis.)

The normal distribution has serious shortcomings as a model for changes in market prices of an asset. In virtually all markets, the distribution of observed market price changes displays far higher peaks and heavier tails than can be captured with a normal distribution. Nassim Taleb, a former Wall Street derivatives trader, developed the *black swan theory of rare events* to explain the disproportionate role of high-impact, hard-to-predict, and rare events that are beyond the realm of normal expectations in history, science, finance, and technology—such as the rise of the personal computer, the Internet, World War I, and the

September 11 attacks. Taleb summarizes his criteria for identifying a black swan event as follows:[3]

> *What we call here a Black Swan is an event with the following three attributes. First, it is an outlier, as it lies outside the realm of regular expectations, because nothing in the past can convincingly point to its possibility. Second, it carries an extreme impact. Third, in spite of its outlier status, human nature makes us concoct explanations for its occurrence after the fact, making it explainable and predictable.*

Taleb contends that banks and trading firms are very vulnerable to hazardous black swan events and are exposed to losses beyond those predicted by their defective models (often based on normal and log-normal distributions). In connection with the third attribute of black swans, note that Taleb advanced his theory *before* the mortgage crisis of 2008.

Creating Random Variables and Distributions

Before dealing with real financial data, it is useful first to generate random variables on a computer and create distributions from these generated numbers. The generated random variables can be seen as *synthetic data* and the methods of creating distributions from them will lead naturally to the moment-matching calibration methods of the following section. Because real data is rarely clean and precise, some practice with idealized data is helpful in learning the concepts. The generation of *pseudo* and *quasi* random numbers and their use in *Monte Carlo* simulations is of fundamental importance in financial engineering (Chapter 4). This section will describe a very straightforward method for generating random numbers using Microsoft Excel.[4]

The Inverse Transform Method

The *inverse transform method* is a basic method for sampling random numbers from a specific distribution given an *invertible* expression for the distribution and a method for generating continuous uniform random numbers between 0 and 1. The idea is to sample a uniform number U between 0 and 1, interpret it as a probability, and assign a random number x from a pre-specified distribution by choosing the smallest number x such that $P(X \leq x) \geq U$. As probabilities lie between 0 and 1, U needs to lie between 0 and 1.

[3]Nassim Taleb, *The Black Swan: The Impact of the Highly Improbable*. Random House, 2007.
[4]For other methods of generating random numbers, see Paul Glasserman, *Monte Carlo Methods in Financial Engineering*. Springer, 2003.

Suppose one wants to sample random numbers from a cumulative distribution function $F_X(x) = P(X \le x)$. Because $F_X(x)$ is a non-decreasing function, the inverse function, if it exists, may be defined for any value U between 0 and 1 as $F_X^{-1}(U) = \inf\{x : F_X(x) \ge U, 0 < U < 1\}$.

Note that $P(F_X^{-1}(U) \le x) = P(U \le F_X(x))$ by the fact that $F_X(x)$ is monotonic. Finally, $P(U \le F_X(x)) = F_X(x)$ as U is assumed to be uniformly distributed, indicating that $F_X^{-1}(U)$ follows the desired distribution.

One can generate a uniform random number between 0 and 1 and interpret it as a probability. This probability can then be seen as the result of a specific random number that can be found by using the *inverse* distribution on the generated probability. The inverse transform sampling method has the following steps:

1. Generate a probability from a standard uniform distribution

$$\rightarrow U \sim Unif\,[0,1]$$

2. Associate a random number with this probability using the inverse transformation (3.45)

$$\rightarrow x = F_X^{-1}(U)$$

This can be done for any random variable where the inverse transformation is known and can be calculated explicitly. For Gaussian variables $N(\mu,\sigma^2)$, the distribution function

$$F_X(x) = \int_{-\infty}^{x} f(s)\,ds = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{x} \exp\left(-\frac{(s-\mu)^2}{2\sigma^2}\right) ds \tag{3.46}$$

does not have an explicit expression for the inverse distribution. Yet approximate formulas exist for the Gaussian distribution inverse function for any desired accuracy. Microsoft Excel has the following two functions, indicated by the arrow, that allow one to create $N(\mu,\sigma^2)$ random variables in a spreadsheet

$$\begin{gathered} U \sim \text{Unif}\,[0,1] \rightarrow \text{RAND}\,() \\ x = F^{-1}(U) \rightarrow \text{NORMINV}(U, \mu, \sigma) \end{gathered} \quad . \tag{3.47}$$

These Excel functions RAND() and NORMINV() are good for demonstration and testing purposes but are not accurate enough for real-world use. Figure 3-3 shows an example of the spreadsheet use of these functions (see Problem 3-1). The empirical moments of the random numbers generated must be compared to the theoretical moments. It is critical that one checks the moments of the random variables produced versus the theoretical ones given by the inputs μ and σ. Remember that the purpose here is to produce random numbers that have the characteristics of the Gaussian distribution. This convergence will depend upon how many random numbers are produced and how accurate the random number generator is. The random number generator in Excel, RAND(), is poor and is not recommended for genuine use. The number of random observations one needs to generate is dependent on the distribution in question, but in general for fat-tailed distributions more random numbers are needed to fully "experience" the distribution. An order-of-magnitude estimate is that one needs to produce at least one million numbers for true convergence (which is not possible in Excel).

INPUTS	
Mean	0.05
Sigma	0.25

Excel Function	OUTPUTS		Excel Function
average()	Mean	0.050195	
stdev()	Stdev	0.24993	
kurt()	Ex-Kurt	-0.00087	
	Uniform	Normal	
RAND()	0.101931	-0.26766	NORMINV(U, Mean, Sigma)
RAND()	0.6825198	0.168689	NORMINV(U, Mean, Sigma)
RAND()	0.5792655	0.100004	NORMINV(U, Mean, Sigma)
RAND()	0.7289174	0.202386	NORMINV(U, Mean, Sigma)
RAND()	0.4770257	0.035595	NORMINV(U, Mean, Sigma)
RAND()	0.2032693	-0.1575	NORMINV(U, Mean, Sigma)
RAND()	0.4303552	0.006133	NORMINV(U, Mean, Sigma)
RAND()	0.4106028	-0.0065	NORMINV(U, Mean, Sigma)
RAND()	0.7526418	0.220707	NORMINV(U, Mean, Sigma)
RAND()	0.8049862	0.264892	NORMINV(U, Mean, Sigma)
RAND()	0.0697711	-0.31937	NORMINV(U, Mean, Sigma)
RAND()	0.7581072	0.225057	NORMINV(U, Mean, Sigma)
RAND()	0.6383251	0.138496	NORMINV(U, Mean, Sigma)
RAND()	0.0770063	-0.30638	NORMINV(U, Mean, Sigma)
RAND()	0.7943755	0.255424	NORMINV(U, Mean, Sigma)

Figure 3-3. *Excel functions for the inverse transform method for normal variables*

■ **Note** RAND() will produce a new uniformly [0,1] distributed random number every time the Excel spreadsheet is recalculated.

Creating a Density Function: Histograms and Frequencies

Theoretical density functions are mimicked by histograms when dealing with finite amounts of empirical data. A *histogram* is a graphical representation of an approximate density function for an underlying random variable. The histogram groups nearby data into discrete buckets and counts the number of data points, called *frequencies*, in a predetermined bucket. Graphically, a histogram is a set of these frequencies pictured as adjacent rectangles over discrete intervals called *bins*, as shown in Figure 3-4. Each rectangle has a height equal to the frequency over that interval and a width equal to the bin size. The appropriate bin size and number of bins depend on the empirical data at hand. If the bin size is too gross, the histogram will not represent some of the subtle points of the true density function because too many data points will put into the same bin. If the bin size is too fine, one will not get a histogram because the frequencies in each bucket will be too small. Unfortunately, there is no systematically best method to choose the number of bins and bin size. It depends on the number of empirical points as well as how spread out these points are in the sample space. You could use the empirical standard deviation to estimate this spread of data, but you would miss the fat tail characteristics of the data unless you calculate the empirical kurtosis and any skew properties. You need to experiment with different ranges of bins and bin sizes and choose a histogram that best communicates the shape of the empirical distribution.

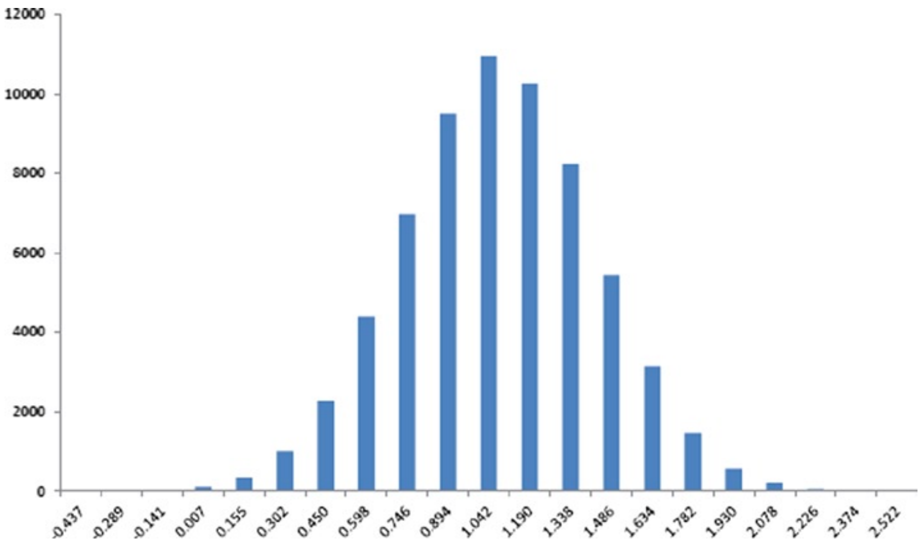

Figure 3-4. *Raw histogram*

Some rules of thumb are good to start with when creating a histogram. If $x_1, x_2, \ldots, x_n$ are the set of n data points, the basic relationship between bins and bin width is

$$\text{bin width} = \frac{\max(x) - \min(x)}{\text{number of bins}} \quad , \tag{3.48}$$

where one can start with the estimate

$$\text{number of bins} = \sqrt{n} \text{ or } n^{1/3} \quad . \tag{3.49}$$

(See Problem 3-1.)

There are two methods in Excel for creating a histogram, depending on the nature of the data—whether *static* or *dynamic*.

Excel Histogram-Creating Method: Static Data

For static data, one can use the Histogram tool in the Excel Data Analysis toolkit (Figure 3-5). The inputs are the predetermined bin range that has the number of bins with their specific attachment and detachment points and the empirical data that will be counted against these bins (Figure 3-6). Figure 3-7 shows the output bin range with the respective frequency count. These frequencies need to be *normalized*, as discussed in the section after next.

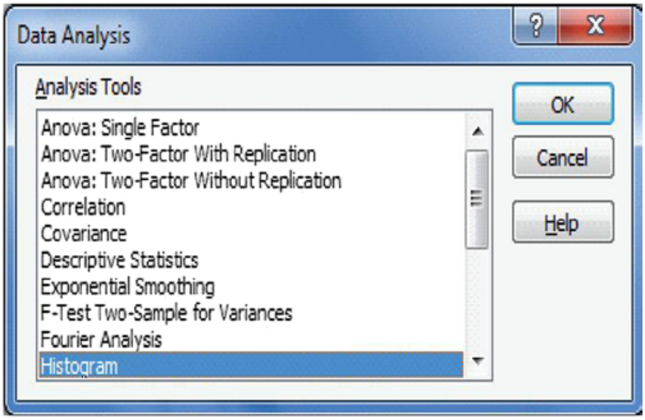

Figure 3-5. *Excel Histogram tool in Data Analysis*

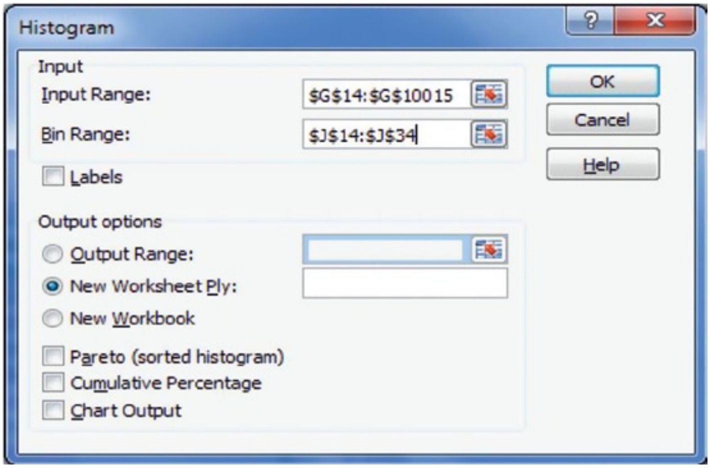

Figure 3-6. *Excel Histogram tool inputs*

Bin	Frequency
-1.000	0
-0.900	0
-0.800	0
-0.700	0
-0.600	0
-0.500	0
-0.400	1
-0.300	9
-0.200	56
-0.100	167
0.000	448
0.100	928
0.200	1496
0.300	1917
0.400	1909
0.500	1533
0.600	860
0.700	453
0.800	166
0.900	48
1.000	8
More	3

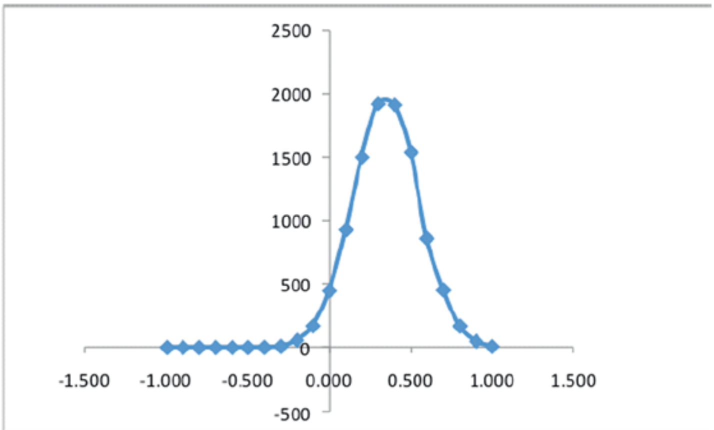

Figure 3-7. *Excel Histogram tool output*

Excel Histogram-Creating Method: Dynamic Data

If one's empirical data are continually changing, the static histogram method described above can become tiresome. The Excel array function *FREQUENCY(data array, bin array)* can be used and is readily able to recalculate when the inputs change, making it a dynamic function. For array functions, one must use CTRL-SHIFT-ENTER over the array space in Excel (usually beside the bin array). This method is ideal for creating a density function from randomly generated normal variables, such as the ones illustrated in Figure 3-3, as the function RAND() produces new numbers on every update of the spreadsheet.

Normalization of a Histogram

Raw frequencies do not equal a density function. A density function must have the property $\int_{-\infty}^{\infty} f(x)dx = 1$. Therefore, one must convert raw frequencies into normalized frequencies. Because one is dealing with discrete data, all integrals are turned into summations whereby

$$\int_{-\infty}^{\infty} \to \Sigma \quad , \tag{3.50}$$

such that

$$dx \to \Delta x \to (Bin\ Size) \tag{3.51}$$

The first step is to find the total area under the raw histogram. One can then normalize the raw frequencies to get normalized frequencies such that the resulting area under the normalized empirical density curve is one (Figure 3-8). The steps are as follow.

Bin Size 0.144777723		Area under Curve	Normalized
Bins	Excel Frequency Function	9409.03	1.0000
-1.048	1	0.14	0.00
-0.903	5	0.72	0.00
-0.758	25	3.62	0.00
-0.613	109	15.78	0.01
-0.469	303	43.87	0.03
-0.324	856	123.93	0.09
-0.179	1885	272.91	0.20
-0.034	3750	542.92	0.40
0.110	6323	915.43	0.67
0.255	8717	1262.03	0.93
0.400	10417	1508.15	1.11
0.545	10457	1513.94	1.11
0.690	8961	1297.35	0.95
0.834	6246	904.28	0.66
0.979	3796	549.58	0.40
1.124	1912	276.82	0.20
1.269	809	117.13	0.09
1.413	287	41.55	0.03
1.558	107	15.49	0.01
1.703	20	2.90	0.00
1.848	5	0.50	0.00

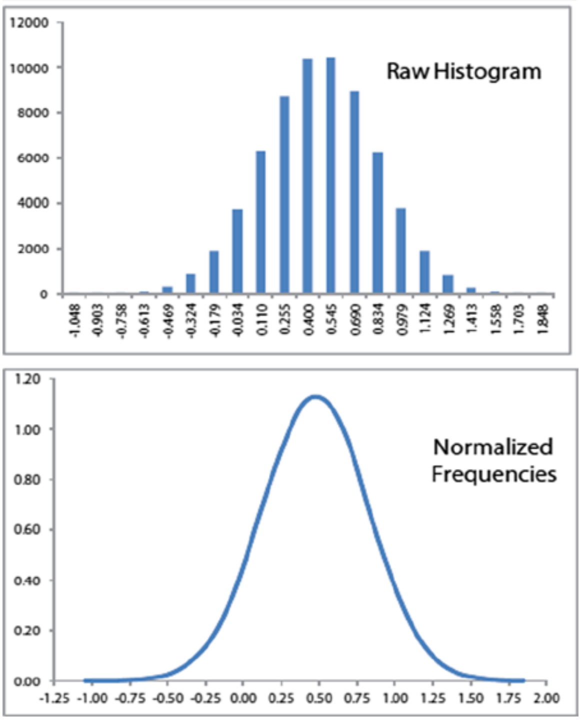

Figure 3-8. *Example of bins, frequencies, and normalization of empirical densities*

1) Area under the normalized empirical density curve:

$$\begin{aligned} \text{Area} &= \sum (\textit{Frequency}) \cdot (\textit{Bin Size}) \\ \text{Area} &= \int_{-\infty}^{\infty} f(x)\,dx \;\rightarrow\; \sum f(x)\,\Delta x \\ &= \sum (\textit{Frequency}) \cdot (\textit{Bin Size}) \end{aligned} \tag{3.52}$$

2) Normalize frequencies by the area:

$$\begin{aligned} \textit{Frequency} &\rightarrow \frac{\textit{Frequency}}{\textit{Area}} \\ &= \textit{Normalized Frequency} \end{aligned} \quad . \tag{3.53}$$

3) Perform a check of the area of the final distribution:

$$\begin{aligned} \text{Area} &= \sum (\textit{Normalized Frequency}) \cdot (\textit{Bin Size}) = 1 \\ &\rightarrow \int_{-\infty}^{\infty} f(x)\,dx = 1 \end{aligned} \quad . \tag{3.54}$$

With these normalized frequencies, one can calculate empirical moments—the mean, variance, and kurtosis—as follows, using the ordered bins of the discretized x-space (Figure 3-8) and assuming the random variable x is either in the middle or at one endpoint of the appropriate bin.

1. $\text{Mean} = E[x] = \int_{-\infty}^{\infty} x f(x)\,dx \rightarrow \sum x(\textit{Normalized Frequency})(\textit{Bin Size})$

2. $$\begin{aligned} \text{Var} &= E(x-\mu)^2 = E[x^2] \\ &= \int_{-\infty}^{\infty} x^2 f(x)\,dx \rightarrow \sum x^2 (\textit{Normalized Frequency})(\textit{Bin Size}), \\ \sigma &= \sqrt{\text{Var}} \end{aligned}$$

3. $\text{Kurt} = \dfrac{E[x^4]}{\sigma^4} = \dfrac{\int_{-\infty}^{\infty} x^4 f(x)\,dx}{\sigma^4} \rightarrow \dfrac{\sum x^4 (\textit{Normalized Frequency})(\textit{Bin Size})}{\sigma^4}$

Mixture of Gaussians: Creating a Distribution with High Kurtosis

Modeling heavy tails is a critical skill in risk management. High peaks and heavy tails are a characteristic of many financial markets with small to medium price changes during most times, interrupted by occasional large price moves. Even though the Gaussian distribution does not capture these features, a simple extension of the Gaussian distribution exhibiting a higher kurtosis than three is a mixture of two or more normal density functions. Consider the following dual mixture,

$$q\,N(0,\sigma_1^2)+(1-q)\,N(0,\sigma_2^2),\quad q\in(0,1) \tag{3.55}$$

$$f(x)\ \rightarrow\ \frac{q}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{x^2}{2\sigma_1^2}\right)+\frac{1-q}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{x^2}{2\sigma_2^2}\right)\quad . \tag{3.56}$$

This is *not* simply adding two random variates. What this produces is a distribution of a random variable drawn from $N(0, \sigma_1^2)$ with probability q and drawn from $N(0, \sigma_2^2)$ with probability $1-q$. A mixed Gaussian distribution is the weighted sum of Gaussian *densities*. It is a three-parameter (q, σ_1, σ_2) distribution. The moments can be calculated using the moments for simple normal distributions. First, check the normalization,

$$\int_{-\infty}^{\infty} f(x)\mathrm{dx}=\int_{-\infty}^{\infty}\left[\frac{q}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{x^2}{2\sigma_1^2}\right)+\frac{1-q}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{x^2}{2\sigma_2^2}\right)\right]dx=q+(1-q)=1 \tag{3.57}$$

Next, calculate the mean

$$E[x]=\int_{-\infty}^{\infty} x\,f(x)=\int_{-\infty}^{\infty}\left[\frac{q}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{x^2}{2\sigma_1^2}\right)+\frac{1-q}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{x^2}{2\sigma_2^2}\right)\right]x\,dx=0 \tag{3.58}$$

This result is not surprising because the individual means of the two densities are zero. Next comes the variance,

$$\begin{aligned}E[x^2]=\int_{-\infty}^{\infty} x^2\,f(x)&=\int_{-\infty}^{\infty}\left[\frac{q}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{x^2}{2\sigma_1^2}\right)+\frac{1-q}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{x^2}{2\sigma_2^2}\right)\right]x^2\,dx\\&=q\sigma_1^2+(1-q)\sigma_2^2=\sigma^2\end{aligned} \tag{3.59}$$

Like the mean, the skew is zero because the individual densities have zero skew. Finally, the kurtosis, using the above result is

$$\mathrm{kurt}=\frac{E[x^4]}{\sigma^4}=\frac{\int_{-\infty}^{\infty} x^4\,f(x)}{\sigma^4}=\frac{3(q\sigma_1^4+(1-q)\sigma_2^4)}{[q\sigma_1^2+(1-q)\sigma_2^2]^2}\quad . \tag{3.60}$$

Therefore, the variance and kurtosis of the three-parameter mixed Gaussian distribution are, respectively,

$$\sigma^2=q\sigma_1^2+(1-q)\sigma_2^2 \tag{3.61}$$

and

$$\text{kurt} = \frac{3(q\sigma_1^4 + (1-q)\sigma_2^4)}{\sigma^4} \quad . \tag{3.62}$$

Figures 3-9 and 3-10 compare a mixed Gaussian density function with that of $N(0,1)$.

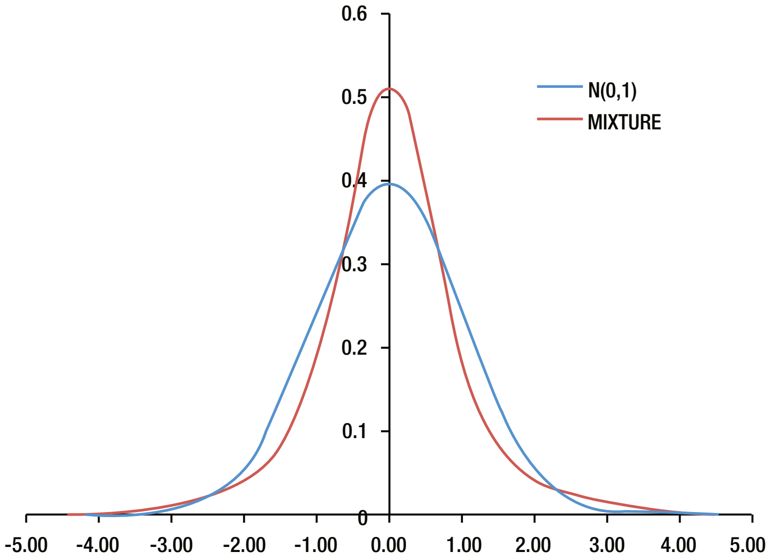

Figure 3-9. *Mixed Gaussian density function compared with that of N(0,1)*

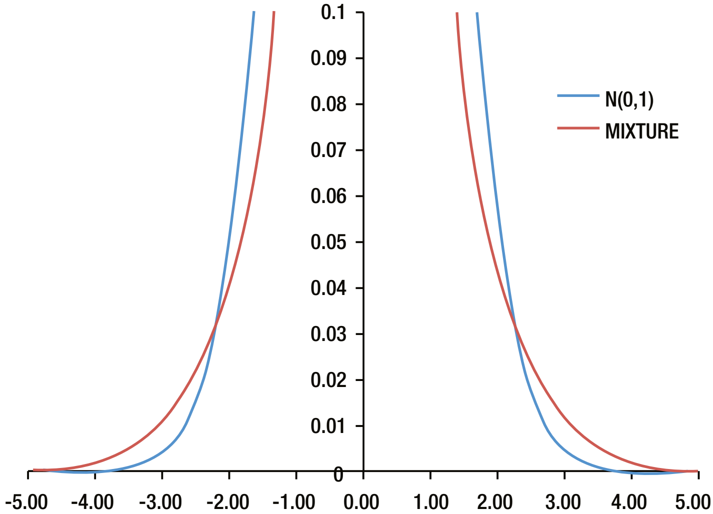

Figure 3-10. *Detail of mixed Gaussian density function compared with that of N(0,1)*

The next two sections describe two methods for creating a mixed Gaussian distribution function in Excel: a *random variable approach* using the inverse transform method and a *density approach* (see Problem 3-2).

Random Variable Approach

One can use the inverse transform method to create two series of normal variables with standard deviations of σ_1 and σ_2, each involving a series of calls to RAND(). Then one can use a third series of calls to RAND() to determine which of the two normal variables just created make up the mixture, such that

$$\text{if} \quad (RAND() < q) \quad choose \quad N(0,\sigma_1^2), \quad else\ choose \quad N(0,\sigma_2^2) \quad . \qquad (3.63)$$

Figure 3-11 shows a sample spreadsheet with this structure, in which the final column displays the outcome of the (3.63) relation. Here again, it is critical that you check the moments of the random variables produced versus the theoretical ones given by expressions (3.58), (3.59), and (3.60) based on the three inputs q, σ_1, and σ_2. Remember that the purpose here is to produce random numbers that have the characteristics of the mixed Gaussian distribution. The distribution is described by its moments (3.58), (3.59), and (3.60) (the mean and skew are zero), and the accuracy of this method is determined by the proximity of the moments resulting from the random numbers to the theoretical density function. The same strictures in respect of the inadequacy of RAND() and the deficiency in the number of random variables generated in Excel apply here as are noted in the section "The Inverse Transform Method."

	sigma1	30.0%	
	Mean	0.00	
	Stdev	30.0%	
	Ex Kurt	0.02	
RAND()	0.763206	0.214996	NORMINV()
RAND()	0.896701	0.378892	NORMINV()
RAND()	0.588111	0.066807	NORMINV()
RAND()	0.872683	0.34175	NORMINV()
RAND()	0.192086	-0.261071	NORMINV()
RAND()	0.877642	0.348983	NORMINV()
RAND()	0.581804	0.061953	NORMINV()
RAND()	0.371439	-0.098413	NORMINV()
RAND()	0.28791	-0.16785	NORMINV()
RAND()	0.942911	0.473907	NORMINV()
RAND()	0.667973	0.130296	NORMINV()
RAND()	0.752297	0.20452	NORMINV()
RAND()	0.155501	-0.303937	NORMINV()

	sigma2	10.0%	
	Mean	0.00	
	Stdev	10.0%	
	Ex Kurt	0.03	
RAND()	0.438846	-0.01539	NORMINV()
RAND()	0.492447	-0.00189	NORMINV()
RAND()	0.010766	-0.22985	NORMINV()
RAND()	0.661572	0.041676	NORMINV()
RAND()	0.202936	-0.08312	NORMINV()
RAND()	0.666807	0.043111	NORMINV()
RAND()	0.856947	0.10667	NORMINV()
RAND()	0.962552	0.17811	NORMINV()
RAND()	0.470122	-0.0075	NORMINV()
RAND()	0.860661	0.108329	NORMINV()
RAND()	0.644384	0.03702	NORMINV()
RAND()	0.589336	0.022584	NORMINV()
RAND()	0.623826	0.031555	NORMINV()

	q	30.0%
	sigma	18.4%
	EX KURT	3.488
	Mean	0.00
	Stdev	18.5%
	Ex Kurt	3.35
RAND()	0.832596	-0.01539
RAND()	0.231416	0.378892
RAND()	0.464458	-0.22985
RAND()	0.07792	0.34175
RAND()	0.089089	-0.26107
RAND()	0.506796	0.043111
RAND()	0.741928	0.10667
RAND()	0.478481	0.17811
RAND()	0.971649	-0.0075
RAND()	0.102874	0.473907
RAND()	0.453096	0.03702
RAND()	0.155939	0.20452
RAND()	0.473341	0.031555

Figure 3-11. *Random variable approach in Excel to generate mixed normal variables*

Density Approach

The density approach directly models the density function rather than the random numbers themselves. For a density function $f(x)$, one begins with discretizing the x-space. This will depend on the level and range of the moments of the density function in question. For a Gaussian density function, the range of x is theoretically $-\infty$ to ∞. Yet for most empirical values of the first two moments μ and σ, the realistic range will be substantially smaller because the tails of the density function exponentially decay. For example, for a normal distribution with a μ=0 and σ=100%, the discretized x-space will be from about –4.0 to 4.0. Density values outside this range will be essentially zero. Therefore, one must be careful when discretizing the x-space. Once the x-values are known, it is a simple matter to calculate the density directly using the expression for $f(x)$. Afterwards, one should check that the theoretical moments have actually been produced by the discrete mixed density function using the formulas (3.64)–(3.66) (also in Figure 3-12), analogous to those adduced in the "Normalization of a Histogram" section.

sigma1	160.0%
sigma2	35.0%
q	15.0%
Theor. sigma	70%
Theor. KURT	12.54

$binsize = 0.1$

$$\frac{1}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{x^2}{2\sigma_1^2}\right)$$

$$\frac{1}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{x^2}{2\sigma_2^2}\right)$$

$$\sum(Mixture)\cdot x\cdot(bin\ size)$$

$$\sqrt{\sum(Mixture)\cdot x\cdot x\cdot(bin\ size)}$$

$$\frac{\sum(Mixture)\cdot x^4\cdot(bin\ size)}{(Sigma)^4}$$

Mean	Sigma	Kurtosis
0.000%	**70%**	**12.42**

x	Dist1	Dist2	MIXTURE	x*x	x*x*x*x
-6.00	0.000220372	1.7464E-64	0.00003306	36.000	1296.00
-5.900	0.000278033	2.24684E-62	0.00004170	34.810	1211.74
-5.800	0.000349413	2.6641E-60	0.00005241	33.640	1131.65
-5.700	0.000437408	2.91123E-58	0.00006561	32.490	1055.60
-5.600	0.000545427	2.93189E-56	0.00008181	31.360	983.45
-5.500	0.00067747	2.72125E-54	0.00010162	30.250	915.06
-5.400	0.000838199	2.32775E-52	0.00012573	29.160	850.31
-5.300	0.001033018	1.83506E-50	0.00015495	28.090	789.05
-5.200	0.001268155	1.33325E-48	0.00019022	27.040	731.16
-5.100	0.001550744	8.92733E-47	0.00023261	26.010	676.52
-5.000	0.001888911	5.50907E-45	0.00028334	25.000	625.00
-4.900	0.002291851	3.13316E-43	0.00034378	24.010	576.48
-4.800	0.002769905	1.64223E-41	0.00041549	23.040	530.84
-4.700	0.003334624	7.93293E-40	0.00050019	22.090	487.97
-4.600	0.003998825	3.53167E-38	0.00059982	21.160	447.75
-4.500	0.004776629	1.44902E-36	0.00071649	20.250	410.06
-4.400	0.005683477	5.47917E-35	0.00085252	19.360	374.81
-4.300	0.006736126	1.90943E-33	0.00101042	18.490	341.88
-4.200	0.007952613	6.13252E-32	0.00119289	17.640	311.17

$$\frac{q}{\sqrt{2\pi\sigma_1^2}}\exp\left(-\frac{x^2}{2\sigma_1^2}\right)+\frac{1-q}{\sqrt{2\pi\sigma_2^2}}\exp\left(-\frac{x^2}{2\sigma_2^2}\right)$$

Figure 3-12. *Mixed Gaussian density approach in Excel*

$$\text{Mean} = E[x] = \int_{-\infty}^{\infty} x\, f_{mixture}(x)\, dx \to \sum x\,(Mixture)(Bin\ Size) \tag{3.64}$$

$$\text{Var} = E[(x-\mu)^2] = E[x^2] = \int_{-\infty}^{\infty} x^2\, f_{mixture}(x)\, dx \to \sum x^2\,(Mixture)(Bin\ Size)\,, \quad \sigma = \sqrt{\text{Var}} \tag{3.65}$$

$$\text{Kurt} = \frac{E[x^4]}{\sigma^4} = \frac{\int_{-\infty}^{\infty} x^4\, f_{mixture}(x)\, dx}{\sigma^4} \to \frac{\sum x^4\,(Mixture)(Bin\ Size)}{\sigma^4} \tag{3.66}$$

Skew Normal Distribution: Creating a Distribution with Skewness

Recall the definition of *skewness*,

$$\frac{E[(X-\mu)^3]}{\left(\sqrt{E[(X-\mu)^2]}\right)^3} \quad . \tag{3.67}$$

Skewness is another very important moment that is often ignored on Wall Street at great risk. Commonly used distributions and their third moments follow:

Gaussian distribution skew = 0

Mixed Gaussian distribution skew = 0

Student's-t distribution skew = 0

Figure 3-13 shows several density functions with positive and negative skew. The centered symmetric one (dashed line) has no skew. The two distributions to the left of center have negative skew even though they are leaning to the right. The negative skew here is manifested by the longer tail to the left. The reverse is true for the density functions with positive skew.

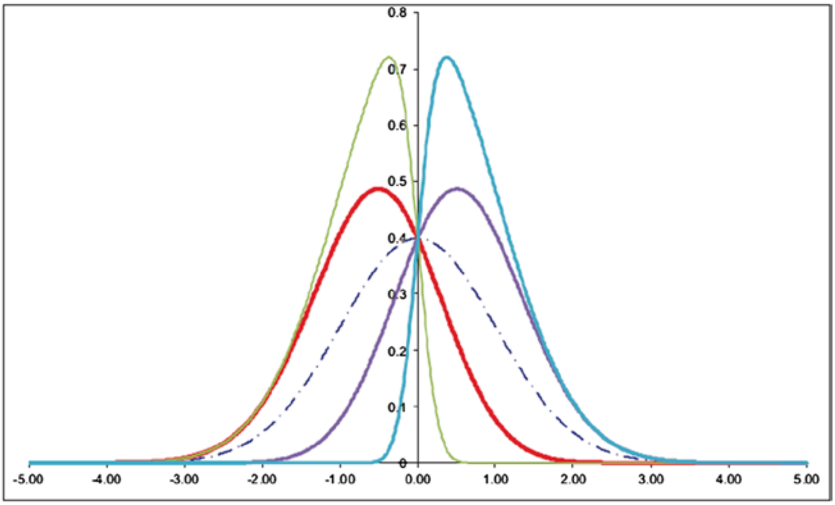

Figure 3-13. *Density functions with skew*

A skewed normal distribution is created by taking a $N(0,1)$ density function and changing the shape of it with a skewing function. A simple skewing function is given by

$$\Phi_{\text{skew}}(\alpha x) = \frac{1}{2}\left[1 + \text{erf}\left(\frac{\alpha x}{\sqrt{2}}\right)\right] = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\alpha x} \exp\left(-\frac{s^2}{2}\right) ds \quad . \tag{3.68}$$

where erf is the error function. This is easily recognizable as the distribution function of a Gaussian variable. Multiply this skewing function with that of a $N(0,1)$ density function

$$N(0,1) \rightarrow \phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad ,$$

$$f(x) = 2\phi(x)\Phi_{\text{skew}}(\alpha x) = \frac{1}{\pi}\exp\left(-\frac{x^2}{2}\right)\int_{-\infty}^{\alpha x} \exp\left(-\frac{s^2}{2}\right) ds \quad . \tag{3.69}$$

This is a one-parameter (α) skewed density function. It is illustrated in Figure 3-13 for several values of α. Positive values of α provide positive skew, and negative values of α provide negative skew. The moments of this distribution can be written in terms of its one parameter as follows:

$$\begin{aligned}
&\textit{Define}\quad \delta = \frac{\alpha}{\sqrt{1+\alpha^2}}\sqrt{\frac{2}{\pi}} \\
&E[X] = \delta \\
&Var[X] = \left(1-\delta^2\right) \\
&Skew[X] = \frac{4-\pi}{2}\frac{\delta^3}{\left(1-\delta^2\right)^{3/2}} \\
&Kurt[X] = 2(\pi-3)\frac{\delta^4}{\left(1-\delta^2\right)^2}+3
\end{aligned} \tag{3.70}$$

Calibrating Distributions through Moment Matching

Finding an appropriate probability distribution that can describe the future events of a certain type of financial instrument comes down to properly characterizing the statistical properties of some aspect of that financial instrument—for instance, the return characteristics of a stock. The historical financial time-series data of stock returns are best described by their empirical moments such as mean, variance, skew, kurtosis as well as empirical temporal moments such as the autocorrelation of returns, squared returns, and so forth (discussed in the "Autocorrelation" section). Any distribution that purports to describe data with a certain set of empirical moments must possess similar theoretical moments to characterize the data properly. If any key moments are missing from the distribution, any analysis based on the distribution must be faulty. The key objection of the black swan critique to the commonly used distributions is that their poor representation of the fat tails of financial data resulted in a colossal risk management failure at many Wall Street firms. To calibrate a distribution properly to represent empirical data, one must match the crucial empirical moments to the theoretical moments of the distribution. If the theoretical moments are lacking, the resulting statistical analysis of these data using the theoretical distribution may be faulty.

Calibrating a Mixed Gaussian Distribution to Equity Returns

A key aspect in analyzing financial data is to fit the distribution of returns for some asset to a known fat-tailed distribution such as the mixed Gaussian distribution. Here we use the density approach. The mixed Gaussian has three parameters, σ_1, σ_2, and q. One can use these to match the second and fourth moments of the empirical distribution (variance and kurtosis). The empirical distribution can be scaled to have zero mean as the moments are all mean adjusted.

Step 1: Create an Empirical Distribution

i. Obtain the financial time-series data of the stock in question. The amount of data will depend on the nature of the analysis. Generally, if you want to represent all historical features of the data, using as much data as possible is suggested.

ii. If P_t is the asset price at time t, create either a natural-log return time series $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right)$ or simple return time series $r_t = \left(\frac{P_t}{P_{t-1}} - 1\right)$. For small stock moves, these returns will be almost similar.

iii. Estimate the mean of these returns $\bar{r} = E[r_t]$.

iv. Calculate mean adjusted returns $\tilde{r}_t = r_t - \bar{r}$. These zero-mean returns will be fit to a mixed normal distribution.

v. Estimate the empirical standard deviation $\sigma_{empirical}$ and $\text{kurt}_{empirical}$ of $\tilde{r}_t$. Note that other empirical moments (such as skew) exist but cannot be fit to the mixed Gaussian distribution.

vi. To get some "visuals" on the data, create an appropriately normalized density function of the mean adjusted returns (as described previously). Calculate the variance and kurtosis from the empirical density function to make sure that it matches the variance and kurtosis of the empirical data. The accuracy will depend on the number of data points, number of bins, and so on. (The purpose of this substep is really for visual inspection of the data versus the theoretical distribution. It is not a requirement for moment matching.)

$$
\begin{aligned}
\sigma^2_{\text{emp den}} &= \sum x^2\,(\textit{Empirical Density})(\textit{Bin Size}) \\
\text{Kurt}_{\text{emp den}} &= \frac{\sum x^4\,(\textit{Empirical Density})(\textit{Bin Size})}{\sigma^4_{\text{emp den}}}
\end{aligned}
\tag{3.71}
$$

Step 2: Match the Theoretical Moments to the Empirical Moments

Because there are two moments to match with three parameters for the Gaussian mixture, you must fix one of the parameters and solve for the other two. Solving for all three will give a family of solutions, some of which may not be the fit you are looking for (check this by inspecting the graph of the fit, such as Figure 3-14). Assuming that σ_2 is fixed, one must solve for q and σ_1 such that

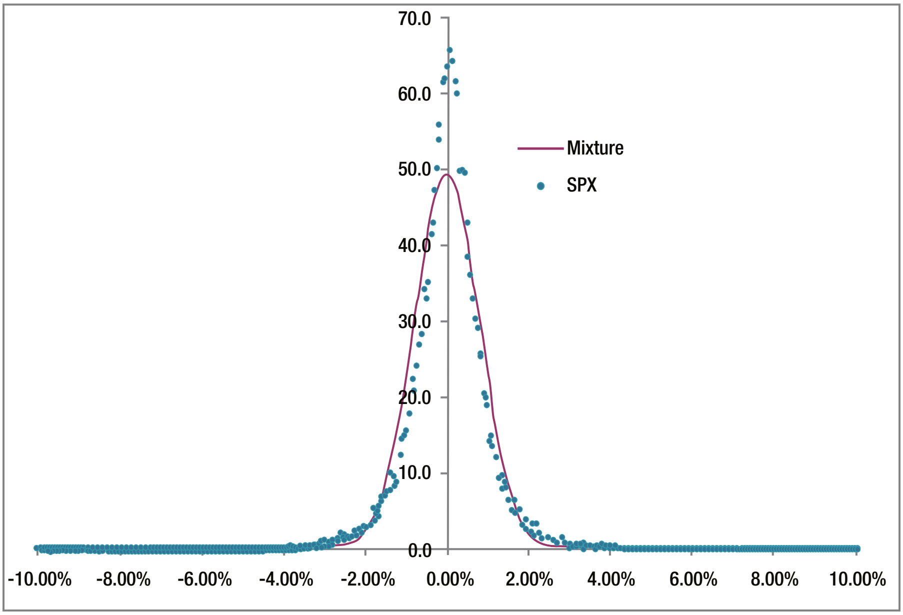

Figure 3-14. *S&P 500 discrete empirical density function versus mixed Gaussian*

$$\sigma_{theoretical}(q,\sigma_1)=\sqrt{q\sigma_1^2+(1-q)\sigma_2^2}=\sigma_{empirical} \tag{3.72}$$

$$\mathrm{kurt}_{theoretical}(q,\sigma_1)=\frac{3(q\sigma_1^4+(1-q)\sigma_2^4)}{[q\sigma_1^2+(1-q)\sigma_2^2]^2}=\mathrm{kurt}_{empirical} \quad . \tag{3.73}$$

Unfortunately, these equations cannot be inverted explicitly to solve for q and σ_1 and must be solved by alternate means, such as those described in the next two sections.

Fitting by Hand

i. Create a graph of both the empirical density function and the theoretical mixed normal density function, as illustrated in Figure 3-14. (See Problem 3-3.)

ii. Adjust σ_1 and q to best-fit the standard deviation and kurtosis of the empirical data. Do this by adjusting (by hand) these parameters until the theoretical moments match the empirical ones. A visual inspection of the graph will help. This fit will not be perfect.

Chi-Squared Fitting

Use the Solver function in Excel to calibrate σ_1, σ_2, and q to the moments. Create the following weighted *chi-squared* function

$$\chi^2(q,\sigma_1) = w_1 \cdot [\text{kurt}_{theoretical}(q,\sigma_1) - \text{kurt}_{empirical}]^2 + w_2 \cdot [\sigma_{theoretical}(q,\sigma_1) - \sigma_{empirical}]^2 \quad , \tag{3.74}$$

which needs to be made as small as possible to have a good match of moments. This method is called *chi-squared minimization*, (3.75). The weights w_i are chosen to scale the levels properly, such that each term in the chi-squared function is on equal footing (see Problem 3-3). Use the Solver function in Excel to solve for σ_1 and q, as shown in Figure 3-15.

$$\underset{\text{By changing } q \,\&\, \sigma_1}{\text{Minimize}} \ \chi^2(q,\sigma_1) \tag{3.75}$$

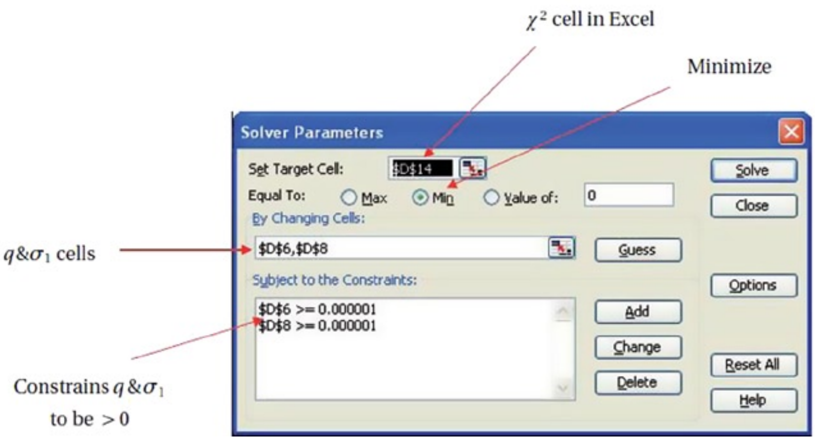

Figure 3-15. *The Solver function in Excel*

Calibrating a Generalized Student's-*t* Distribution to Equity Returns

The basic Student's-*t* density function is given by one parameter ν,

$$f(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad . \tag{3.76}$$

The generalized version is given by three parameters μ, ν, and λ. The mean is captured by μ, whereas λ provides a length scale,

$$f(t)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\sqrt{\frac{\lambda}{\nu\pi}}\left(1+\frac{\lambda(t-\mu)^2}{\nu}\right)^{-\frac{\nu+1}{2}} \quad . \qquad (3.77)^{5}$$

Note that t is the random variable keeping with tradition and should not be confused with time. The mean, variance, and kurtosis of this distribution are given by

$$E[t]=\mu, \quad \mathrm{Var}[t]=\frac{1}{\lambda}\frac{\nu}{\nu-2}, \quad \mathrm{kurt}[t]=\frac{6}{\nu-4}+3 \quad . \qquad (3.78)$$
$$\nu>4$$

This distribution has no skew. Note that the variance only exists for $\nu > 2$, whereas the kurtosis exists for $\nu > 4$. Unlike the mixed Gaussian distribution, the moment-matching equations

$$\mathrm{Var}[t]=\frac{1}{\lambda}\frac{\nu}{\nu-2}=\sigma^2_{empirical} \qquad (3.79)$$

$$\mathrm{kurt}[t]=\frac{6}{\nu-4}+3=\mathrm{kurt}_{empirical} \qquad (3.80)$$

can be inverted explicitly to give the following moment-matching formulas:

$$\begin{aligned}\nu &= \frac{6}{\mathrm{kurt}_{empirical}-3}+4 \\ \lambda &= \frac{1}{\sigma^2_{empirical}}\left[\frac{\nu}{\nu-2}\right]\end{aligned} \qquad (3.81)$$

Figures 3-16 and 3-17 illustrate a method of moments calibration of the Euro Stoxx 50 index known as SX5E (see Problem 3-4). Figure 3-17 is a log graph (base 10) of Figure 3-16 (y-axis) to visually bring out the fit of the tails of the distribution. The method to create a log scale in Excel is shown in Figure 3-18.

[5]Repeats (3.26).

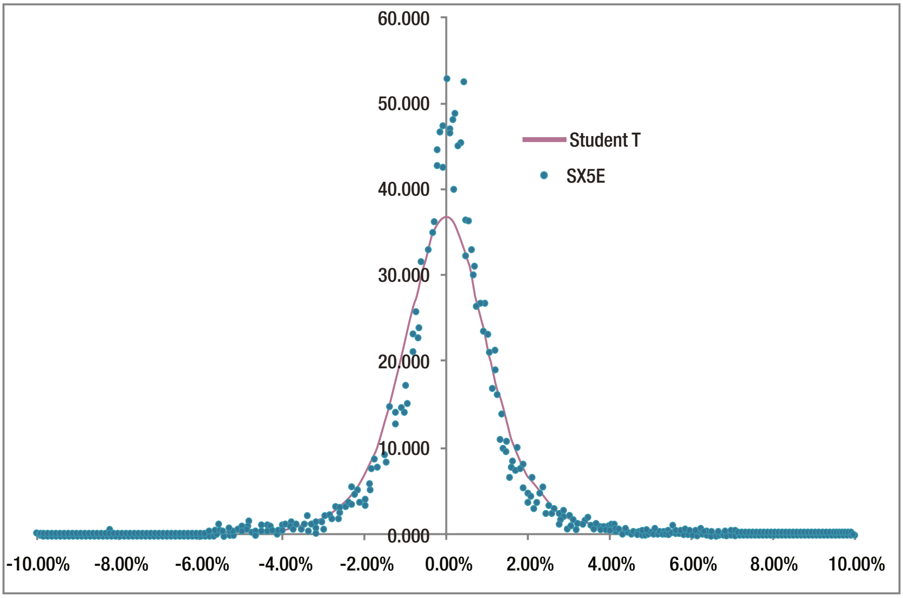

Figure 3-16. *Method of moments fit for the SX5E Index (Euro Stoxx 50) to a Student's-t distribution*

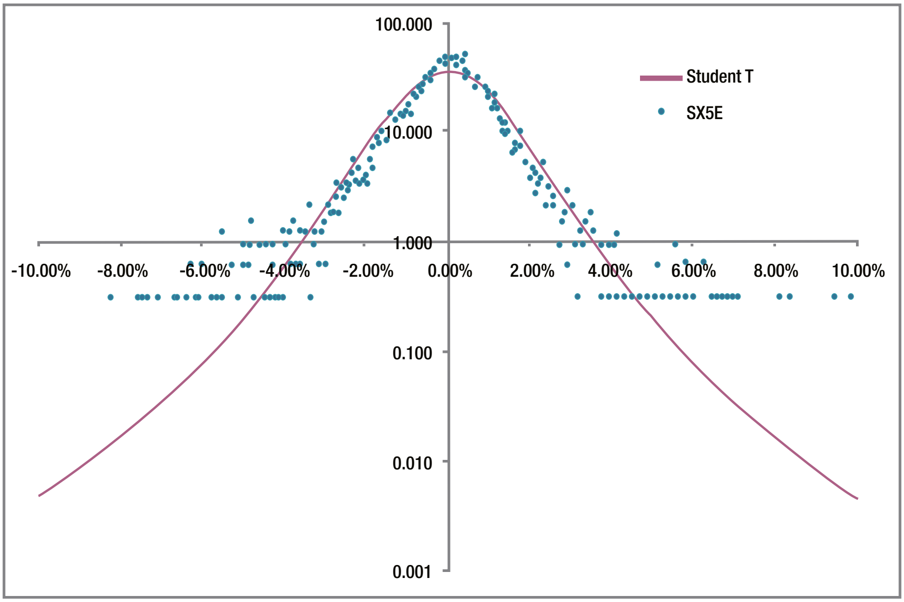

Figure 3-17. *Log graph of a method of moments fit for the SX5E Index to a Student's-t distribution*

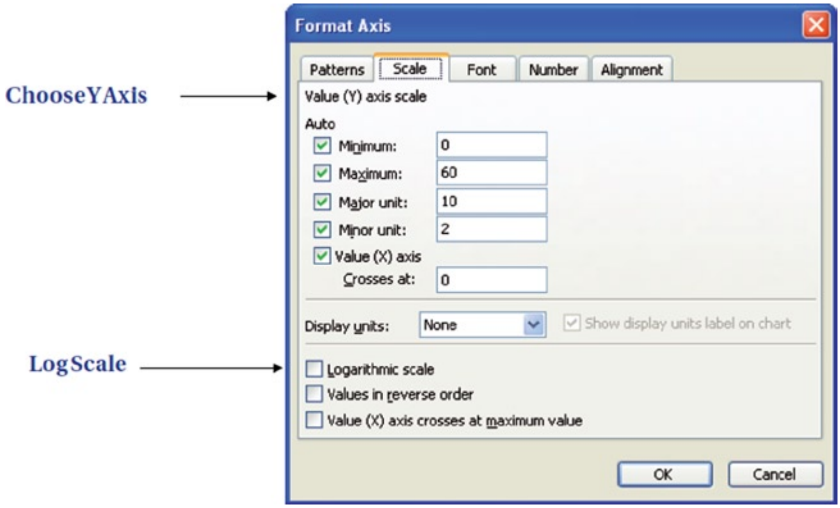

Figure 3-18. *Creating a log scale in Excel*

Calibrating a Beta Distribution to Recovery Rates of Defaulted Bonds

Probability distributions are used for all sorts of financial analysis, not just modeling financial returns. When corporate or government bonds default, they stop paying their coupon payments (see Chapter 6 for further discussion). The value of the defaulted bond does not necessarily become zero but has a value known as a *recovery rate*. This recovery rate is based on many factors, including the possibility of recovering some amount of cash from the issuer in bankruptcy court. These recovery rates range from 0% to 100% and are random variables because they are unknown until a default event happens. Therefore, modeling recovery rates is important when dealing with defaultable bonds, and one needs a distribution where the variables go from 0 to 1. An example of such a distribution is the two-parameter beta distribution introduced in (3.22),

$$f(x) = \frac{x^{p-1}(1-x)^{q-1}}{B(p,q)} \qquad x \in [0,1]$$

$$\text{Beta Function} = B(p,q) = \int_0^1 t^{p-1}(1-t)^{q-1}\, dt = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)} \tag{3.82}$$

$$\Gamma(p) = \int_0^\infty t^{p-1} e^{-t}\, dt, \quad \Gamma(p+1) = p\,\Gamma(p), \quad \Gamma(p) = (p-1)\,\Gamma(p-1)$$

Its moments are given by

$$E[X^k] = \int_0^1 x^k f(x)dx = \int_0^1 \frac{x^{(p+k)-1}(1-x)^{q-1}}{B(p,q)}dx = \frac{B(p+k,q)}{B(p,q)}$$
$$E[X] = \frac{B(p+1,q)}{B(p,q)} = \frac{\dfrac{\Gamma(p+1)\Gamma(q)}{\Gamma(p+q+1)}}{\dfrac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}} = \frac{p\Gamma(p)\Gamma(p+q)}{(p+q)\Gamma(p+q)\Gamma(p)} \tag{3.83}$$

$$E[X] = \frac{p}{p+q}$$
$$Var(X) = \frac{pq}{(p+q+1)(p+q)^2}$$
$$Skew(X) = \frac{2(q-p)\sqrt{p+q+1}}{(p+q+2)\sqrt{pq}} \tag{3.84}$$
$$Kurt(X) = \frac{6[(p-q)^2(p+q+1)-pq(p+q+2)]}{pq(p+q+2)(p+q+3)} + 3$$

Assuming one has an historical recovery rate data R_i per bond type (see Table 3-1), one can calculate the empirical moments,

$$\bar{R} = \frac{1}{N}\sum_{i=1}^{N} R_i, \quad \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(R_i - \bar{R})^2 \quad . \tag{3.85}$$

Using the moment-matching method to calibrate the beta distribution,

$$\frac{p}{p+q} = \bar{R}, \quad \frac{pq}{(p+q+1)(p+q)^2} = \sigma^2 \quad , \tag{3.86}$$

one can solve for the parameters p and q explicitly,

$$p = \bar{R}\left(\frac{\bar{R}(1-\bar{R})}{\sigma^2} - 1\right) \tag{3.87}$$

$$q = (1-\bar{R})\left(\frac{\bar{R}(1-\bar{R})}{\sigma^2} - 1\right) \quad . \tag{3.88}$$

Table 3-1 presents the results of this calibration.

***Table 3-1.** Statistics for Recovery Rates (%) Upon Default (1970–2000)*

	Mean	SD	p	q
Bank Loans				
Senior Secured	64.0	24.4	1.84	1.03
Senior Unsecured	49.0	28.4	1.03	1.07
Bonds				
Senior Secured	52.6	24.6	1.64	1.48
Senior Unsecured	46.9	28.0	1.02	1.16
Senior Subordinate	34.7	24.6	0.95	1.79
Subordinate	31.6	21.2	1.20	2.61
Junior Subordinate	22.5	18.7	0.90	3.09
Preferred Stock	18.1	17.2	0.73	3.28

Source: A. M. Berd and V. Kapoor, "Digital Premium," Journal of Derivatives, 10 (2003), 66–76.

Figures 3-19 and 3-20 illustrate two calibrated density functions of recovery rates.

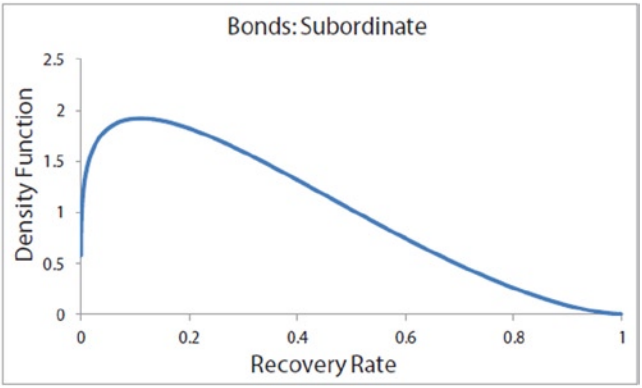

***Figure 3-19.** Beta distribution for recovery rates from defaulted subordinate bonds*

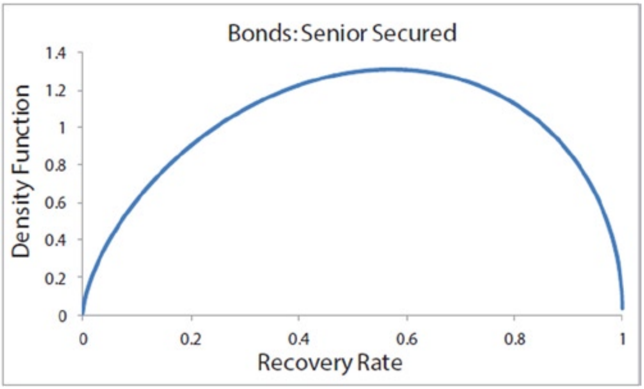

Figure 3-20. *Beta distribution for recovery rates from defaulted senior secured bonds*

Basic Risk Measures

Measuring and controlling risk is a major concern for all financial institutions. With better measurement, an institution can chart and take a more prudent path to risk management. Financial risk is typically associated with the statistical uncertainty of the final outcome of the valuation of financial positions. An objective measure of this risk is needed to handle the "rare events" in which the true financial risk resides. No single risk measurement can give the complete picture of potential losses in the future. There are generically three large types of financial risk: *market risk*, *credit risk*, and *operational risk*. Furthermore, many subdivisions of these risk types exist. Market risk encompasses all the well-known risks that attend the trading of financial securities, such as the price risk of stocks, the volatility risk of options (Chapter 5), the interest risk of fixed income instruments, the foreign exchange risk of currency transactions, and the prepayment risk of mortgages. Credit risk deals with losses arising from defaults and bankruptcy. Operational risk deals with losses arising from business operations failures such as in the informational technology (IT) area or the legal department. Credit and operational risk will be further explored in Chapters 6 and 7.

The *probability of extreme losses* is of great concern. A typically used but poor risk measurement of extreme losses used on Wall Street and prescribed by many regulators around the globe is the *value at risk* (VaR). Suppose one has a distribution of daily returns of some asset (say the S&P 500). One can ask, "What is the probability of a daily return being less than 7%?" Statistically, this means

$$F(-0.07) = P(r \le -0.07) = \int_{-\infty}^{-0.07} f(r)\,dr \quad . \tag{3.89}$$

■ **Caution** This value is highly dependent on the tail of the distribution $F(x) = P(X \le x)$ that is being used to model the returns (such as Gaussian, mixed Gaussian, or Student's-t).

The more typical question is, "What is the loss amount (negative return) at the 95% confidence interval?" This is referred to as a *95% VaR number*. It means "What is the return level r_{95} where 95% of the returns are greater than this level?" Mathematically, this is

$$1-0.95=0.05=P(r\le r_{95})=\int_{-\infty}^{r_{95}} f(r)\,dr \quad , \tag{3.90}$$

or equivalently

$$0.95=P(r>r_{95})=\int_{-r_{95}}^{\infty} f(r)\,dr \quad . \tag{3.91}$$

VaR is the level of loss (e.g., r_{95}) corresponding to a certain probability of loss (e.g., 5%). VaR is usually calculated at the 95%, 99%, and 99.9% levels. These levels are highly dependent on the tail of the distribution. Figure 3-21 shows a distribution function of returns with the VaR(95) point indicated.

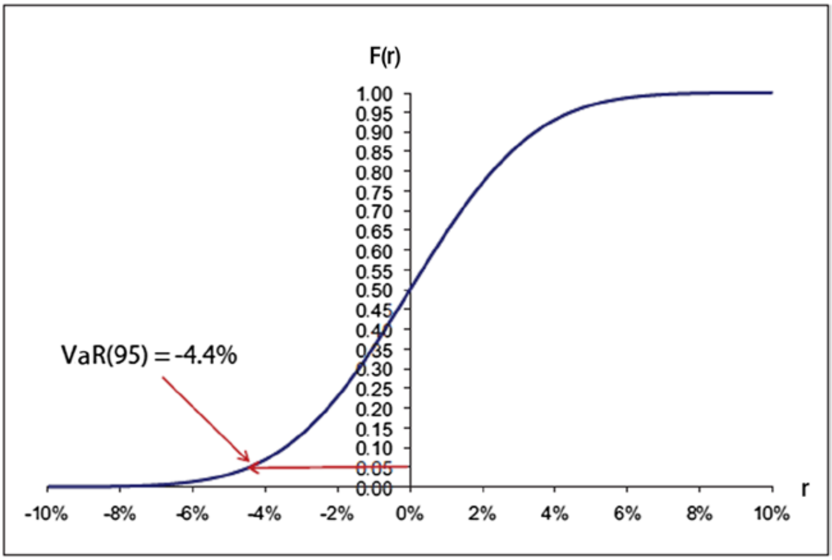

***Figure 3-21.** A cumulative distribution function of returns with VaR(95) indicated*

Generally speaking, given a distribution of returns $F(r)$ and a confidence level (percentile) α, VaR (α) is the smallest number r_α such that the probability of the return r is less than r_α is at most 1-α,

$$\mathrm{VaR}_\alpha=\sup\{r_\alpha\in\Re : P(r\le r_\alpha)\le 1-\alpha\}=\sup\{r_\alpha\in\Re : F(r_\alpha)\le 1-\alpha\} \quad . \tag{3.92}$$

■ **Caution** Much of the literature and nomenclature on Wall Street deals with "loss" distributions in which the loss is mathematically a positive number, such as "a loss of 100K." The foregoing definition for VaR must be changed appropriately.[6]

One of the main problems with VaR is that it cannot answer the following question: "If the VaR(80) return is (negatively) exceeded (a 20% chance), what is the realized loss?" This VaR tells one the cut-off at which 80% of the returns are higher than the cut-off. The real question is, what are the negative returns in the remaining 20% of the distribution? This number will depend on the thickness of that tail of the distribution. Figure 3-22 shows two distributions having the same VaR(80) point but clearly different loss levels once this common point is reached. The fat-tailed distribution has a thicker tail and is more risky than the normal distribution even though they both have the same VaR(80).

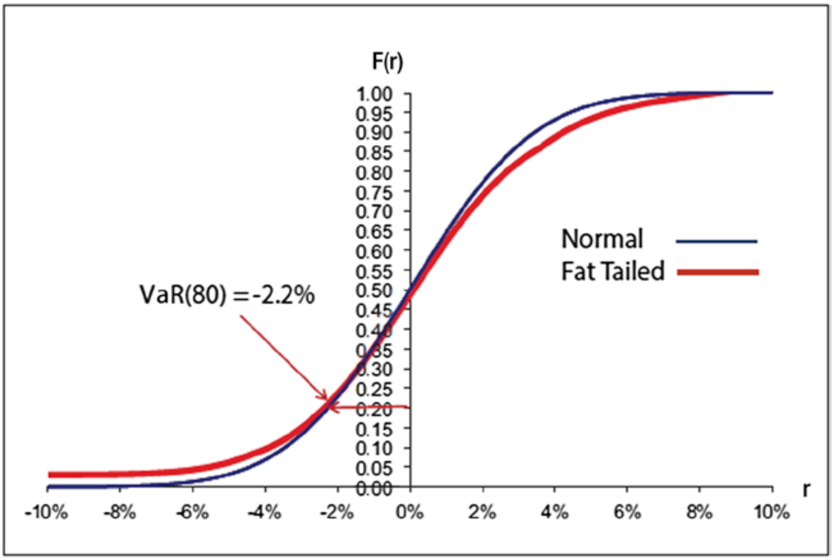

Figure 3-22. *A comparison of two cumulative return distribution functions with identical VaR(80) points indicated*

[6]See A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management.* Princeton University Press, 2005.

The most straightforward way to get a better risk measure is to calculate the area under the tail to the left of the VaR_a point. That is, "What is the average of the losses *if* the VaR_a point is exceeded?" This is called *conditional VaR* (CVaR) *or expected shortfall*,

$$\text{CVaR}(\alpha) = \frac{\int_{-\infty}^{r_\alpha} r \cdot f(r)\,dr}{\int_{-\infty}^{r_\alpha} f(r)\,dr} \quad . \tag{3.93}$$

Note that this formula is similar to taking the expectation value of r in the region $-\infty$ to r_α except the density function is *not* normalized in this region. Therefore, one must divide by the area of this region, which is the denominator in (3.93). For instance, CVaR(95) equals

$$\text{CVaR}(95) = \frac{\int_{-\infty}^{r_{95}} r \cdot f(r)\,dr}{\int_{-\infty}^{r_{95}} f(r)\,dr} = \frac{\int_{-\infty}^{r_{95}} r \cdot f(r)\,dr}{0.05} \quad . \tag{3.94}$$

Calculating VaR and CVaR from Financial Return Data

Estimating VaR and CVaR from return data, whether historic or simulated, is a straightforward process. This section describes how to calculate VaR and CVaR from historical data as an educational exercise. In reality, all risk is in the future and therefore VaR and CVaR are calculated using simulations of events in the future, which are described in Chapter 4.

Note that the time scale of the returns dictates the time scale of the risk measure. Daily returns provide daily VaR and CVaR numbers. Weekly returns provide weekly VaR and CVaR numbers. Here are the following steps for historical VaR and CVaR:

1. Get the set of return data (historic or simulated).
2. Sort the returns from lowest to highest and number them as in Figure 3-23. The sorted data are called the "order statistics" of the set of return data, where 1 = L(1) = the smallest value among the returns, 2 = L(2) = the next larger value, ...

1	-22.928%
2	-9.498%
3	-9.382%
4	-9.229%
5	-8.670%
6	-7.951%
7	-7.141%
8	-7.072%
9	-7.037%
10	-6.977%

Figure 3-23. *Sorted and numbered returns (a count of 15,432)*

3. Get the total number of returns (15,432 in the example in Figures 3-23 to 3-25).

4a. Approximate the VaR(99) as $\rightarrow L(\lfloor(1-99\%)*(Total\ Count)\rfloor) = L(\lfloor .01*15432 \rfloor = L(154) = 2.638\%$. It is an approximation because the data are discrete and finite.

4b. Approximate the VaR(99.9) as $\rightarrow$ $L(\lfloor(1-0.999)*\text{Total Count}\rfloor) = L(\lfloor .001*15432\rfloor) = L(15) = -6.324\%$

151	-2.647%	12	-6.876%
152	-2.645%	13	-6.341%
153	-2.643%	14	-6.339%
154	-2.638%	15	-6.324%
155	-2.634%	16	-6.033%
156	-2.633%	17	-5.939%
157	-2.625%	18	-5.560%

Figure 3-24. *VaR(99) and VaR(99.9) sorted returns*

5a. CVaR(99)

$$\text{CVaR}(99) = \frac{\int_{-\infty}^{r_{99}} r \cdot f(r)\,dr}{0.01} \approx \left(\frac{1}{154}\right)\sum_{i=1}^{154} return \quad . \tag{3.95}$$

5b. CVaR(99.9)

$$\text{CVaR}(99.9) = \frac{\int_{-\infty}^{r_{99.9}} r \cdot f(r)\,dr}{0.001} \approx \left(\frac{1}{15}\right)\sum_{i=1}^{15} return \tag{3.96}$$

$$\text{Sum} = -128.70\%$$

$$\text{CVaR}(99.9) \approx \left(\frac{1}{15}\right)\text{Sum} = -8.580\% \tag{3.97}$$

1	-22.928%	
2	-9.498%	
3	-9.382%	
4	-9.229%	
5	-8.670%	
6	-7.951%	
7	-7.141%	
8	-7.072%	Average these for CVaR(99.9)
9	-7.037%	
10	-6.977%	
11	-6.937%	
12	-6.876%	
13	-6.341%	
14	-6.339%	
15	-6.324%	VaR(99.9)

Figure 3-25. *Calculating VaR(99.9) and CVaR(99.9) from sorted returns*