Chapter 3. Quantitative Finance Tools

"*Beware (of) geeks bearing gifts.*" Stephen Harper (1994)[1]

Introduction

In this chapter, we introduce several important quantitative tools and illustrate how they are used in finance. One of the critical aspects of most financial analysis is time measurement. Module 3.1 reviews a variety of issues related to calendar-based calculations. Module 3.2 introduces the standard normal cumulative distribution estimation, $N(d)$, widely used in option valuation and its related inverse, $N^{-1}(d)$, widely used with dynamic risk measures (DRMs), such as value-at-risk. Monte Carlo simulations are at the heart of numerous DRMs and thus are thoroughly reviewed in Module 3.3.

Throughout this material, we seek to apply Occam's razor where we seek the simplest solution with the minimal number of related factors. Instrumental in this approach is the LSC model denoted for level, slope, and curvatures or LSC. Module 3.4 introduces the LSC model and thoroughly applies it to various term structure of interest rates.

Module 3.5 briefly introduces data sorting as well as selected data insights related to stock prices. Many finance applications involve solving for embedded parameters, so we introduce useful solving routines. One example is solving for the yield to maturity for a bond. We explore embedded parameter estimation in Module 3.6.

We conclude this chapter, with exhaustive coverage of the normal and lognormal distributions as they are the most widely used in finance.

[1]Quoted by Don Chance at his very popular web site that contains numerous great quotes. See http://www.bus.lsu.edu/academics/finance/faculty/dchance/MiscProf/DerivaQuote/Qt16.htm.

Module 3.1: Managing the Calendar

Learning objectives

- Contrast the 30/360 day count method with the ACT/365 day count method
- Explain day counting with each date having a separate, sequential integer value

Module overview

Investments can be defined as the reallocation of consumption through time. Time is easy to describe in generic terms, but there are several complexities that arise when dealing with specific cases. The philosophical issues related to time are set aside here and the focus is on calculating the number of days between two dates by two methods.

Central to most finance calculations is the movement of money through time. For example, to compute the present value or future value, you need to know the length of the period for the computation. There are numerous different approaches for calculating the number of days between two dates and there are numerous different approaches for calculating the total number of days in a single year. Often, one wants to compute the fraction of the year.

For bonds, this day count is used to compute the number of days in accrued interest and the number of days in a coupon period. For financial derivatives, the method of counting days has an important influence on some derivative instrument's valuation.

Meticulously accounting for day count and payment frequency is very important. For example, consider a 5 percent rate environment and quarterly pay interest rate swap with $100,000,000 notional amount and 10-year term. If the fixed leg of the swap is actual days divided by a 360 day year, the floating leg of the swap is 30 day months divided by a 365 day year, then over $1,500,000 of swap value is attributable to just day counting differences.

Two popular day types are referred to as "ACT/365" and "30/360". Understanding these two basic types will be foundational in understanding all of the other types.

Day type ACT/365

Intuitively, this method is easy to grasp. You compute the literal number of days between two dates. To arrive at the fraction of a year, you divide by 365 (that is, ignore leap year).

Counting days is rather tedious and most software packages contain modules that take care of day counting computation for you. These modules typically use the 'Julian' date method. The Julian method converts each day to an integer and the difference between these integers give you the correct day count.

Day type 30/360

This day counting convention is much less intuitive. The general assumption is that each month has 30 days and hence a year has 360. Obviously, this is not the case and so some adjustments are incorporated. We adopt the following notation:

M1 – month of first date,	M2 – month of second date,
D1 – day of first date,	D2 – day of second date,
Y1 – year of first date, and	Y2 – year of second date.

The following adjustments are required:

1. If D1 and D2 are the last day of February (leap year - 29, non-leap year - 28), then change D2 to 30.
2. If D1 is the last day of February, then change D1 to 30.
3. If D1 is 30 or 31 and D2 is 31, then change D2 to 30.
4. If D1 is 31, then change D1 to 30.

After these adjustments, the number of days between two dates is

$$\text{Day Count} = (Y2 - Y1) * 360 + (M2 - M1) * 30 + (D2 - D1).$$

For example, assume the first date is September 11, 2015, and the second date is December 15, 2015. No adjustments are required, and the number of days is

$$\text{Day Count} = (2015 - 2015) * 360 + (12 - 9) * 30 + (15 - 11) = 94.$$

Several additional points must be remembered. First, the number of days in the year is always 360 regardless of whether it is a leap year or not. Second, the number of days in a period is always 360 divided by the number of periods in a year. For example, if a bond pays quarterly, then the number of days in a quarter is 360/4 = 90. Third, the day count procedure is used to compute accrued interest within a period. Hence, the remaining days in a period is just the number of days in the period minus the number of days that have accrued. This point will be covered in more detail when we discuss interest calculations.

Selected other day types

There are a vast number of other day types and day counting conventions. Here a just a few.

ACT/ACT: This day type requires the actual number of days be computed for the period as well as the accrual days. Hence, a leap year would make a difference.

ACT/360: This day type requires the actual number of days be computed for the accrual period but assumes a 360 day year. Eurodollar futures contracts use this method.

30/ACT: This day type requires the actual number of days be computed for the entire period, but the accrual period is computed using the 30/360 method.

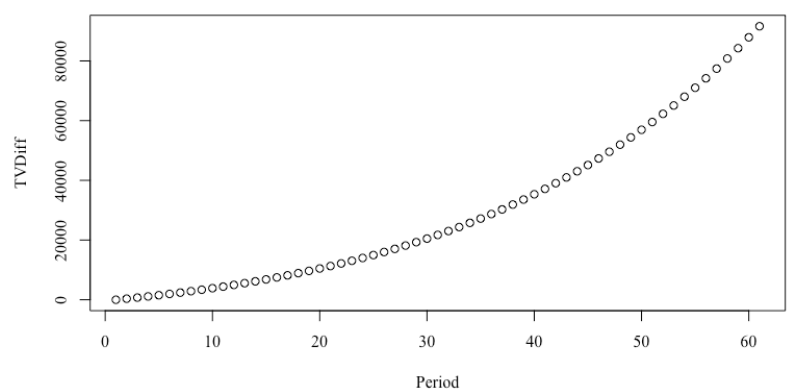
30/365: This day type requires the 30/360 method for the accrual period but assumes a 365 day year.

End-of-month rule

Almost all debt securities maturing at the end-of the month followed what is called the end-of-month rule. For example, if a U.S. Treasury security matures on June 30th and is semi-annual coupon paying, then the other coupon date is December *31st*. Historically, some securities issued by the Federal Home Loan Bank do not follow the end-of-month rule. For these securities, if they are semi-annual coupon paying and it matures on June 30th, then the other coupon date is December 30th. It is important to pay attention to the details.

Several other calendar functions could be developed. Figure 3.1.1 illustrates the future dollar value of the difference between ACT/360 day count and ACT/365 day count with a 5% interest rate, quarterly compounding and $1 million notional amount. Clearly, day counts are an important consideration.

Figure 3.1.1 Future dollar value difference between ACT/360 day count and ACT/365
Five percent interest rate, quarterly compounding, and $1 million notional amount



Module 3.2. Cumulative Normal Distribution Function and its Inverse

Learning objectives

- Explain how to compute the cumulative distribution function of the normal distribution
- Understand approximations for N(d) and $N^{-1}(d)$
- Introduce an iterative test routine for functions and their inverses

Module overview

Many option valuation models rely on the ability to solve for the value of the cumulative normal distribution (CDF, denoted $N(d)$) when given the upper limit or percentage point (d). For example, the standard Black, Scholes, Merton option valuation model (BSMOVM) had two calculations usually denoted $N(d_1)$ and $N(d_2)$. Many risk management calculations rely on the ability to perform the inverse calculation, $N^{-1}(d_1)$. That is, estimate the percentage point d, when given the prescribed CDF probability $N()$.

Computing N(d)

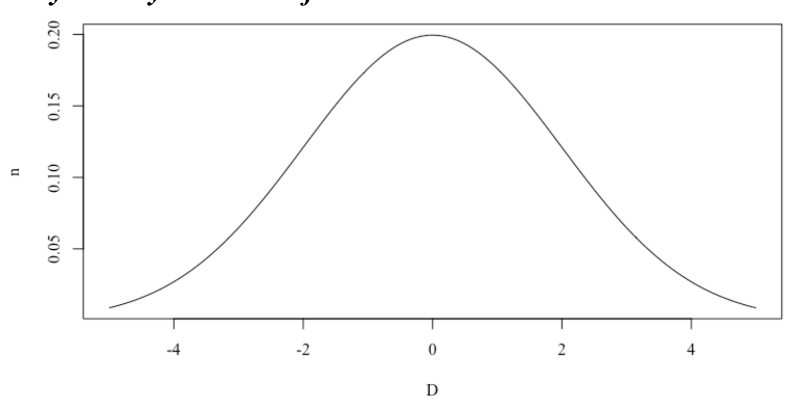
The solution to the CDF value, denoted $N()$, given a known value of the percentage point (d), can be expressed as

$$N\left(d\right)=\int_{-\infty}^{d}\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}dx. \tag{3.1}$$

Often, option models such as the BSMOVM were labeled ‘closed form’ equations. “In mathematics, a closed-form expression is a mathematical expression that is formed with constants, variables and a finite number of standard operations and functions, such as +, –, ×, ÷, *n*th root, exponentiation, logarithm, trigonometric functions, and inverse hyperbolic functions. Usually, no limits or integrals are accepted.”[2] It is generally assumed that the solution for the expression above for $N(d)$ is ‘well-known.’ Thus, we need to know it. Unfortunately, an exact analytic expression to solve this open integral does not exist. With a computer, however, it can be easily solved.

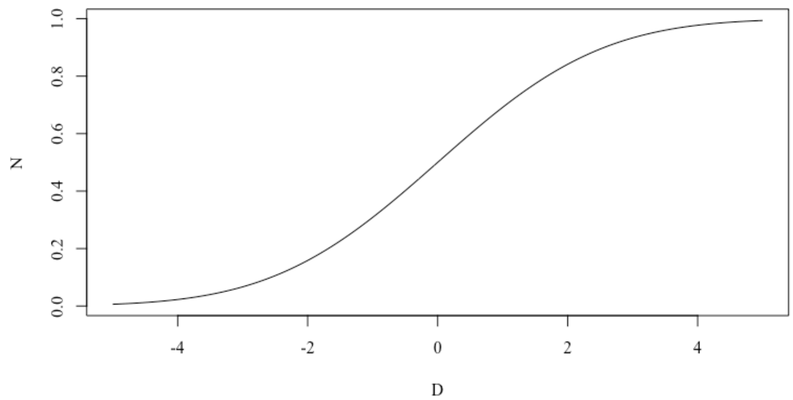
Note that the range of the percentage point d is $-\infty < d < \infty$ and the range of the CDF probability is $0 \le N(d) \le 1$. The probability density function (PDF) is illustrated in Figure 3.2.1 and the cumulative distribution function (CDF) is illustrated in Figure 3.2.2. The purpose of the R code illustrated here is to estimate $N(d)$, given d (using the method ND(d)) or to estimate d given $N(d)$ (using the method D(n)). In the example below, $d = -1.644853$ and $N(d) = 0.05$.

Figure 3.2.1. Probability Density Function of Standard Normal Distribution



[2]Closed-form expression, www.wikipedia.com, July 5, 2023.

Figure 3.2.2. Cumulative Distribution Function of Standard Normal Distribution



This CDF can be approximated by (d>0)

$$N\left(d\right)=0.5+\frac{1}{\pi}\sum_{r=0}^{12}\frac{e^{-\frac{\left(r+\frac{1}{2}\right)^2}{9}}\sin\left\{\frac{d\left(r+\frac{1}{2}\right)\sqrt{2}}{3}\right\}}{\left(r+\frac{1}{2}\right)}. \tag{3.2}$$

According to Stuart and Ord, this approximation is accurate to nine decimal places. The method, ND(d), is based on this approximation. See Alan Stuart and J. Keith Orr, *Kendall's Advanced Theory of Statistics*, 5th Edition, Volume 1 Distribution Theory, page 185.

Computing $N^{-1}(d)$

Wichura (1988) provides a FORTRAN version of an approximation for the inverse normal CDF computation that he claims is accurate to the 16th decimal place. The method, D(n), is based on this approximation converted from FORTRAN.

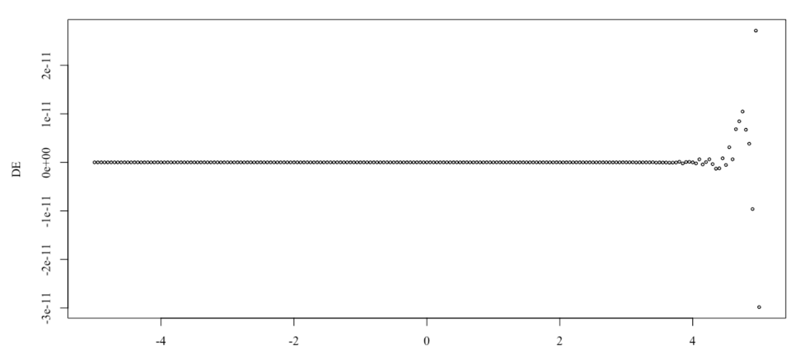
Wichura (1988) breaks the CDF into two regions below $N(d)$ < 0.5. Within these two regions, seven polynomials are estimated. The goal is to provide an accurate algorithm for estimating d, given $N(d)$. The interface code discussed below provides a test of the accuracy of the D(n) and ND(d) programs.

Estimation error illustrated

Estimation error illustrated

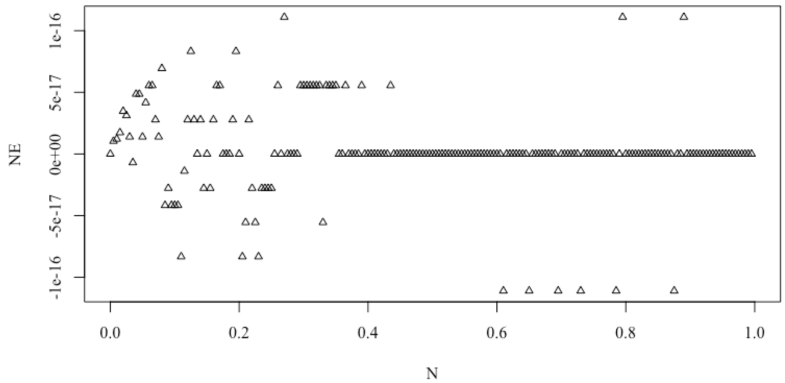
With the development of estimation methods for both $N(d)$ as well as its inverse, we can examine estimation error. Figure 3.2.1 illustrates the estimation error produced by starting with a value of d, estimating the value of $N(d)$, and then estimating based on this $N(d)$ the value of d.

Figure 3.2.1. Estimation error in D



Similarly, Figure 3.2.2 illustrates the estimation error produced by starting with a value of $N(d)$, estimating the value of d, and then estimating based on this d the value of $N(d)$.

Figure 3.2.2. Estimation error in N



In either case, the estimation error is essentially zero.

Module 3.3. Univariate Random Numbers

Learning Objectives

- Explain how to generate univariate random numbers, specifically uniform and normally distributed random variables
- Develop the capacity to generate random numbers indicating likelihood
- Illustrate how to generate random numbers for simulating rare events

Module overview

The ability to generate random numbers is very important in financial analysis. We use random number generating in several modules related to dynamic risk management.

Of course, generating random numbers with a deterministic computer poses some technical problems. Hence, the purist will refer to generating "pseudo" random numbers. For ease of exposition, the pseudo disclaimer is dropped. The main objective here is to introduce the ability to generate random numbers using R. Four different random number generating tasks are presented here, uniform integer, uniform real, likelihood, and normal. Each random number generator has to be initialized with a seed value—a task embedded in the R function.

Computing Uniform Integer Random Numbers

Whenever interacting with integers, it is important to remember the numerical limitations of integers. In R, the upper bound is 2,147,483,647 and lower bound is –2,147,483,647.[3]

The discrete uniform distribution has a finite set of possible outcomes and each outcome is equally likely. The parameters of the discrete uniform distribution in this application are the lower bound (L) and upper bound (U). Each integer within and including the bounds are assumed to be equally likely. The population mean is $(L + U)/2$ and the variance is $[(U - L + 1)^2 - 1]/12$.

Figure 3.3.1 illustrates a random sample based on a uniform distribution with limits of zero and one.

Figure 3.3.1. Random sample illustrations of a uniform distribution with limits of zero and one

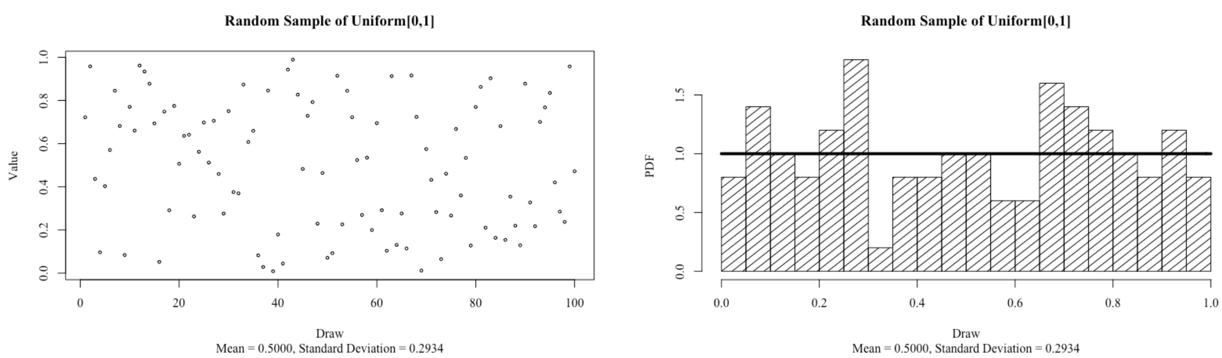
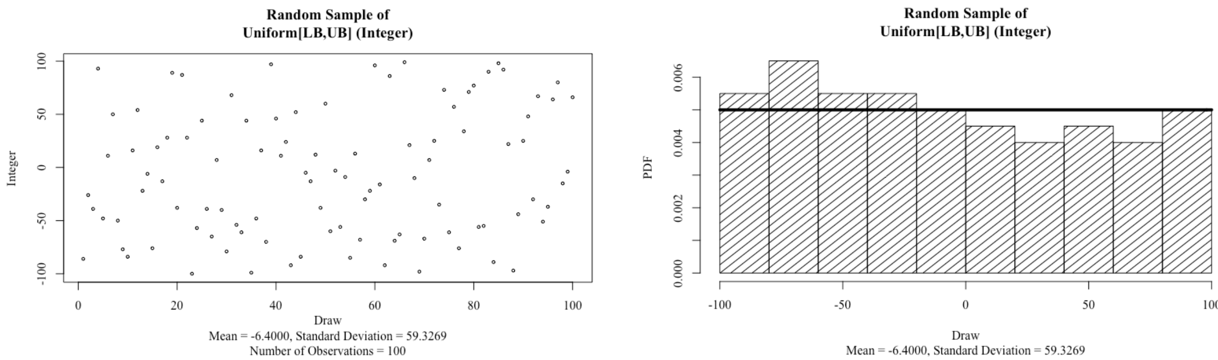


Figure 3.3.2 illustrates a random sample based on a uniform integer distribution with limits of –100 and 100.

[3]These limits may change over time.

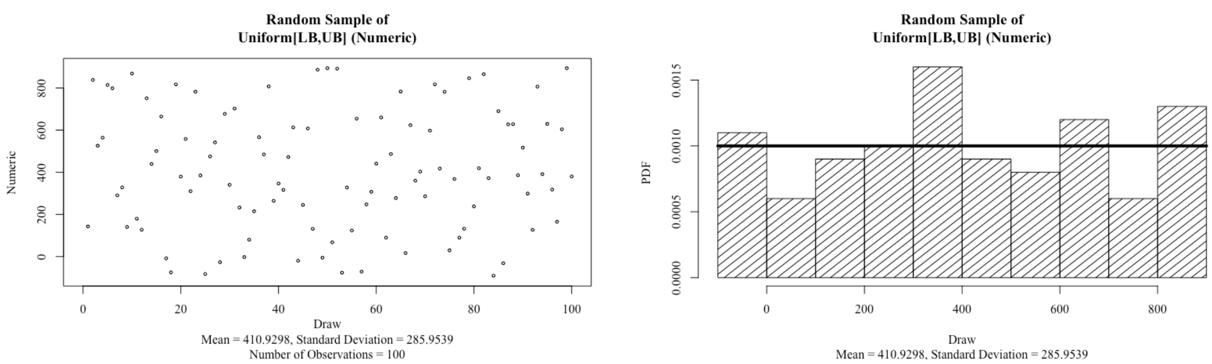
Figure 3.3.2. Random sample illustrations of a uniform integer distribution with limits of –100 and 100



Computing Uniform Real Random Numbers
The continuous uniform distribution has an infinite set of possible outcomes. The parameters of the continuous uniform distribution in this application are the lower bound (L) and upper bound (U). The population mean is (L + U)/2 and the variance is $(U – L)^2/12$.

Figure 3.3.3 illustrates a random sample based on a uniform real distribution with limits of –100 and 900.

Figure 3.3.3. Random sample illustrations of a uniform real distribution with limits of –100 and 900



Computing Normal Random Numbers
The continuous normal distribution has an infinite set of possible outcomes. The parameters of the continuous normal distribution in this application are the population mean and the standard deviation.

Figure 3.3.4 illustrates a random sample based on a normal distribution with mean zero and standard deviation one.

Figure 3.3.4. Random sample illustrations of a normal distribution with mean 0 and standard deviation 1

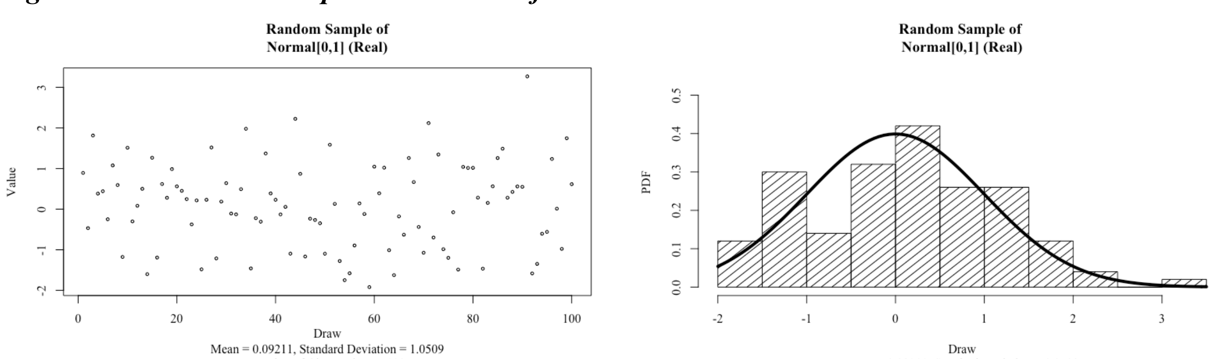
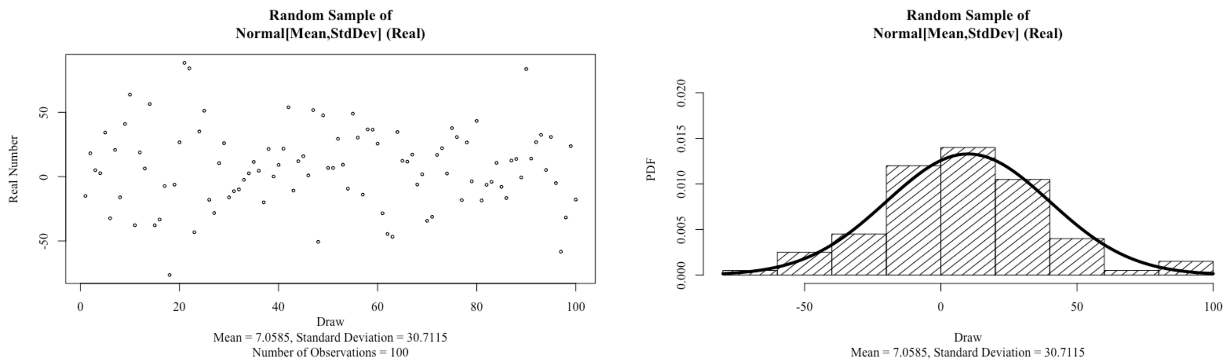


Figure 3.3.5 illustrates a random sample based on a normal distribution with mean 10 and standard deviation 30.

Figure 3.3.5. Random sample illustrations of a normal distribution with mean 10 and standard deviation 30

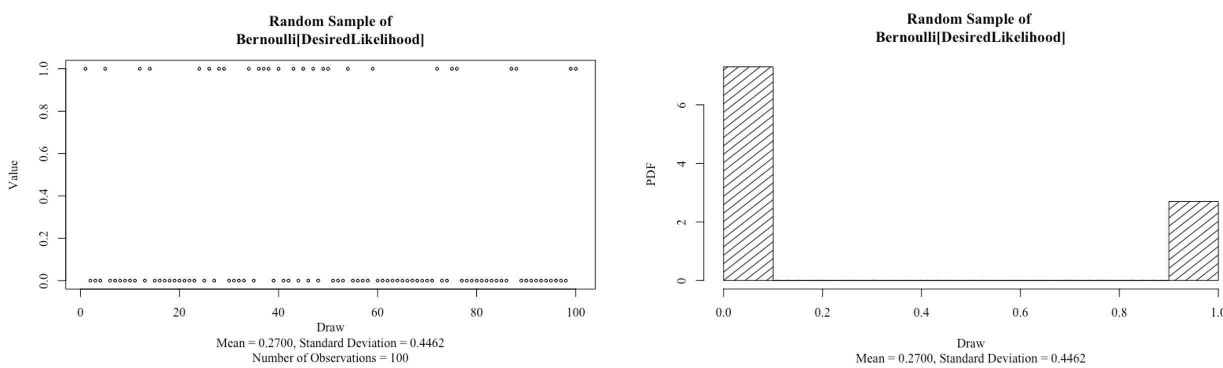


Computing Likelihood

The method to generate likelihood values are based on a uniform distribution with zero lower bound and one upper bound. With each draw, the outcome is either zero or one depending on whether the uniform draw was below the desired likelihood (DL) value. The mean is just the desired likelihood, DL, and the variance is (DL – DL^2).

Figure 3.3.6 illustrates a random sample based on a Bernoulli distribution with the desired likelihood of 25%.

Figure 3.3.6. Random sample illustrations of a Bernoulli distribution with the desired likelihood of 25%



Module 3.4. The LSC Model: Curve Fitting Using Linear Regression

Learning Objectives

- Introduce the LSC model used widely throughout this material
- Define and illustrate numerous properties of the LSC model

Module Overview

We motivate this module with three examples. Chuck was responsible for managing the risk of an interest rate swap book comprising over 60,000 different swaps tied to LIBOR. Even without knowledge of swaps, it is important to know this swap book had numerous future complex cash flows on almost every single day over the next 30 years. How should Chuck estimate the base discount rate to apply for every cash flow? Further, to manage the resultant risk, Chuck needs a parsimonious model (that is, one with only a few parameters to measure and manage).

Stephen, a quant in the Treasury Department of a national bank is responsible for $800 million U.S. Treasury portfolio with over 30 different bonds. How does Stephen measure and manage this risk?

Andy, a quant in the Risk Department of a major energy company, has identified over 30,000 separate risk factors from over five commodities, 120 different monthly maturities, and 50 different geographic locations. How does Andy reduce the dimensionality of this risk challenge?

Chuck, Stephen, and Andy all can apply a simple regression model to parsimoniously fit various data so as to reduce the dimensionality of the challenge and hence improve decision-making. A three factor LSC model will reduce both Chuck's and Stephen's problem to three variables to manage, based on either the swap curve or the U.S. Treasury curve. Although Andy's problem is a bit more difficult, one could deploy the LSC model in various ways to reduce dimensionality to 20 to 30 major factors.

A brief review of ordinary least squares regression is provided in Appendix A to this module. We thoroughly introduce the powerful LSC model in this module. We will be illustrating the LSC model with extensive use of an R package that solves regressions of this nature. We now turn to the LSC model.

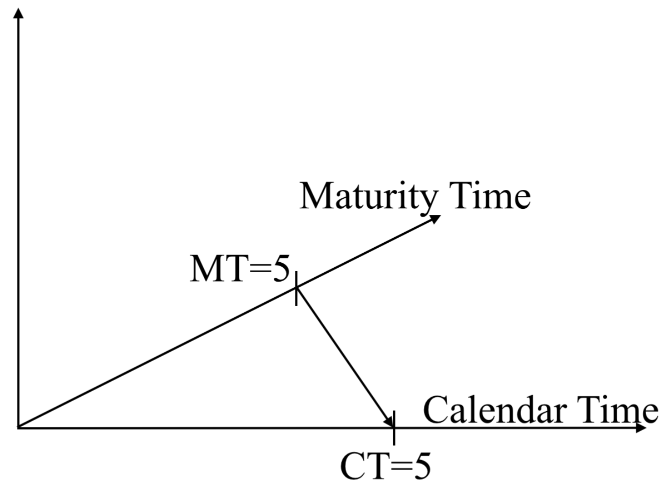
LSC model motivated through term structure of interest rates

Fitted term structure models can be viewed from two perspectives, calendar time and maturity time. The calendar time perspective is focused on the behavior of the term structure over time. For example, explaining the cross-sectional differences in bond returns is measured in calendar time, such as the past month.

The maturity time perspective is focused on the shape of the term structure at a particular point in calendar time. For example, explaining the shape of U. S. Treasury notes and bonds yields with different time to maturity. There are numerous other applications of the maturity time perspective, including the term structure of futures prices, the term structure of implied option volatility, and the term structure of dividends.

The calendar time perspective addresses the stochastic nature of the term structure of interest rates whereas the maturity time perspective is solely focused on the current observed relationship of observed yields and maturity. Although our focus is maturity time, we review the literature on both perspectives, as they are important for our purposes. Once the term structure can be reasonably estimated from a maturity time perspective, only then can the stochastic nature of the time series perspective be reasonably understood. Figure 3.4.1 illustrates these two perspectives with a five year bond.

Figure 3.4.1. Contrasting maturity time and calendar time



Fitted calendar time term structure models

Prior empirical studies of the term structure of interest rates have documented several well-known observations. Crack and Nawalkha (2000) summarize that "(u)p to 95 percent of the returns to U. S. Treasury security portfolios are explained by term-structure level shifts, slope shifts, and curvature shifts (Litterman and Scheinkman 1991; Jones 1991; Willner 1996; Jamshidian and Zhu 1997)." (p. 34)

Jamshidian and Zhu (1997) apply principal components analysis to the yield curve in three countries, Germany, Japan and the United States. They find that about 94 percent of the variation in "yield curve movements" is explained by only three components. Because this analysis was based on Riskmetrics, it is unclear how these results were influenced by various smoothing techniques. It is important to note that yield curve movements are not the same as bond returns, although they are related.

Litterman and Scheinkman (1991) examine weekly excess returns to U. S. Treasury bonds from February 22, 1984 through August 17, 1988 and finds that a three factor model explains on average 97% of the cross-sectional variation of excess bond returns. The factor model employed unobservable factors where "each factor has a mean of zero and a unit variance, and that the covariance between any two distinct factors is zero." (p. 57)

Jones (1991) reports that of the variation in U. S. Treasury bond portfolio returns, "86.6% of the return has been attributable to parallel shifts in the yield curve, 9.8% to twists, and 3.6% to butterfly changes." (p. 43) Jones' results are based on annual observations of six maturity ranges provided by the Merrill Lynch Treasury bond Indexes from 1979 through 1990.

Knez, Litterman and Scheinkman (1994) examine money market security (U. S. Treasury bills, commercial paper, certificates of deposit, Eurodollar certificates of deposit, and bankers' acceptances) returns from January 1985 to August 1988 and "find that three factors explain, on average, 86 percent of the total variation in returns and four factors explain 90 percent." (p. 1880)

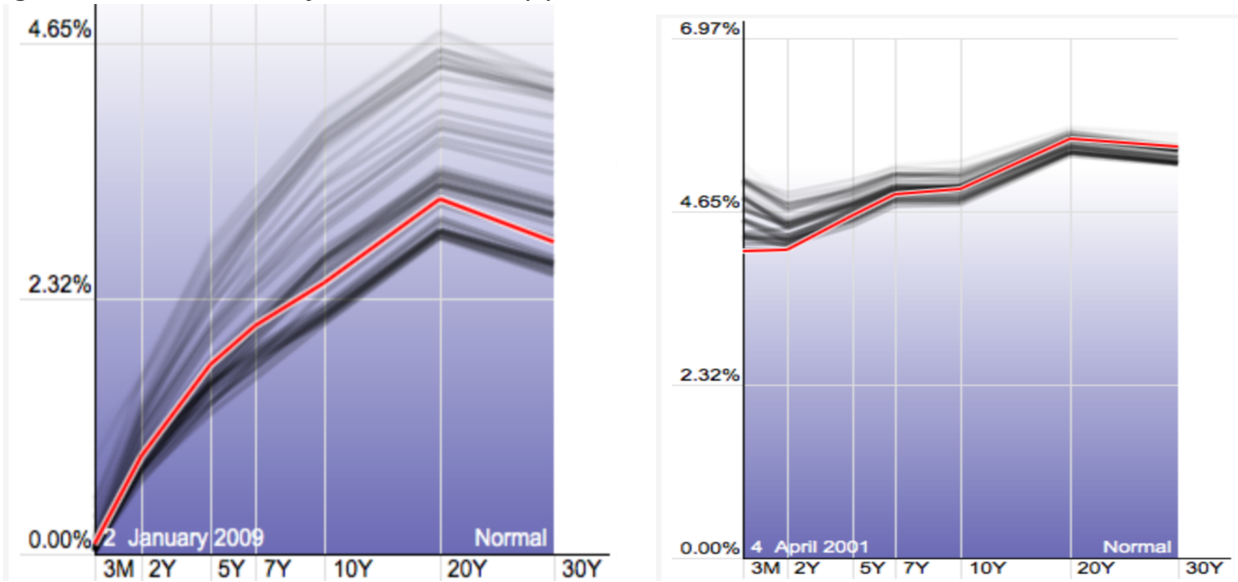
Fitted maturity time term structure models

The goal of this strand of research is to represent the term structure by some mathematical function that has desirable properties. As quoted in Nelson and Siegel (1987), Milton Friedman recognized the benefits of a parsimonious term structure model when he states, "Students of statistical demand functions might find it more productive to examine how the whole term structure of yields can be described more compactly by a few parameters."[4] There is a large literature on fitting the term structure dating at least back to Durand (1942) and includes piecewise polynomial splines (McCulloch (1971, 1975)), various parametric models (Fisher (1966), Echols and Elliott (1976), Cooper (1977), Dobson (1978), and Chambers, Carleton and Waldman (1984)), and exponential splines (Vasicek and Fong (1982)). Several authors offer subjectively drawn curves, including Woods (1983), Malkiel (1966), and Durand (1942). Figure 3.4.2 highlights the behavior of U.S.

[4]Quoted in Nelson and Siegel (1987), page 474.

Treasury yields over time. It is clearly that these yields do not move parallel, rather at time the long end of the curve is more volatility and at other times the short end of the curve is more volatile. Further, the curve is not always upward or downward sloping. It takes on a wide array of shapes.

Figure 3.4.2. Illustration of the U.S. Treasury yield curve over time[5]



Willner (1996) posits that the desirable properties of a curve fitting routine must address the bond "portfolio manager's need for *intuitive*, *descriptive*, and *comprehensive* risk exposure information." (p. 49, italics in original) Nelson and Siegel (1987) provide one such model and appeared to be motivated by the mathematical relationship between spot rates and forward rates. They put forward a parsimonious model that was "solved from differential equations describing rational interest rate behavior" (p. 50, Willner (1996)). Specifically, based on our notation

$$y_i = \sum_{j=0}^{2} x_{i,j} f_j, \tag{3.3}$$

where y_i denotes some input maturity time variables such as an interest rate for some maturity corresponding to i, $x_{i,j}$ denotes some input coefficients based on some maturity and some factor, and f_j denotes the output factors. Nelson and Siegel's original model assumed

$$x_{i,0} = 1, \; x_{i,1} = \frac{s_1}{\tau_i}\left(1 - e^{-\tau_i/s_1}\right), \text{ and } \; x_{i,2} = \frac{s_1}{\tau_i}\left(1 - e^{-\tau_i/s_1}\right) - e^{-\tau_i/s_1} \tag{3.4}$$

where s_1 denotes a scalar that applies various weights to different locations on the term structure (termed the time constant by Nelson and Siegel (1987) and the hump position parameter by Willner (1996) and "determines the rate at which the regressor variable decay to zero"[6]), $x_{i,1}$ and $x_{i,2}$ denotes maturity coefficients, a parameter that depends solely on maturity time and the selected scalar, and f_j denotes the model factor, an output parameter that is typically found using ordinary least squares regression applied to maturity time spot rates.

This model has several desirable properties when applied to the term structure of interest rates:

- As maturity approaches infinity, the spot rate approaches f_0, the level of the term structure
- As maturity approaches zero, the spot rate approaches $f_0 + f_1$, where $-f_1$ is the slope of the term structure
- f_2 measures the curvature that appears in the intermediate maturities

[5]See https://support.stockcharts.com/doku.php?id=other-tools:yieldcurve.
[6]See Nelson and Siegel (1987), page 478.

Barret, Gosnell and Heuson (1995) and Willner (1996) both report that fitted yield curve functions are not that sensitive to the choice of the scalar.[7]

Steeley (2008) used daily UK government bond coupon STRIPS from December 8, 1997 to May 15, 2002 and thoroughly examines a variety of curve fitting methodologies. Spot yield curve fitting methodologies include cubic spline, polynomial, Vasicek as well as the LSC model below (referred by Steeley as the extended Svensson model). Based on a three factor model as used below, Steeley documents that the LSC model has the lowest "average (across the sample) mean (across the curve) absolute yield error". (p. 1502) With six factors, Steeley reports that the cubic spline has the best fit, but the LSC model is a close second.

We now focus on the LSC model due to our interest in robust bond portfolio risk measures.

LSC model

An accurate methodology was developed by Svensson (1995) based on the work of Nelson and Siegel (1987). We call this approach the LSC model for level, slope, and curvature(s). We use a general form that can be expressed as

$$y_i = \sum_{j=0}^{N} x_{i,j} f_j , \tag{3.5}$$

where y_i denotes some input maturity time variables such as an interest rate for some maturity corresponding to i, $x_{i,j}$ denotes input LSC coefficients based on some maturity and some factor, and f_j denotes the output factors. The LSC model in general form assumes

$$x_{i,0} = 1 ,\ x_{i,1} = \frac{s_1}{\tau_i}\left(1 - e^{-\tau_i/s_1}\right), \text{ and } x_{i,j} = \frac{s_j}{\tau_i}\left(1 - e^{-\tau_i/s_j}\right) - e^{-\tau_i/s_j} ; j > 1 . \tag{3.6}$$

Following the literature, we assume the input scalars, s_j, are defined where $s_1 = s_2$. Again s_j denotes scalars that applies various weights to different locations on the term structure, $x_{i,j}$ denotes LSC maturity coefficients, a parameter that depends solely on maturity time and selected scalars as illustrated in Equation (3.6), and f_j denotes the output LSC factor, a parameter that is typically found using ordinary least squares regression applied to maturity time spot rates. (See Appendix A for mathematical details related to the LSC model.)

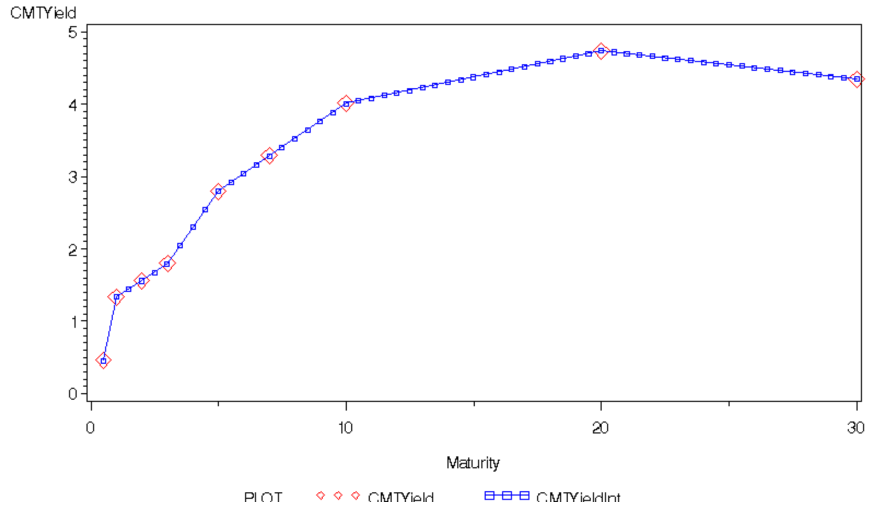
Again, note that as maturity goes to infinity, $\tau_i \to +\infty$, then $y_i \to f_0$. Thus, f_0 is interpreted as the output level parameter. As maturity goes to zero, $\tau_i \to 0$, then $y_i \to f_0 + f_1$. Thus, $-f_1$ is interpreted as the output slope parameter. Note that if the interest rate term structure is upward sloping then f_1 is negative.

To illustrate this empirical approach, consider the market information available on October 31, 2008. Figure 3.4.3 illustrates the nine CMT yields (6 month, 1, 2, 3, 5, 7, 10, 20, and 30 year) as well as 6 month spaced, linearly interpolated yields on a yearly basis between provided CMT yields. The goal is to provide a smooth set of observations from only nine CMT yields.

We first walk through a few LSC deployments that are not coded here. The goal is to understand more clearly the attributes of the LSC model. Thus, we first compute the linearly interpolated values and then fit the LSC model.

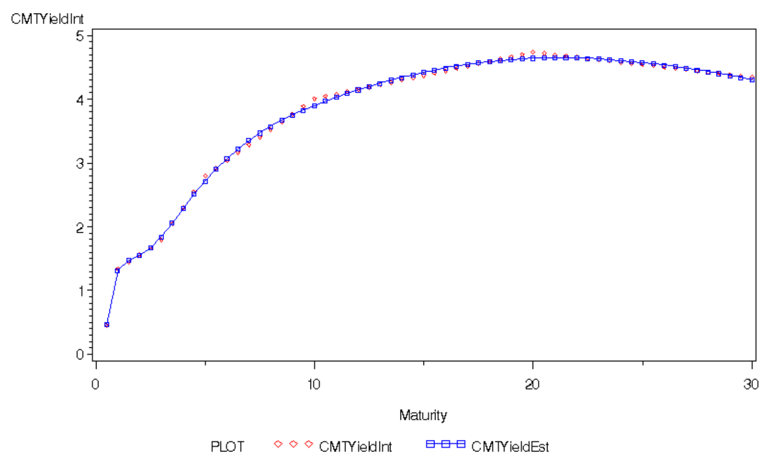
[7]See Willner (1996), p. 51.

Figure 3.4.3. Linear interpolation



To achieve the maximum fit, we chose to use a nine-factor LSC model at this point. The eight scalars used are 0.5, 1.5, 2, 5, 7, 10, 15, and 20.[8] Figure 3.4.4 illustrates this result. It is not surprising that this model fit the data very well.

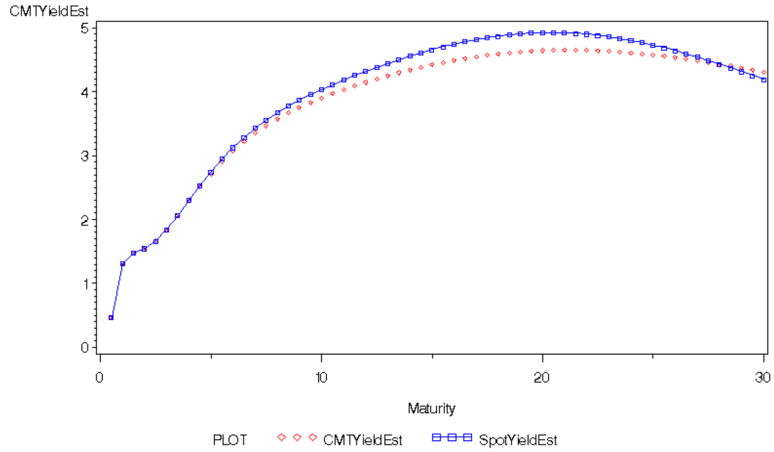
Figure 3.4.4. LSC model with nine factors



As an illustration of the LSC model and with this complete set of approximated CMT yields, we compute the implied discount factors as well as the implied, continuously compounded, spot rates. Figure 3.4.5 illustrates these implied spot rates. With this complete set of spot rates, we are now ready to estimate the three-factor LSC model defined above.

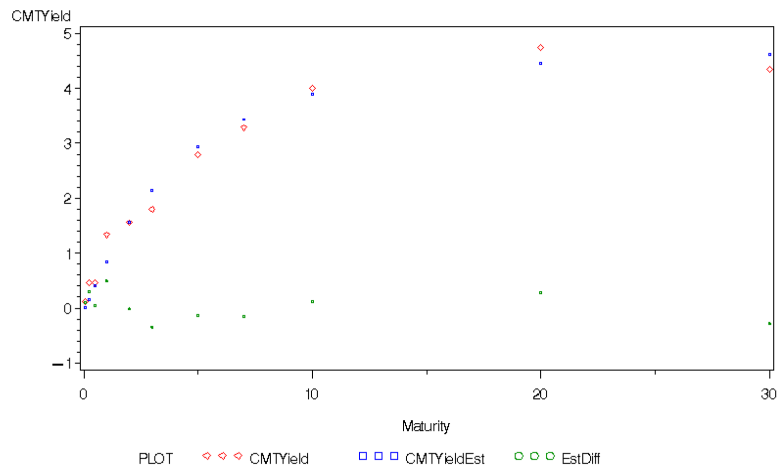
[8]The results are not sensitive to the choice of scalars, due to the large number of factors.

Figure 3.4.5. Implied spot rates inferred from LSC model with nine factors



The three-factor LSC model with a single scalar set to 3.0 is applied to the spot rates. Figure 3.4.6 presents the original CMT data (all 11 observations, including both 1- and 3-month CMT). Note that the fit is far from precise. By design, a three-factor model will not fit a complex interest rate data perfectly. One potential objective is bond risk measurement; hence, it should be clear that the LSC model is capturing more than just parallel shifts in spot rates.

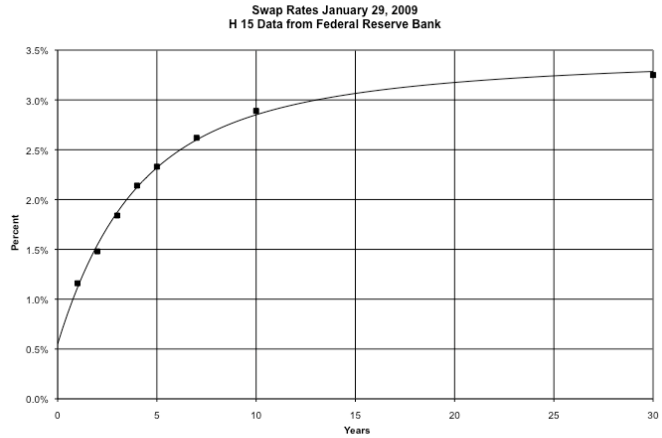
Figure 3.4.6. Original CMT compared with fitted CMT



We now examine a similar problem based on interest rate swap data. The LSC model is illustrated below with interest rate swap data for January 29, 2009. Note that swap rates are not annualized, continuously compounded spot rates. The general curve fitting approach of the LSC model works well for most shapes of the term structure. Figure 3.4.7 illustrates the three factor LSC model alone with numerical values.

Figure 3.4.7. U. S. swap rates with fitted LSC model along with numerical values

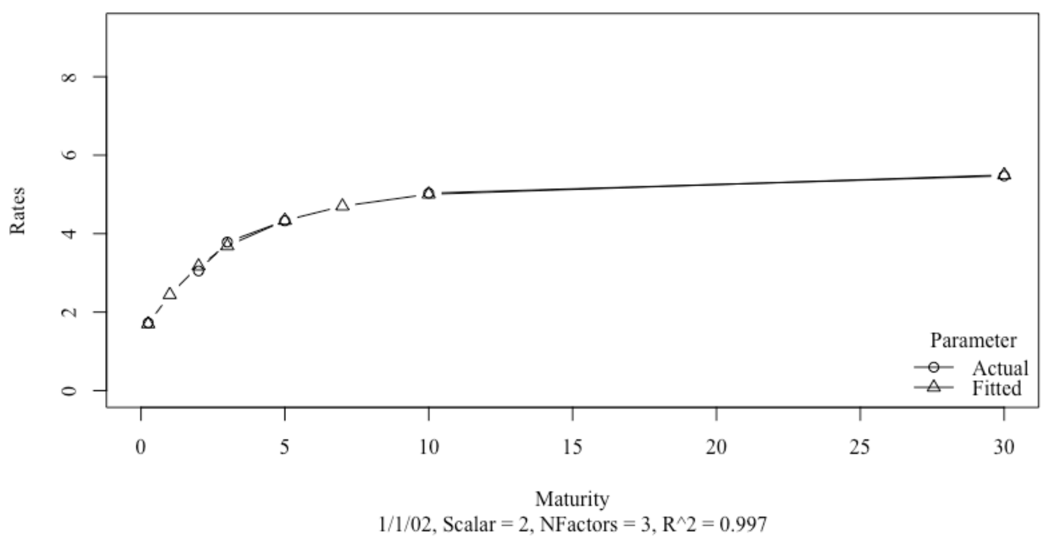
Level	3.50618%		
Slope	-2.95534%		
Curvature	-0.35538%		
Scalar	2.00		
	Regression Variables		
Maturity	**y**	**x1**	**x2**
0.00001	na	1.00000	0.00000
1	1.16%	0.78694	0.18041
2	1.48%	0.63212	0.26424
3	1.84%	0.51791	0.29478
4	2.14%	0.43233	0.29700
5	2.33%	0.36717	0.28508
7	2.62%	0.27709	0.24689
10	2.89%	0.19865	0.19191
30	3.25%	0.06667	0.06667



In the LSC-related data above, Maturity provides the τ_i, y denotes the observed swap rates, $x_{i,0}$ (not shown) is the intercept terms, x1 ($x_{i,1}$) denotes the first vector of input LSC coefficients, and $x2$ ($x_{i,2}$) denotes the second vector of input LSC coefficients. The resulting LSC parameters are f_0 = 3.50618% (Level), f_1 = –2.95534% (Slope), and f_2 = –0.35538% (Curvature). Thus, the LSC curve fitting module is very flexible and can be solved using ordinary least squares regression.

One measure of how well the LSC model fit the data is r-square (denoted R^2 or R^2) or correlation coefficient squared. Recall R^2 measures the percentage of variability in the regression dependent variable (interest rates here) and the regression independent variables (functions of maturity and scalars) from the linear model. Figure 3.4.8 illustrates typical U.S. Treasury CMT data and an LSC three factor model with the scalar set to equal 2. Note the R^2 is nearly equal to 1 (0.997) indicating a good fit. Thus, with only three factors, Level, Slope, and Curvature1, we can estimate the infinite number of potential rates from overnight to perpetual.

Figure 3.4.8. U. S. swap rates with fitted LSC model in R



One weakness of the LSC model that is encountered is when the financial data lacks cross-sectional variability. Figure 3.4.9 illustrates U.S. Treasury CMT rates on January 1, 2019 when the CMT rates were all nearly the same. The R^2 is only 0.792 yet the fit is very good for most applications in quantitative finance.

Figure 3.4.9. U. S. swap rates with fitted LSC model when data lacks variability

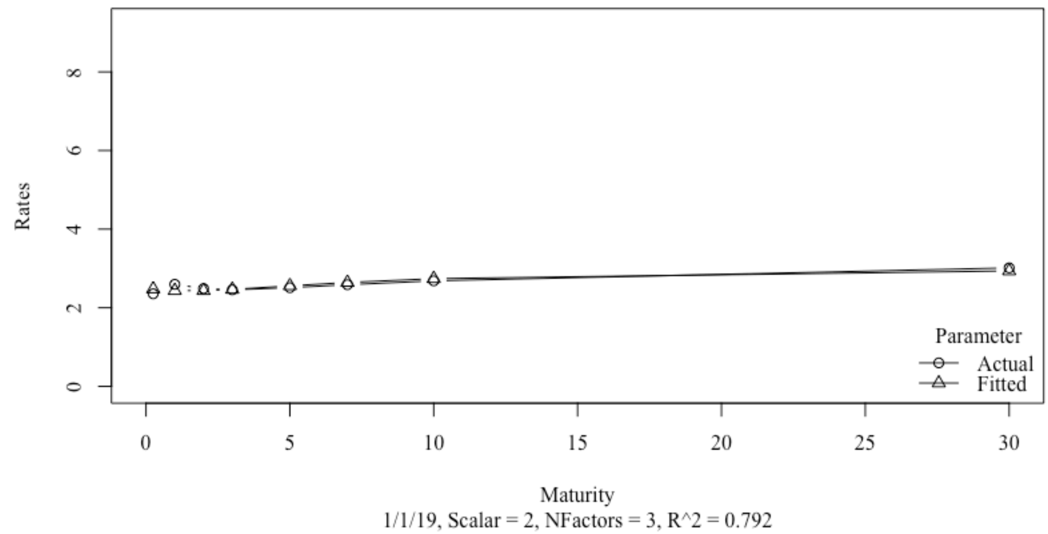
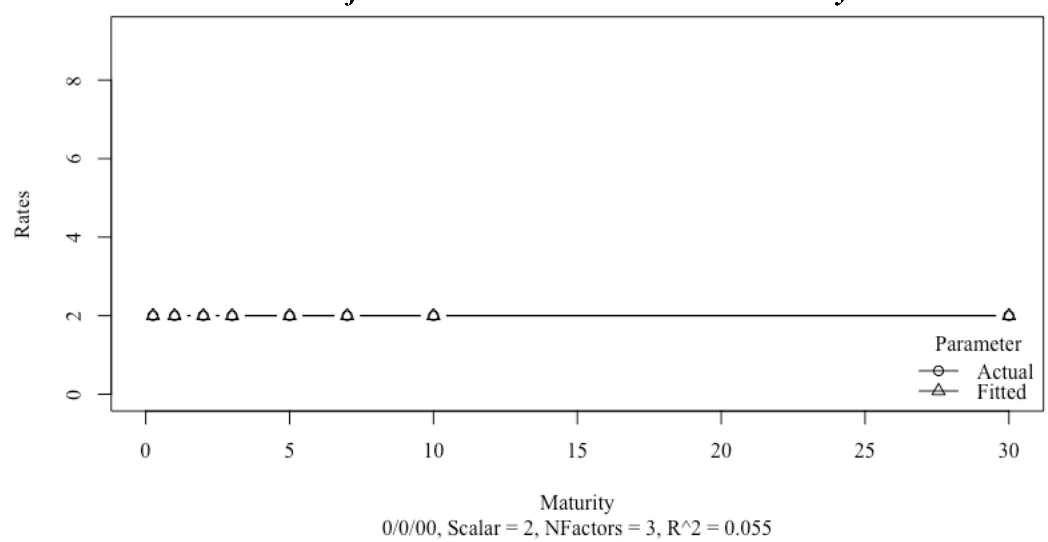


Figure 3.4.10 illustrates fabricated data that oscillates up and down by 0.1 basis points. In this case, the R^2 is nearly zero but yet again the fit is very close. Thus, visual inspection is important when applying the LSC model to actual financial data.

Figure 3.4.10. Fabricated rates with fitted LSC model with minimal variability



The LSC model is flexible enough to have as many factors as an analyst wants. There are clearly tradeoffs. More factors will require more non-linear scalars. Our experience is that the more factors used, the more instability exhibited in the LSC model coefficients. Further, if you choose to solve for the scalars, the LSC model coefficients also exhibit more instability. Some analysts will first fit a nonlinear model to understand the implied scalars and then use the average of these scalars in their model. In most applications, the end results indicated in the study are not influenced by the choice of scalars but are often heavily influenced by the number of factors. A single factor model with just level is typically inadequate. Further, more than three factors (level, slope, and curvature1) is typically unnecessary and results in parameter instability over time. We explore these issues further.

LSC model independent variables and scalars
When deploying the LSC model, used widely in this material, you will face many choices. First, you must decide how many factors to choose and what scalars to apply. Figure 3.4.11 illustrates up to five regression independent variables. The regression independent variables essentially apply different weighting to different maturities. Level applies constant weight. Slope applies more weight to the lower maturities. Curvature1 applies the most weight to the near term but is humped. Note that the peak of Curvature1 is not 2. Curvature2 and Curvature3 illustrate how the curvatures are influenced by the scalar choice.

Figure 3.4.11. LSC model regression independent variables, $x_{i,j}$

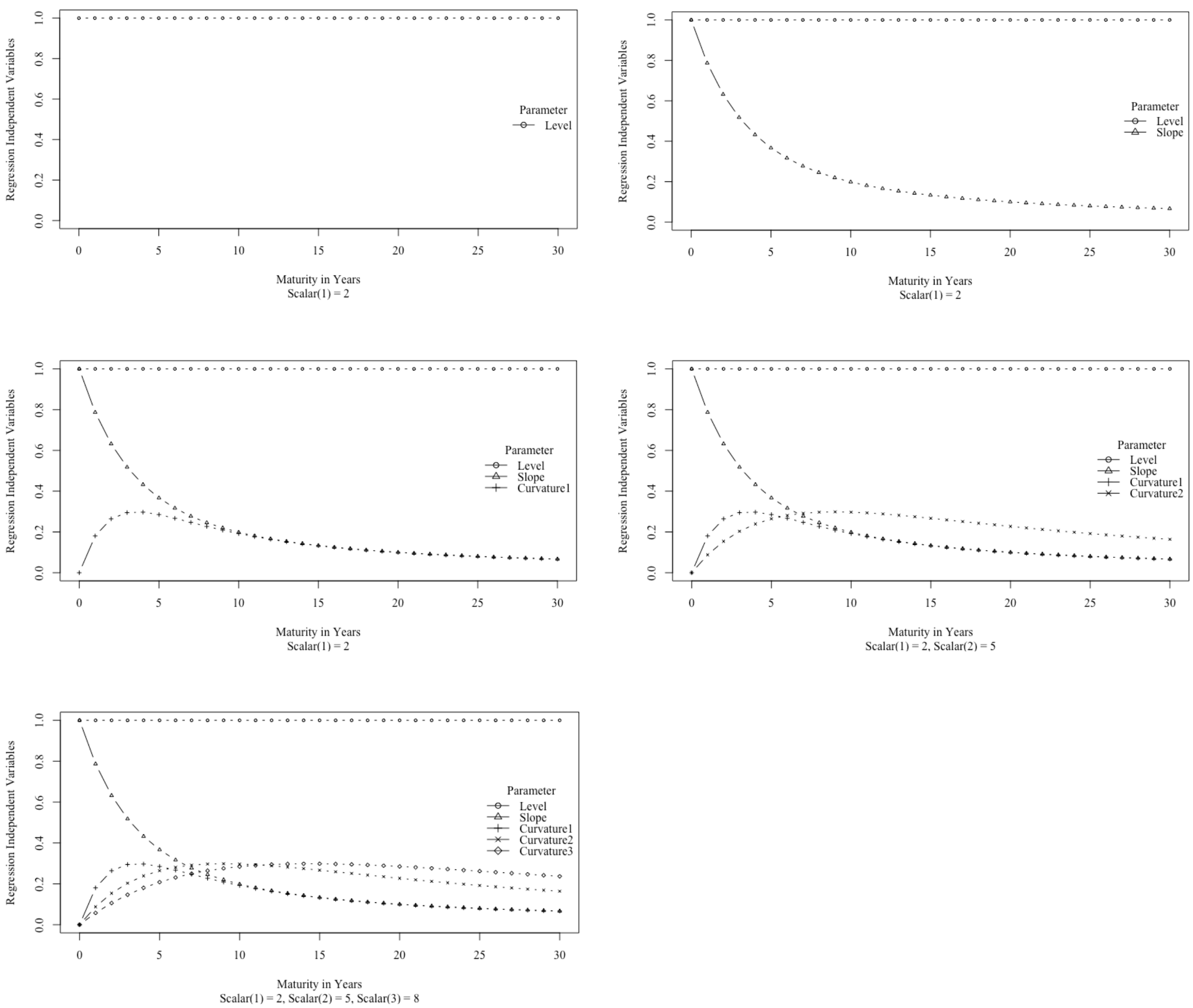


Figure 3.4.12 focuses on a three factor model but applying different scalars. Notice that scalar1 influences both slope and curvature1 regression independent variables. Many academic writers as well as practitioners have found setting scalar1 to 2.0 to be adequate and at times optimal.

Figure 3.4.12. Three factor LSC model regression independent variables, $x_{i,j}$ with different scalars

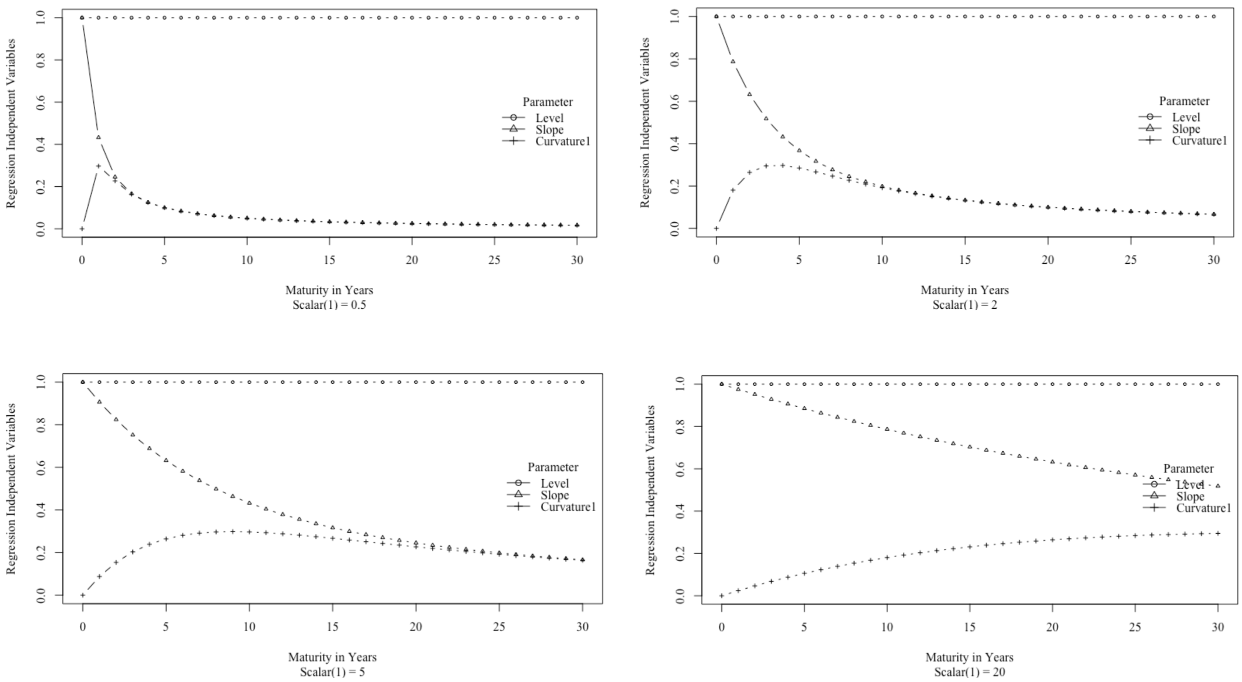


Figure 3.4.13 illustrates different slope coefficients within a three factor LSC model, where level is set to five percent and curvature1 is set to zero. Notice that a negative slope results in an upward sloping function and a positive slope results in a downward sloping function. Although arbitrary in how it is represented, this is the form that was originally taken when this model was being developed.

Figure 3.4.13. Three factor LSC model with different slope coefficients

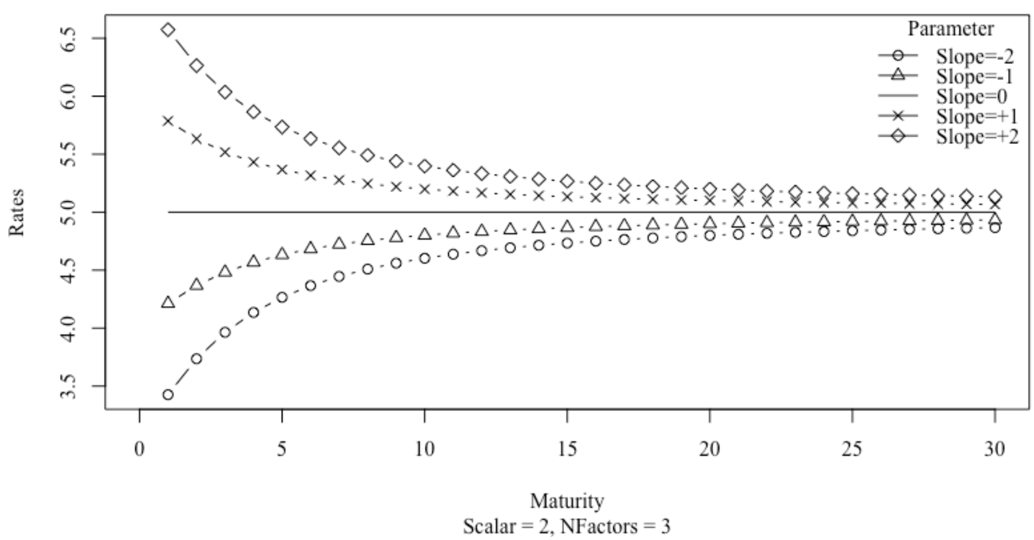


Figure 3.4.14 illustrates different curvature1 coefficients within a three factor LSC model, where level is set to five percent and slope is set to zero. Notice that a negative curvature1 results in an initially downward

sloping function that turns upward and a positive curvature1 results in an initially upward sloping function that turns downward. Again, remember that the LSC model is fit using linear regression, but the resulting functions are quite flexible in its non-linear shape.

Figure 3.4.14. Three factor LSC model with different coefficient1 coefficients (Slope = 0)

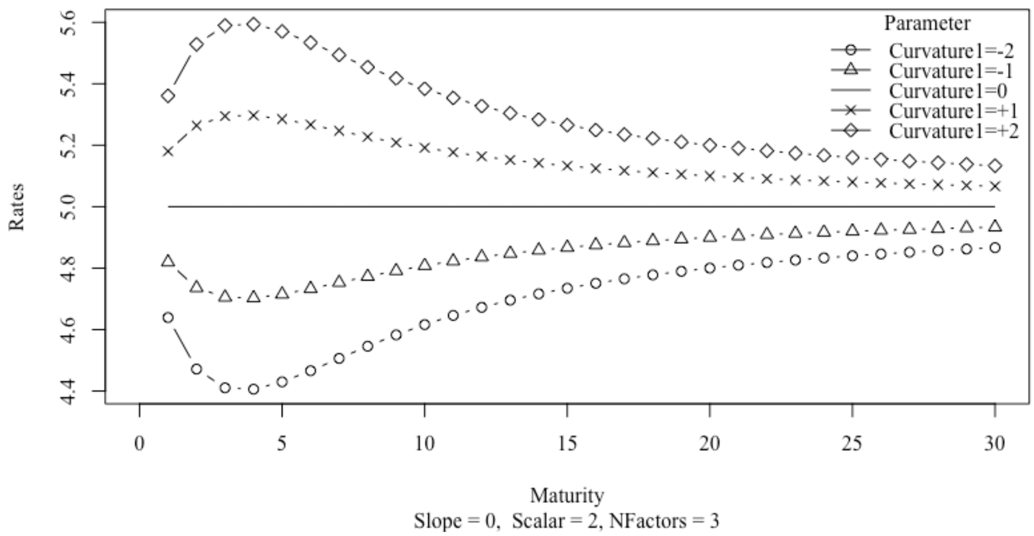
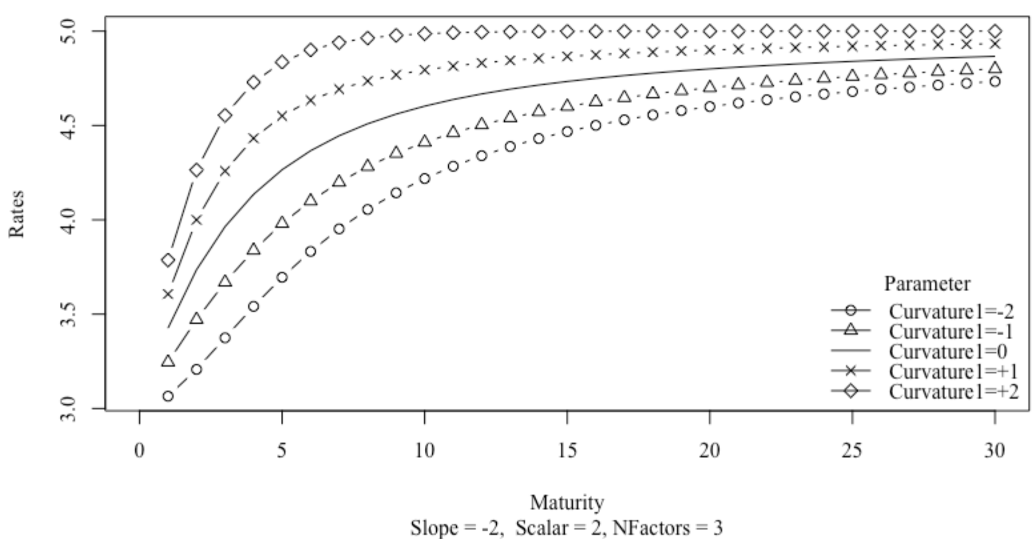


Figure 3.4.15 illustrates different curvature1 coefficients within a three factor LSC model, where level is set to five percent and slope is set to –2 (upward sloping). Notice that a negative curvature1 results in a lower initial rate and is slower to converge to the level coefficient whereas a positive curvature1 results in a higher initial rate and is faster to converge to the level coefficient.

Figure 3.4.15. Three factor LSC model with different coefficient1 coefficients (Slope = –2)



The LSC model has been adopted in a wide array of practical finance applications, including fitting various term structures of interest rates, various term structures of volatilities, patters of first differences in maturity-varying futures contracts, as well as percentage changes in maturity-varying futures contracts.

Appendix 3.4A. Ordinary least squares regression review

A standard statistical problem is to find the best solution to a set of linear equations of the form

$$\begin{aligned} &x_{11}\beta_1 + x_{12}\beta_2 + x_{13}\beta_3 + \cdots + x_{1n}\beta_n = y_1 \\ &x_{21}\beta_1 + x_{22}\beta_2 + x_{23}\beta_3 + \cdots + x_{2n}\beta_n = y_2 \\ &\vdots \\ &x_{m1}\beta_1 + x_{m2}\beta_2 + x_{m3}\beta_3 + \cdots + x_{mn}\beta_n = y_m \end{aligned}, \tag{3.7}$$

or in matrix notation

$$\underset{mxn}{\mathbf{X}}\,\underset{nx1}{\mathbf{b}} = \underset{mx1}{\mathbf{Y}}, \tag{3.8}$$

where

$$\underset{mxn}{\mathbf{X}} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}, \ \underset{nx1}{\mathbf{b}} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \text{and } \underset{mx1}{\mathbf{Y}} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}. \tag{3.9}$$

The **X** matrix is known as the independent variables and the **Y** vector is known as the dependent variables. The **b** vector of unknown parameters can be found using ordinary least squares regression.

We assume here that $m > n$ and the set of linear equations is said to be over-determined. Thus, we seek the best fit by solving for the **b** vector.

The normal equations can be written in matrix form as

$$\begin{aligned} &\underset{nxm}{\mathbf{X}^T}\,\underset{mxn}{\mathbf{X}}\,\underset{nx1}{\mathbf{b}} = \underset{nxm}{\mathbf{X}^T}\,\underset{mx1}{\mathbf{Y}} \\ &\underset{nxn}{\mathbf{NX}}\,\underset{nx1}{\mathbf{b}} = \underset{nx1}{\mathbf{NY}} \end{aligned}, \tag{3.10}$$

and the OLS solution to this set of linear equations can be express as

$$\begin{aligned} &\underset{nx1}{\hat{\mathbf{b}}} = \left(\underset{nxm}{\mathbf{X}^T}\,\underset{mxn}{\mathbf{X}}\right)^{-1}\underset{nxm}{\mathbf{X}^T}\,\underset{mx1}{\mathbf{Y}} \\ &\underset{nx1}{\hat{\mathbf{b}}} = \left(\underset{nxn}{\mathbf{NX}}\right)^{-1}\underset{nx1}{\mathbf{NY}} \end{aligned}, \tag{3.11}$$

where $\underset{nxn}{\mathbf{NX}}$ denotes the normalized X matrix and $\underset{nx1}{\mathbf{NY}}$ denotes the normalized Y matrix.

According to Press, et. al. (1992), the LU decomposition approach is very efficient for finding the solution to this set of linear equations. The idea is to decompose the **NX** matrix into lower and upper triangular matrices, where the lower triangular matrix has elements only on the diagonal and below and the upper triangular matrix has elements only on the diagonal and above.

$$
\underset{nxn}{\mathbf{NX}} = \begin{bmatrix} nx_{11} & nx_{12} & \cdots & nx_{1n} \\ nx_{21} & nx_{22} & \cdots & nx_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ nx_{n1} & nx_{n2} & \cdots & nx_{nn} \end{bmatrix}
$$

$$
= \underset{nxn}{\mathbf{L}}\,\underset{nxn}{\mathbf{U}} = \begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n1} & \lambda_{n2} & \cdots & \lambda_{nn} \end{bmatrix} \begin{bmatrix} \upsilon_{11} & \upsilon_{12} & \cdots & \upsilon_{1n} \\ 0 & \upsilon_{22} & \cdots & \upsilon_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \upsilon_{nn} \end{bmatrix}. \tag{3.12}
$$

The basic idea is to solve sequentially for the unknown parameters.

$$
\begin{gathered} \underset{nxn}{\mathbf{NX}}\,\underset{nx1}{\mathbf{b}} = \underset{nx1}{\mathbf{NY}} \\ \underset{nxn}{\mathbf{L}}\,\underset{nxn}{\mathbf{U}}\,\underset{nx1}{\mathbf{b}} = \underset{nx1}{\mathbf{NY}} \end{gathered} \tag{3.13}
$$

That is, we first solve for the vector **Z** such that

$$
\underset{nxn}{\mathbf{L}}\,\underset{nx1}{\mathbf{Z}} = \underset{nx1}{\mathbf{NY}}. \tag{3.14}
$$

Because **L** is a lower triangular matrix, the solution is straightforward based on forward substitution. Based on this result, we can find

$$
\underset{nxn}{\mathbf{U}}\,\underset{nx1}{\mathbf{b}} = \underset{nx1}{\mathbf{Z}}. \tag{3.15}
$$

Because **U** is an upper triangular matrix, the solution is straightforward based on backward substitution. Thus, we solved for the unknown parameters without having to compute the matrix inverse or relying on numerous pivoting routines. Again, for more extensive discussion of this procedure, see Press, et. al. (1992).

Appendix 3.4B. Details of the LSC Model

The LSC model can be used to estimate a wide variety of interest rates related to the term structure. The goal of the estimation exercise is to approximate some rate across a wide array of maturities, typically with only a handful of observed market prices. The set of discount factors, $d\left(\tau_j : t_i\right)$, is defined based on either spot rates, $r\left(\tau_j : t_i\right)$, or instantaneous forward rates, $f\left(\tau_j : t_i\right)$, as

$$
d\left(\tau_j : t_i\right) \equiv e^{-r\left(\tau_j : t_i\right)\tau_j} \tag{3.16}
$$

or

$$
d\left(\tau_j : t_i\right) \equiv e^{-\int_0^{\tau_j} f\left(u, : t_i\right) du}. \tag{3.17}
$$

Thus, the instantaneous forward rate can be expressed as

$$
f\left(\tau_j : t_i\right) = \frac{\partial}{\partial \tau_j}\left[-\ln d\left(\tau_j : t_i\right)\right]. \tag{3.18}
$$

Substituting from Equation (3.16),

$$
f\left(\tau_j : t_i\right) = \frac{\partial}{\partial \tau_j}\left[r\left(\tau_j : t_i\right)\tau_j\right] = \frac{\partial r\left(\tau_j : t_i\right)}{\partial \tau_j}\tau_j + r\left(\tau_j : t_i\right). \tag{3.19}
$$

Hence, the instantaneous forward rate is the current spot rate plus any marginal change in the spot rate curve. The expanded version of Svensson (1995) and Nelson and Siegel (1987) can be expressed in the following theorem.

Theorem A1. If spot rates are expressed in terms of forward rates as

$$r\left(\tau_j : t_i\right)\tau_j = \int_0^{\tau_j} f\left(u,: t_i\right)du, \tag{3.20}$$

and estimated forward rates can be expressed as

$$f\left(\tau_j : t_i\right) = a_0 + a_1 e^{-\tau_j/s_1} + \sum_{i=2}^{n} a_i \left[\frac{\tau_j}{s_{i-1}} e^{-\tau_j/s_{i-1}}\right]. \tag{3.21}$$

then estimated spot rates are expressed as

$$r\left(\tau_j : t_i\right) = a_0 + a_1 \left[\frac{1-e^{-\tau_j/s_1}}{\tau_j/s_1}\right] + \sum_{i=2}^{n} a_i \left[\frac{1-e^{-\tau_j/s_{i-1}}}{\tau_j/s_{i-1}} - e^{-\tau_j/s_{i-1}}\right]. \tag{3.22}$$

Proof: The proof is based on elementary integration properties and follows directly from the following two lemmas.

Lemma 1. Assuming c>0, then

$$\int_0^{\tau} e^{-u/c}du = \frac{1-e^{-\tau/c}}{1/c}. \tag{3.23}$$

Proof of Lemma 1:

$$\int_0^{\tau} e^{-u/c}du = \left.\frac{e^{-u/c}}{-1/c}\right|_0^{\tau} = \frac{e^{-\tau/c}}{-1/c} - \frac{e^{-0/c}}{-1/c} = \frac{1-e^{-\tau/c}}{1/c}.\text{ QED.} \tag{3.24}$$

QED

Lemma 2. Assuming c>0, then

$$\int_0^{\tau} \frac{u}{c} e^{-u/c}du = \frac{1-e^{-\tau/c}}{1/c} - \tau e^{-\tau/c}.$$

Proof of Lemma 2: Let

$$F\left(u\right) = \frac{u}{c} \text{ and } G'\left(u\right) = e^{-u/c}.$$

Thus,

$$F'\left(u\right) = \frac{1}{c} \text{ and } G\left(u\right) = -\frac{e^{-u/c}}{1/c}.$$

Therefore, based on integration by parts

$$\int_0^{\tau} F\left(u\right)G'\left(u\right)du = F\left(u\right)G\left(u\right)\Big|_0^{\tau} - \int_0^{\tau} F'\left(u\right)G\left(u\right)du \text{ and}$$

$$\int_0^{\tau} \frac{u}{c} e^{-u/c}du = \left.\frac{u}{c}\frac{e^{-u/c}}{-1/c}\right|_0^{\tau} - \int_0^{\tau}\frac{1}{c}\frac{e^{-u/c}}{-1/c}du = -\tau e^{-\tau/c} + \int_0^{\tau} e^{-u/c}du.$$

Using Lemma 1 and rearranging. QED.

Module 3.5. Sorting Data

Learning Objectives

- Explain how to sort data very efficiently
- Review the manner in which data files can be read (see R commentary)
- Learn the variety of financial uses of sorting routines

Module Overview

The ability to sort data is very important for a wide array of financial tasks. We first briefly review some of the financial applications and then show some related output.

Financial applications of sorting

There are several financial tools that require numerical data to be sorted. For example, when computing value-at-risk using Monte Carlo simulation or historical simulation, the numerical data is sorted and then the value-at-risk measure is computed.

Investment managers often will produce histograms of a portfolio's historical rates of return. To produce these histograms, the data must first be sorted. When exploring the use of leverage, a variety of portfolios are constructed, and various distributions can be examined. Each of these distributions would first be sorted. Several modules related to dynamic risk management will require data to be sorted.

Sorting illustrations

The R code illustrates managing data related to stock prices. Figure 3.5.1 illustrates the first differences distribution overlaid with an estimate of the normal distribution.

Figure 3.5.1. Illustration of first differences distribution

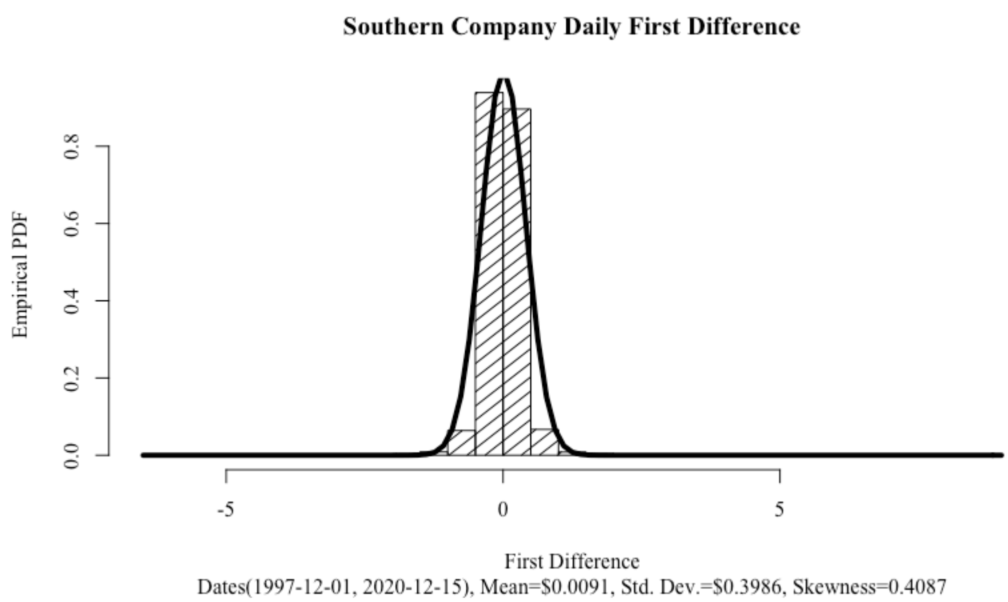


Figure 3.5.2 illustrates the percentage rates of return distribution overlaid with an estimate of the normal distribution.

Figure 3.5.2. Illustration of percentage rate of returns

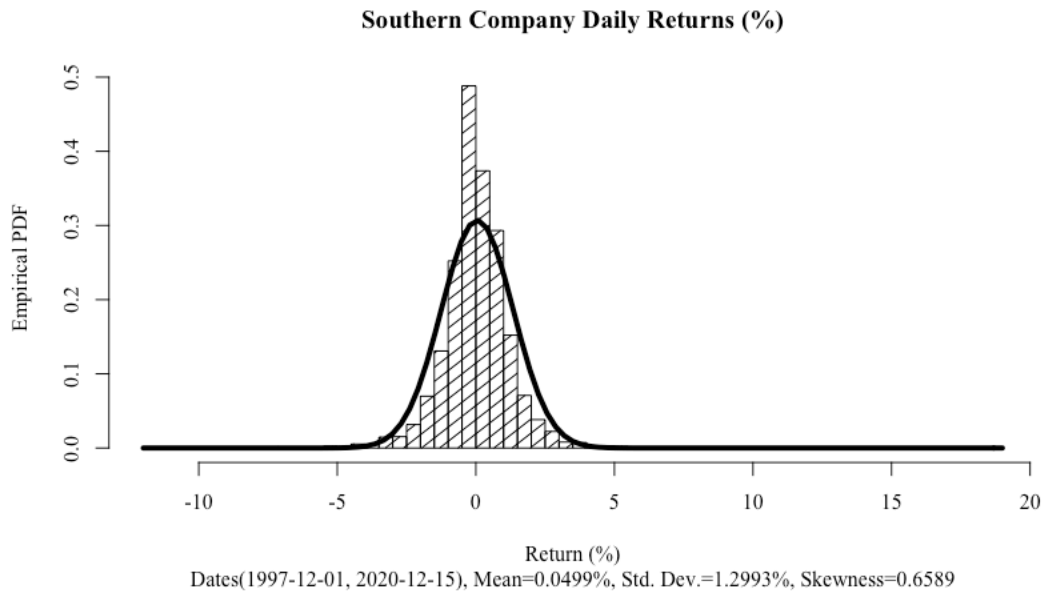
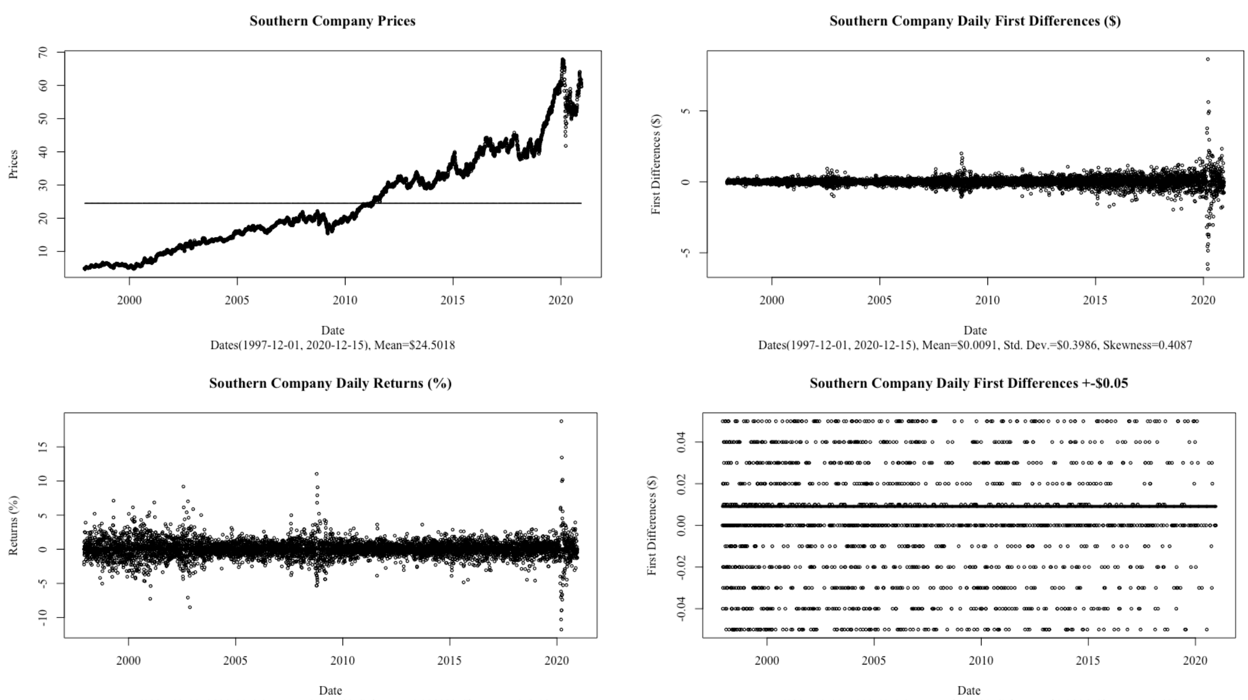
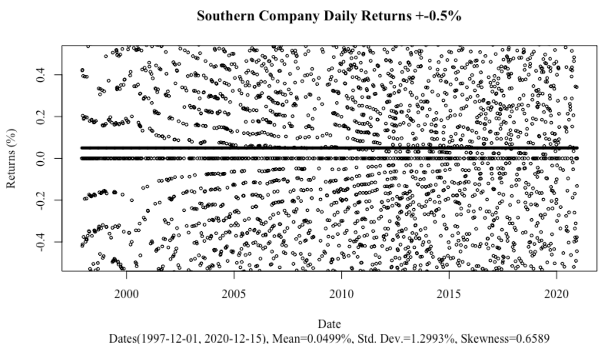


Figure 3.5.3 illustrates the several ways to represent financial data including prices, first differences, and percentage daily returns. Note that $0.01 trading is clear with first differences and the prevalence of holidays (too many zeros) is also apparent.

Figure 3.5.3. Selected illustrations contrasting first differences and rates of return





There is an enormous number of financial insights that are now within your grasp with the combined power of R coding and financial analysis. For example, the assumption that financial distributions are stable is clearly not true for first differences—grows over time with underlying stock price and financial crisis effects. Stability is also clearly not true for rates of return—declines over time and financial crisis effects. Finally, decimalization influences first differences and percentage returns differently.

Module 3.6. Embedded Parameters

Learning Objectives

- Explain how to solve for embedded parameters
- Develop the capacity to solve for any embedded parameter
- Learn how compute the implied yield to maturity for a simple, fixed rate bond

Module Overview

The ability to solve embedded parameters is introduced here illustrated with the problem of solving for the appropriate yield to maturity given the current market price of a bond.

Bond pricing and yield to maturity

The simplest way to express the current value of a bond (V) given a fixed dollar coupon (C) with a given face value (Par) and a stated time to maturity expressed in years (t_N) as well as assumed yield to maturity (y) is

$$V = \sum_{i=1}^{N} \frac{C}{(1+y)^{t_i}} + \frac{Par}{(1+y)^{t_N}}. \tag{3.25}$$

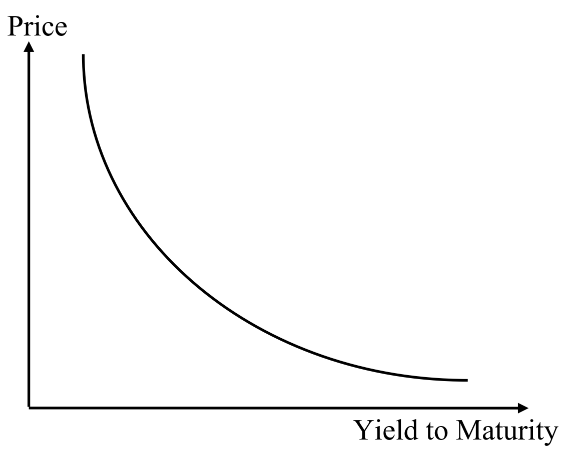
Our focus here is on solving for the yield to maturity, an embedded parameter. This task is accomplished by expressing a function of yield to maturity such that when the correct embedded parameter is used, then this function returns zero. We define the current market price of the bond as P. That is,

$$f(y) = \sum_{i=1}^{N} \frac{C}{(1+y)^{t_i}} + \frac{Par}{(1+y)^{t_N}} - P \tag{3.26}$$

The goal of programs like the one presented here is to solve one-dimensional problems like this one numerically. The yield to maturity that makes $f(y) = 0$ is known as the root of this function.

Recall as yield goes up, bond price goes down; as yield goes down, bond price goes up. The price-yield curve can be described as convex. It means that the curve bows away from the origin of the graph as illustrated in Figure 3.6.1. (Bowed toward the origin would be called concave.) The idea is when we are given a bond price, we can compute the implied yield to maturity. That is, what yield to maturity will return a bond price equal (within some allowable error) to the observed market price of the bond.

Figure 3.6.1. Relationship between bond prices and yield to maturity



Numerical methods

There are no known methods to derive an exact equation for the yield to maturity problem above as well as numerous other embedded functions in finance, such as option model implied volatility. Almost all functions in finance are continuous, thus making root finding easier. There are innumerable methods, however, to solve for embedded parameters. Methods include bisection, false position, secant, Ridders, Brent, and Newton-Rhapson. Source code in C++ is available for each of these methods in *Numerical Recipes in C++*. Source code in R is widely available in different packages. We are using the optimize function within the stats package.

Because the Brent method is built on concepts from other methods, we briefly introduce some root finding concepts and methods. The optimization routine used in the R stats package is a "slight variation" of the Brent method.

First, at least one solution to an embedded parameter problem is said to be bracketed if in the interval (a,b), the function $f(a)$ and $f(b)$ return opposite signs. Remember that the function is configured such that $f(x) = 0$ when x is the correct solution to the implied parameter problem. Simply move through the range (a,b) at sufficiently "small" increments. You have an approximation to the solution when the sign changes.

Second, the method of bisection is very simple and is fail proof. The idea is to take the extremes of the bracket, find the midpoint, and then determine where the opposite signs still occur. That is, consider $f(a)$ and $f(m)$ compared to $f(m)$ and $f(b)$, where $m = (a + b)/2$. Repeat this process, until the difference between the brackets is within the allowable error amount.

Third, the secant method assumes that the slope of the function is approximately linear and, based on this slope information, often converges faster than the method of bisection. The secant method requires an initial guess.

Fourth, Newton-Rhapson is the best choice when the first derivative of the given function is known. Rather than numerically computing the slope of the line, the information contained in the first derivatives is used. We have found, however, that Brent is extremely reliable and very fast. Therefore, we use it even when computing the first derivative is possible.

Fifth, the concept of inverse quadratic interpolation is exploited in the Brent method. Inverse quadratic interpolation relies on three points to fit an approximation function. Thus, it is not linear. This approach assumes the implied parameter (x) is roughly a quadratic function of the known parameter (y). Therefore, the inverse quadratic interpolation is more efficient than simple linear interpolation.

According to Press, et. al. (1992), "Brent's method combines root bracketing, bisection, and *inverse quadratic interpolation* to converge from the neighborhood of a zero crossing. ... Brent's method combines the sureness of bisection with the speed of a higher-order method when appropriate. We recommend it as the method of choice for general one-dimensional root finding where a function's values only (and not its derivative or functional form) are available." (p. 360-361)

The Brent method requires the user to bracket the root(s) prior to starting the analysis. Thus, we select a very high and a very low yield to maturity to assure that $f(y{=}500\%) < 0$ and $f(y{=}0.001\%) > 0$. The programmer rather than the end user determines these bounds in our case. Another decision made by the programmer that must be made is the numerical accuracy where the search will cease. We code an epsilon of 0.000001, thus if the absolute value of $f(y)$ is less than 0.000001, then the Brent method will stop searching. Brent is a combination of the best aspects of bisection, secant, and inverse quadratic interpolation.

Although not directly related to embedded functions, Figure 3.6.2 illustrates the relationship between bond prices and yield to maturity for different maturities.

Figure 3.6.2. Relationship between bond prices and yield to maturity for different maturities

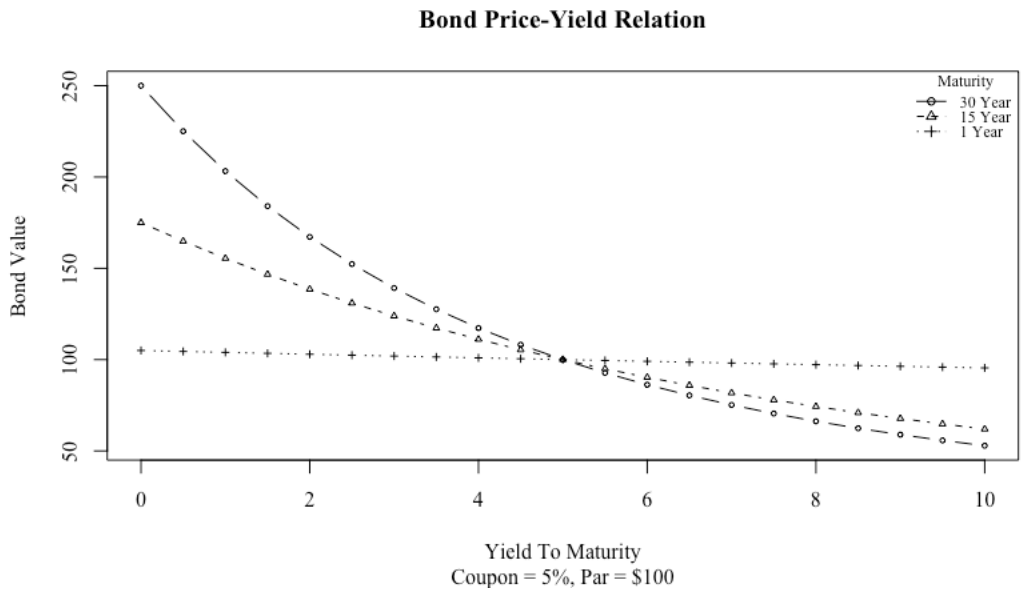
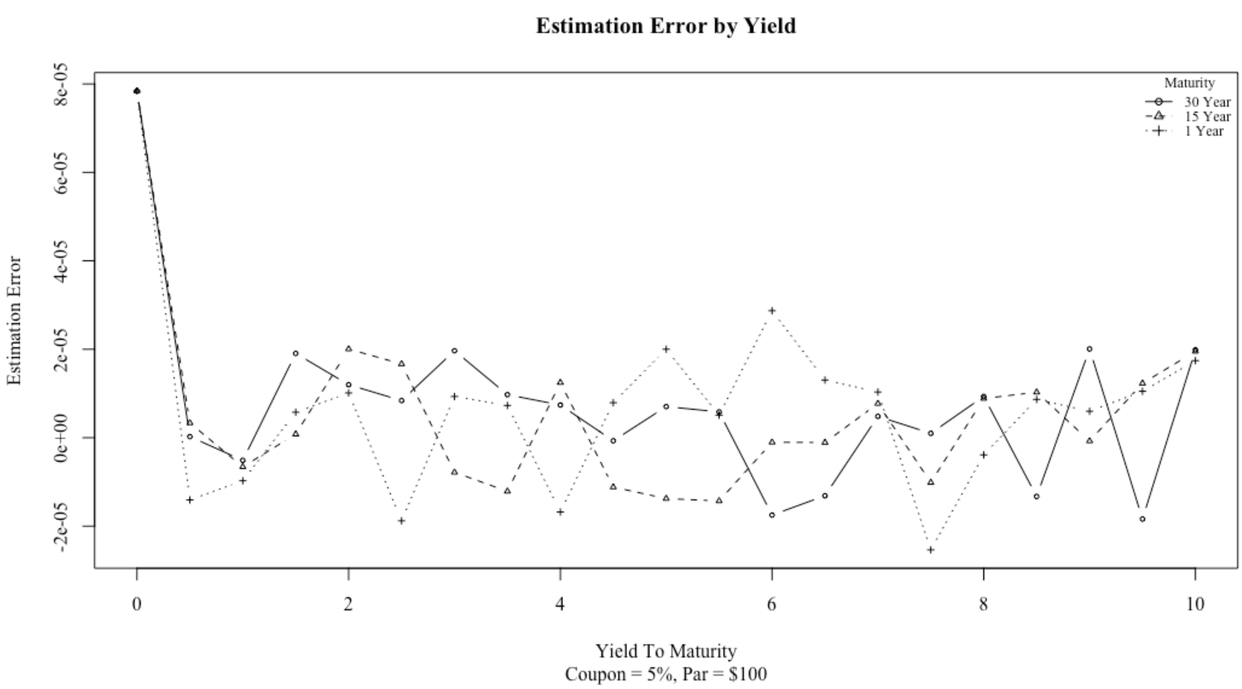


Figure 3.6.3 illustrates the relationship between yield to maturity estimation error and yield to maturity for different maturities. Clearly, there is minimal estimation error.

Figure 3.6.3. Relationship between yield to maturity estimation error and yield to maturity for different maturities



Module 3.7: Numerical Integration and the Lognormal Distribution

Learning objectives

- Explore the use of numerical integration in finance
- Compute the probability of an option being in-the-money based on the lognormal distribution
- Contrast the normal and lognormal distribution
- Understand when assuming the lognormal distribution is inappropriate

Module overview

This module illustrates how to estimate via integration the probability of a call and put option being in-the-money assuming the underlying instrument's terminal value is lognormally distributed. We also provide several summary statistics for both the lognormal and normal distribution. The parameters of this distribution are based on option-related information, and we assume the underlying instrument is expected to grow at the dividend-adjusted risk-free interest rate.[9]

We review the basic properties of the lognormal distribution with a focus on the behavior of the lognormal distribution with very high dispersion. It is common for option prices to imply unreasonable volatilities when we assume a lognormal distribution. One objective here is to identify when the lognormal distribution will likely need to be replaced with an alternative probability distribution.

The lognormal distribution holds a central role in finance. For example, the underlying instrument of financial derivatives is often assumed to follow a lognormal distribution. This distribution is attractive because the underlying instrument cannot be negative due to limited liability and the lognormal distribution limits are zero (not inclusive) and positive infinity (not inclusive).

The following sections explore in detail the properties of the lognormal distribution. Because the lognormal distribution is widely used in finance it is vital to understand its properties. Some of the normal distribution properties are also provided for comparison.

Univariate Normal and Lognormal Distribution

The cumulative distribution function (CDF) of a variable X is defined as

$$F_X(x) \equiv \Pr(X \le x). \tag{3.27}$$

With continuous variables, such as the normal and lognormal distributions, the CDF is related to the probability density function (PDF) as

$$F_X(x) = \int_{-\infty}^{x} f_X(j)\,dj, \tag{3.28}$$

where $f_X(j)$ denotes the corresponding PDF.

The lognormal distribution is directly related to the normal distribution. The lognormal distribution has two parameters, the mean, μ, and the standard deviation, σ. At this point, we are using the symbols for mean and standard deviation generically. Later, we will use these same symbols for very specific finance applications.

The mean must be finite, $-\infty < \mu < +\infty$, and the standard deviation must be positive, $\sigma > 0$. The range of the lognormal distribution is the positive real number line or $0 < x < +\infty$.[10] For many finance applications, x would be related to the underlying instrument, such as a particular stock price. Interestingly, the lognormal distribution does not admit the possibility of $x = 0$. Thus, one weakness of the lognormal distribution being used to model a stock price is that the company can never go bankrupt in such a way that the existing stock price is worthless. The financial marketplace is littered with a vast number of worthless common stock. Recall that no mathematical model of reality is correct, but many models are useful. The skilled financial analyst knows the model's limitations. The purpose of analytical tools, such as the lognormal distribution, is

[9]This assumption is not required it just makes the transition to the Black, Scholes, Merton option valuation model easier.

[10]It is unclear whether the range of the lognormal is actually $0 < x \le +\infty$. That is, is positive infinity included or not? I believe not.

to roughly approximate reality. The quantitative professional is always keen to know when these tools will fail to be useful.

If $y = \ln(x)$ is distributed normal, then *x* is said to have a lognormal distribution. Given this functional relationship, we see that the two distributions are based on the same two underlying parameters. The CDFs are defined as (Λ is the Greek upper case lambda)

$$\Lambda(\mu,\sigma) = \Lambda(d) = \int_{-\infty}^{d} \frac{e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}}}{x\sigma\sqrt{2\pi}} dx \text{ (Lognormal CDF) and} \quad (3.29)$$

$$N(\mu,\sigma) = N(d) = \int_{-\infty}^{d} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} dx \text{ (Normal CDF).} \quad (3.30)$$

The PDFs therefore are (λ is the Greek lower case lambda)

$$\lambda(\mu,\sigma) = \lambda(x) = \frac{e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}}}{x\sigma\sqrt{2\pi}} \text{ (Lognormal PDF) and} \quad (3.31)$$

$$n(\mu,\sigma) = n(x) == \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}} \text{ (Normal PDF).} \quad (3.32)$$

Thus, we distinguish between the PDFs and CDFs with lower and upper case and between the lognormal and normal with lambda (λ) and *n*.

From its definition, the median is

$$\hat{x} \text{ such that } \int_{0}^{\hat{x}} f(x)dx = \int_{\hat{x}}^{+\infty} f(x)dx = \frac{1}{2}. \quad (3.33)$$

The medians are

$$Median_{\lambda} = e^{\mu} \text{ and} \quad (3.34)$$

$$Median_{n} = \mu. \quad (3.35)$$

Notice that the lognormal distribution median is invariant to the standard deviation of the normal distribution.

The mode satisfies the following two properties, $f'(x) = 0$ and $f''(x) < 0$. For the lognormal and normal distributions, the modes are

$$Mode_{\lambda} = e^{\mu-\sigma^2} \text{ and} \quad (3.36)$$

$$Mode_{n} = \mu. \quad (3.37)$$

The lognormal distribution mode is an exponentially decreasing function of the normal distribution variance. This is an important property of the lognormal distribution as it applies to financial applications. As the variance of the normal distribution increases, the peak (mode) of the lognormal distribution is decreasing.

The first four moments of the lognormal and normal distribution are as follows.

First moment about zero (mean) is

$$Mean_{\lambda} = e^{\mu+\frac{\sigma^2}{2}} \text{ and} \quad (3.38)$$

$$Mean_{n} = \mu. \quad (3.39)$$

The mean of the lognormal distribution is an exponentially increasing function of the normal distribution variance. Recall the lognormal distribution can be viewed as an exponential transformation of a normally distributed variable, say *x*. Thus, *x* is symmetric and e^x is asymmetric with positive skewness because higher values of *x* imply values of e^x are further from the mean than *x*. Thus, the lognormal mean is higher for higher normal distribution variance. Therefore, as the normal variance increases, the normal distribution mean has to decline if the goal is to maintain the same lognormal mean.

The next three moments are presented about the mean (and not zero). Second moment about the mean or variance can be expressed as:

$$\mu_{2\lambda} = Variance_{\lambda} = e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right) \text{ and} \tag{3.40}$$

$$\mu_{2n} = Variance_{n} = \sigma^2. \tag{3.41}$$

The variance is an exponentially increasing function of the normal distribution variance. By definition, the standard deviation of the normal distribution is σ. Third moment about the mean can be expressed as

$$\mu_{3\lambda} = e^{3\mu+\frac{3\sigma^2}{2}}\left(e^{\sigma^2}-1\right)^2\left(e^{\sigma^2}+2\right) \text{ and} \tag{3.42}$$

$$\mu_{3n} = 0. \tag{3.43}$$

The third moment is an exponentially increasing function of the normal distribution variance. The third moment of the normal distribution is zero. Fourth moment about the mean is

$$\mu_{4\lambda} = e^{4\mu+2\sigma^2}\left(e^{\sigma^2}-1\right)^2\left(e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-3\right) \text{ and} \tag{3.44}$$

$$\mu_{4n} = 3\sigma^4. \tag{3.45}$$

The fourth moment is an exponentially increasing function of the normal distribution variance. The fourth moment about the mean of the normal distribution is directly related to the normal distribution variance.

A unitless measure of skewness (the third standardized or normalized moment) is expressed as

$$\gamma_{S\lambda} = \frac{\mu_{3\lambda}}{\mu_{2\lambda}^{3/2}} = \sqrt{e^{\sigma^2}-1}\left(e^{\sigma^2}+2\right) \text{ and} \tag{3.46}$$

$$\gamma_{Sn} = \frac{\mu_{3n}}{\mu_{2n}^{3/2}} = 0. \tag{3.47}$$

Note that the normalized lognormal skewness is widely reported incorrectly.[11] To be thorough, substituting from Equations (3.40) and (3.41), we have

$$\begin{aligned}\gamma_{S\lambda} = \frac{\mu_{3\lambda}}{\mu_{2\lambda}^{3/2}} = \frac{e^{3\mu+\frac{3\sigma^2}{2}}\left(e^{\sigma^2}-1\right)^2\left(e^{\sigma^2}+2\right)}{\left[e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)\right]^{3/2}} = \left(e^{\sigma^2}-1\right)^2\left(e^{\sigma^2}-1\right)^{3/2}\left(e^{\sigma^2}+2\right). \\ = \sqrt{e^{\sigma^2}-1}\left(e^{\sigma^2}+2\right)\end{aligned} \tag{3.48}$$

The normalized lognormal skewness is an exponentially increasing function of the normal distribution variance. Symmetrical distributions, such as the normal distribution, will have $\gamma_{S\lambda} = 0$. If $\gamma_{S\lambda} > 0$, then we have the following relationship: mean > median > mode. If $\gamma_{S\lambda} < 0$, then we have the opposite relationship: mean < median < mode. There are, however, some exceptions. See Stuart and Ord (1987), p. 107.

The measure of excess kurtosis (fourth standardized moment minus 3) is

$$\gamma_{K\lambda} = \frac{\mu_{4\lambda}}{\mu_{2\lambda}^2} - 3 = e^{4\sigma^2}+2e^{3\sigma^2}+3e^{2\sigma^2}-6 \text{ and} \tag{3.49}$$

$$\gamma_{Kn} = \frac{\mu_{4n}}{\mu_{2n}^2} - 3 = 0. \tag{3.50}$$

[11]See, for example, Wikipedia. Hopefully, it is corrected by the time you view the lognormal distribution Wikipedia page.

This measure is termed "excess" because we subtract 3, the value of kurtosis for the normal distribution. The excess kurtosis is an exponentially increasing function of the normal distribution variance. Note that $\gamma_{K\lambda} = 0$ is called mesokurtic. The normal distribution and binomial distributions are mesokurtic. When $\gamma_{K\lambda} > 0$, the distribution is called leptokurtic having excess positive kurtosis. Leptokurtic distributions have fatter tails. The lognormal distribution is leptokurtic as well as the Laplace distribution and the logistic distribution. When $\gamma_{K\lambda} < 0$, the distribution is called platykurtic having negative excess kurtosis. Platykurtic distributions have thinner tails. The uniform distribution and Bernoulli distribution (p=1/2) are platykurtic.

Finally, differential entropy or continuous entropy is introduced. Differential entropy developed out of information theory and is supposed to measure the "… average surprisal of a random variable …," where surprisal denotes the surprise of seeing a particular outcome. (For more on differential entropy, see the information theory literature starting with Wikipedia.)

$$\text{Entropy of the lognormal distribution: } h_\lambda = \frac{1}{2}\left[1 + \ln\left(2\pi\sigma^2\right)\right] + \mu. \tag{3.51}$$

$$\text{Entropy of the normal distribution: } h_n = \frac{1}{2}\left[1 + \ln\left(2\pi\sigma^2\right)\right]. \tag{3.52}$$

For clarity, consider the following example related to some asset price distribution.

Asset price distribution example

Recall if $x \sim N\left(\mu_g, \sigma_g\right)$ (normal distribution, subscript g denote the generic mean and standard deviation – not finance specific) and $y = e^x$, then $y \sim \Lambda\left(\mu_g, \sigma_g\right)$ (lognormal distribution). In the context of rates of return (*R*), suppose $S_T = S_t e^{R(T-t)}$. In the material to follow, the mean and standard deviation notation uses the traditional finance form.

If a stock's continuously compounded rate of return is distributed normal $R \sim N\left[\mu\left(T-t\right), \sigma\sqrt{T-t}\right]$, then the terminal stock price is distributed lognormal $S_T \sim \Lambda\left[\ln\left(S_0\right) + \mu\left(T-t\right), \sigma\sqrt{T-t}\right]$. Thus, the terminal stock price can be expressed as $E\left(S_T\right) = S_0 e^{\left(\mu + \frac{\sigma^2}{2}\right)(T-t)}$ and the variance of the terminal stock price can be expressed as $Var\left(S_T\right) = S_0^2\left(e^{2\left(\mu+\sigma^2\right)(T-t)} - e^{\left(2\mu+\sigma^2\right)(T-t)}\right)$. Alternatively, the normal distribution parameters can be expressed as a function of the lognormal distribution parameters, $\mu = \ln\left[\frac{E\left(S_t\right)}{S_0}\right] \Big/ \left(T-t\right)$ and $\sigma^2 = \ln\left[\frac{Var\left(S_t\right)}{E\left(S_t\right)^2} + 1\right] \Big/ \left(T-t\right)$.

We now consider the application to stock returns. Suppose a stock is trading for \$100 and we have a one year horizon. If we say the annualized, continuously compounded expected rate of return on a stock is $\hat{\mu} = 12$ percent, what do we mean? Typically, we intend for the following equality to hold,

$$E\left(S_T\right) = S_0 e^{\hat{\mu}(T-t)} = 100e^{0.12(1)} = 112.749685.$$

Note the two different expressions for the terminal expected value of the stock are

$$E\left(S_T\right) = S_0 e^{\hat{\mu}(T-t)} \text{ and} \tag{3.53}$$

$$E\left(S_T\right) = S_0 e^{\left(\mu + \frac{\sigma^2}{2}\right)(T-t)}. \tag{3.54}$$

These two expressions are a source of much confusion. These two expressions are easily reconciled by setting

$$\mu = \hat{\mu} - \frac{\sigma^2}{2}. \tag{3.55}$$

For example, suppose we have S_0 = \$100, T – t = 1 year, $\hat{\mu}$ = 12%, and σ = 30%. Therefore,

$$E(S_T) = S_0 e^{\left(\mu + \frac{\sigma^2}{2}\right)(T-t)} = S_0 e^{\hat{\mu}(T-t)} = 100e^{0.12(1)} = 112.749685$$

and

$$\mu = \hat{\mu} - \frac{\sigma^2}{2} = 0.12 - \frac{0.3^2}{2} = 0.075.$$

Based on the notation above, we have

$$\begin{aligned} S_T &\sim \Lambda\left[\ln(S_0) + \left(\hat{\mu} - \frac{\sigma^2}{2}\right)(T-t), \sigma\sqrt{T-t}\right] \\ &= \Lambda\left[\ln(S_0) + \mu(T-t), \sigma\sqrt{T-t}\right] \\ &= \Lambda\left[\ln(100) + 0.075(1), 0.30\sqrt{1}\right] = \Lambda(4.680170, 0.30) \end{aligned}.$$

In this example, we can also compute several other statistics for this lognormal distribution:[12]

$$\begin{aligned} Median &= e^{\ln(S_0) + \left(\hat{\mu} - \frac{\sigma^2}{2}\right)(T-t)} = e^{4.680170} = 107.788415, \\ Mode &= e^{\ln(S_0) + \left(\hat{\mu} - \frac{\sigma^2}{2}\right)(T-t) - \sigma^2(T-t)} = e^{4.680170 - 0.09(1)}, \text{ and} \\ &= e^{4.680170 - 0.09(1)} = 98.511194 \\ Variance &= e^{2[4.680170] + 0.09^2(1)}\left(e^{0.09^2(1)} - 1\right) = 1{,}197.189781. \end{aligned}$$

and thus the standard deviation is 34.600430.

We illustrate the remaining statistics graphically in the next section.

Asset price distributions with various standard deviations

On the following pages, we have the probability density functions and the cumulative distribution functions with standard deviations of increasing magnitudes. Several important observations can be made from the following exhibits.

- With increasing volatility, we have an increasing mean, the median remains unchanged, and the mode declines.
- With increasing volatility, the skewness increases.
- With increasing volatility, the likelihood of observing a very low value increases at an increasing rate.

[12]Note that these statistics are reported without rounding error. If you verify these results, which we recommend as a learning exercise, you will have results slightly different.

Figure 3.7.1A. Stock Price = $100, Horizon = 1 year Expected Return = 12%, Standard Deviation = 30%

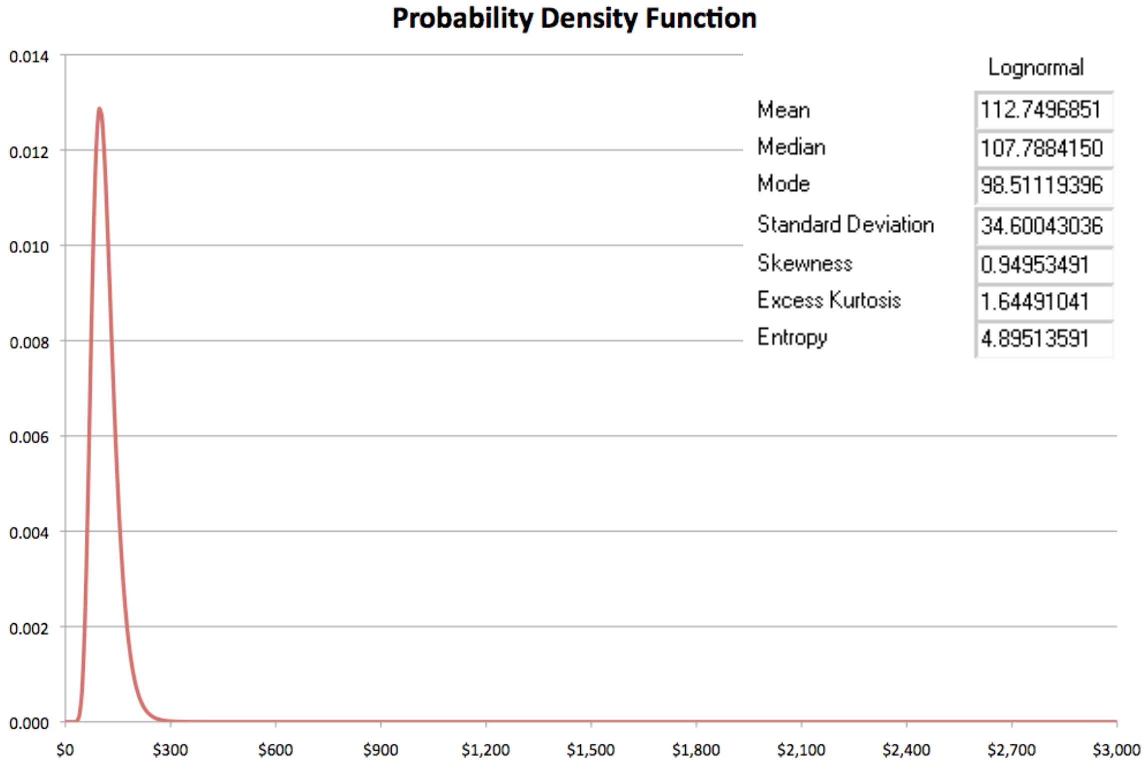


Figure 3.7.1B. Stock Price = $100, Horizon = 1 year Expected Return = 12%, Standard Deviation = 30%

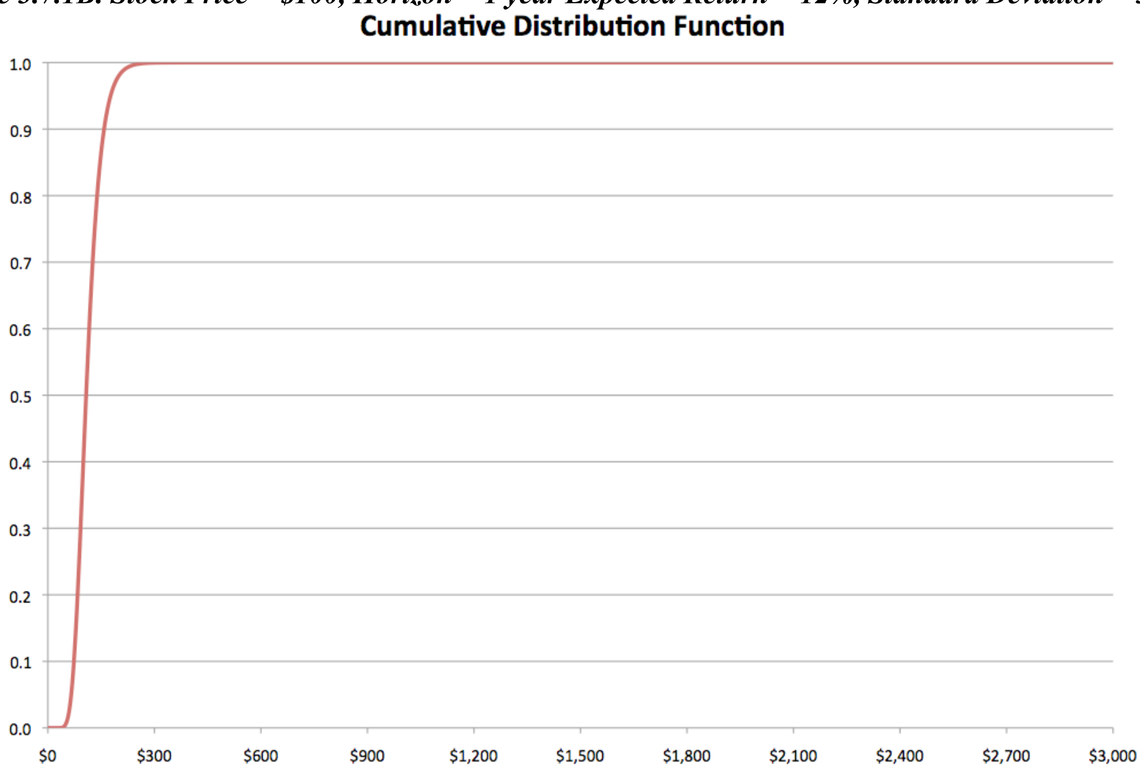


Figure 3.7.2A. Stock Price = $100, Horizon = 1 year Expected Return = 12%, Standard Deviation = 80%

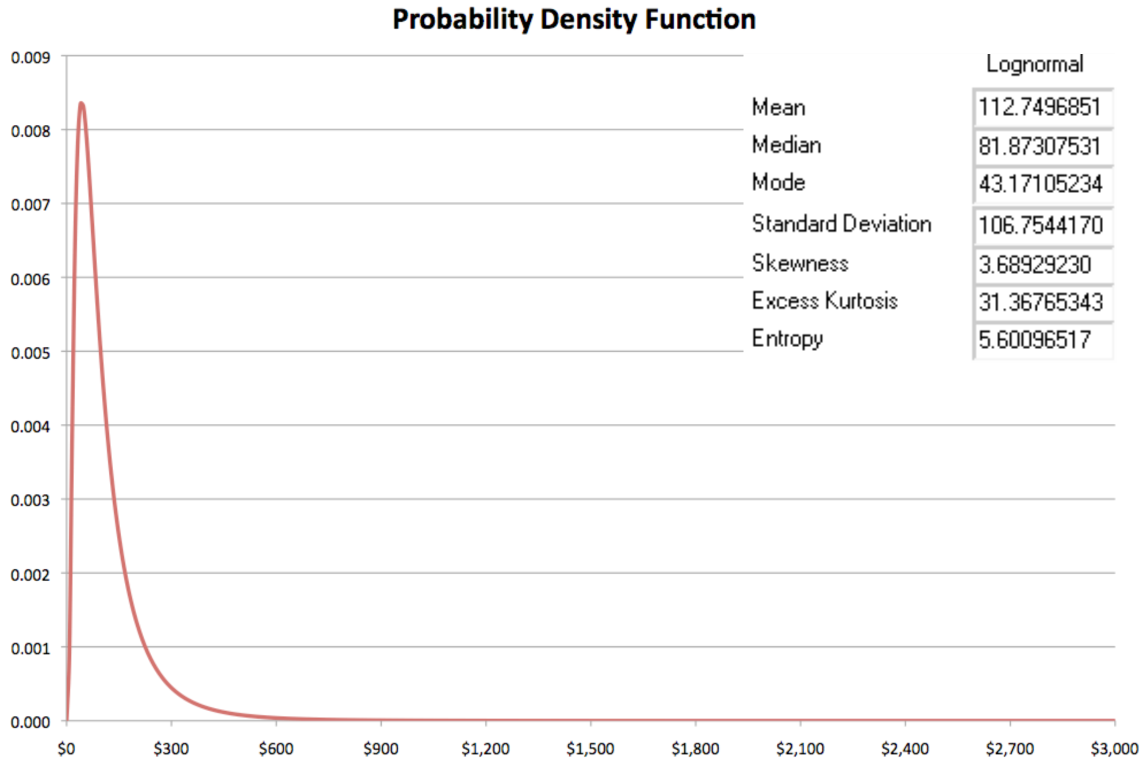


Figure 3.7.2B. Stock Price = $100, Horizon = 1 year Expected Return = 12%, Standard Deviation = 80%

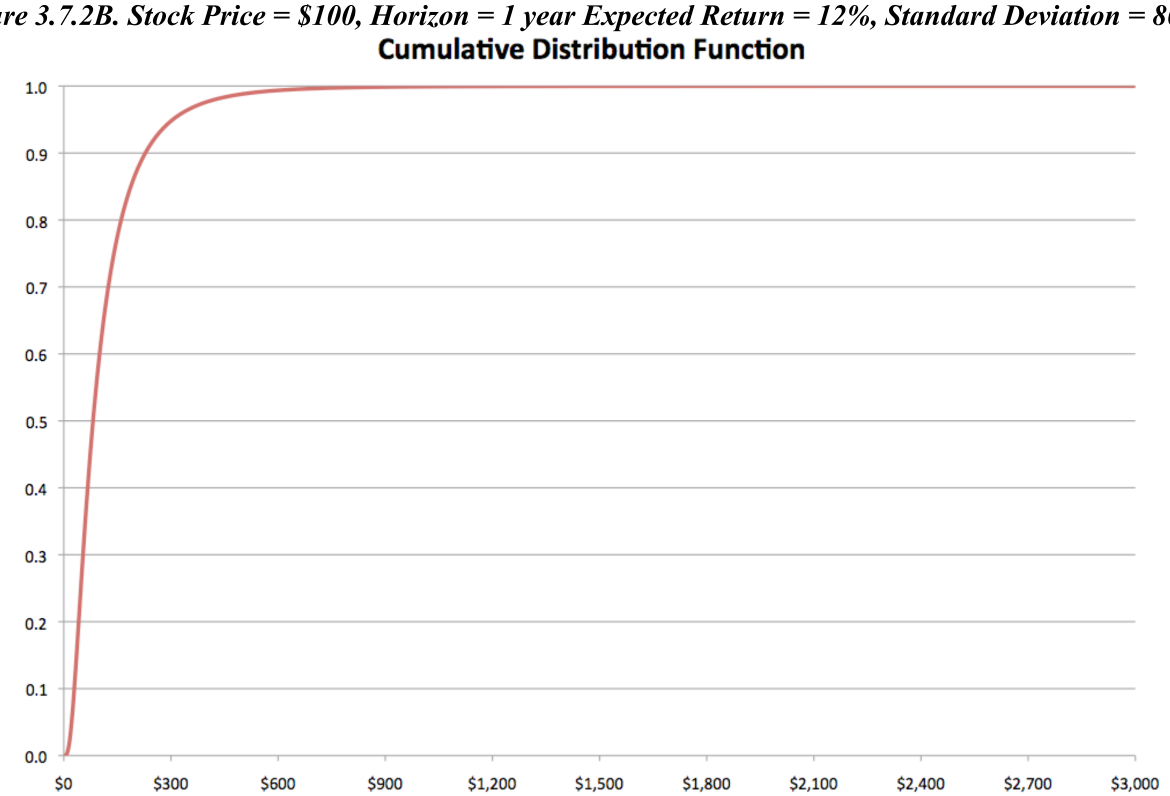


Figure 3.7.3A. Stock Price = $100, Horizon = 1 year Expected Return = 12%, Standard Deviation = 130%

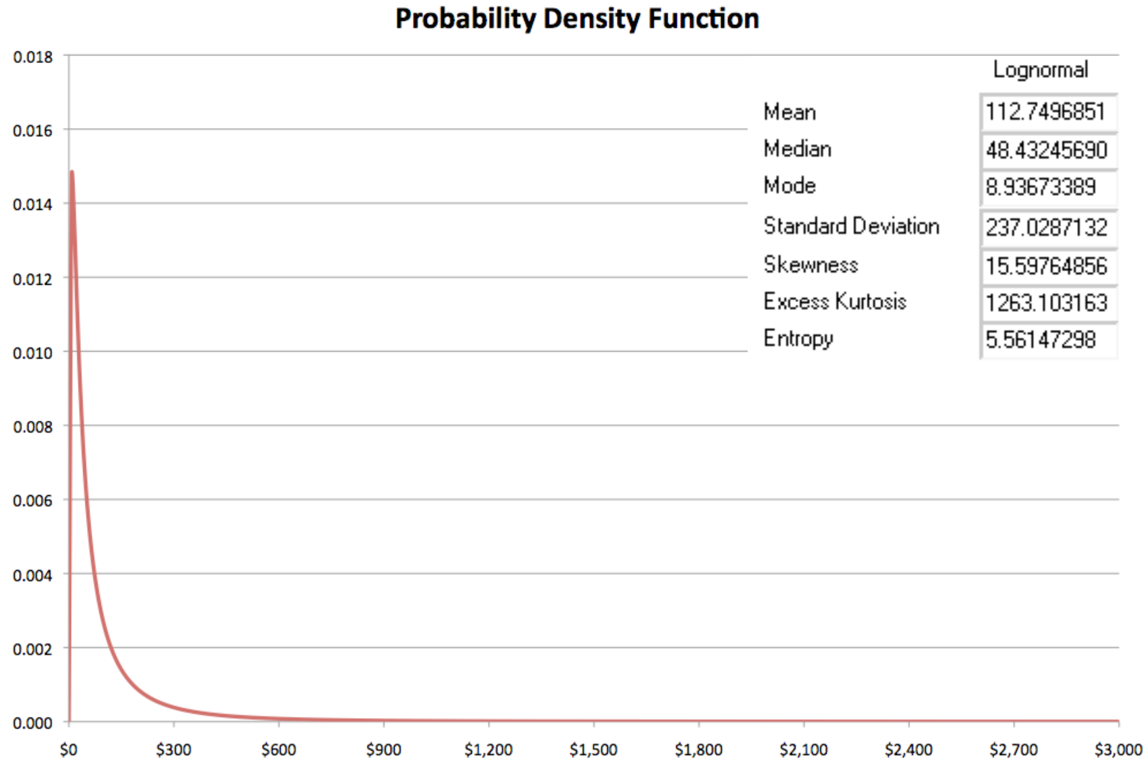


Figure 3.7.3B. Stock Price = $100, Horizon = 1 year Expected Return = 12%, Standard Deviation = 130%

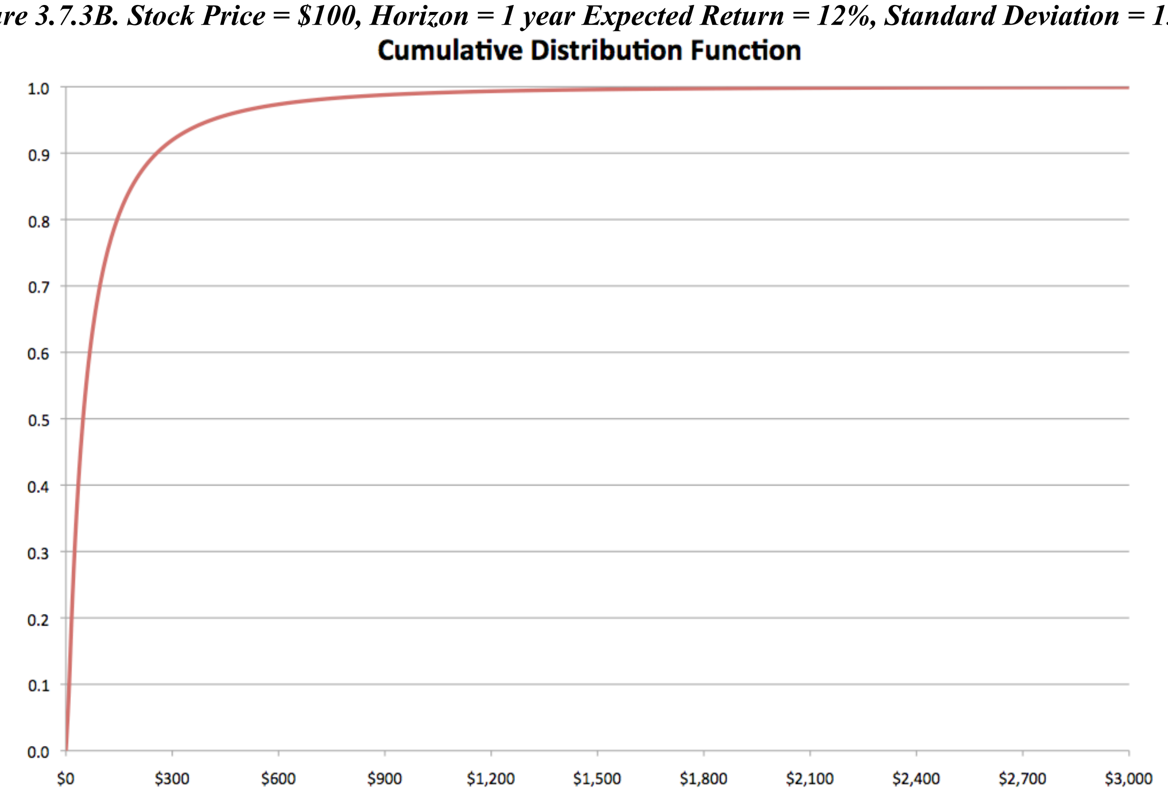


Figure 3.7.4A. Stock Price = $100, Horizon = 1 year Expected Return = 12%, Standard Deviation = 180%

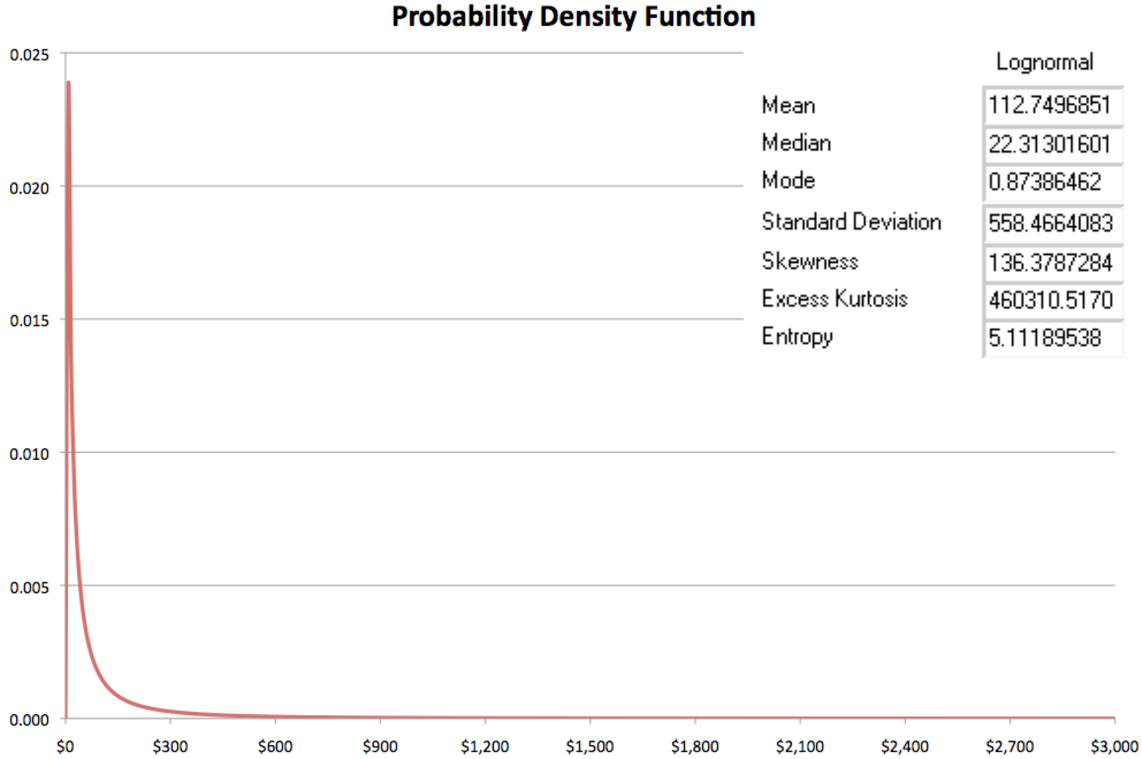
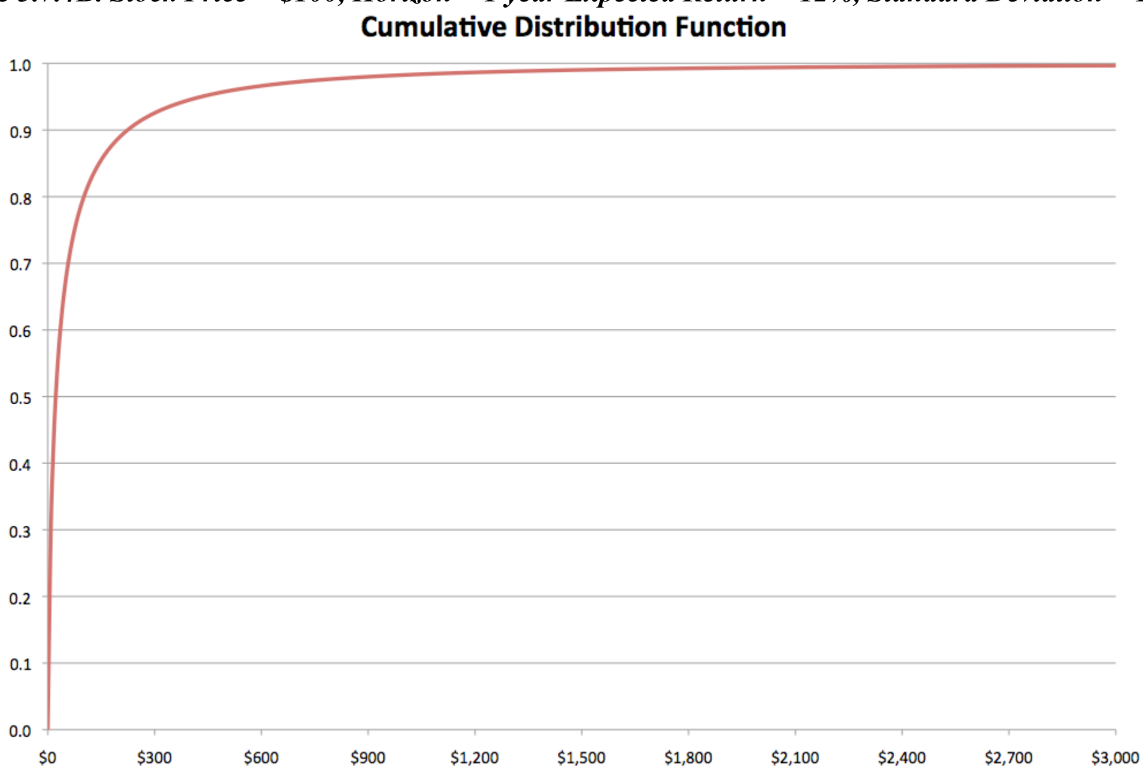


Figure 3.7.4B. Stock Price = $100, Horizon = 1 year Expected Return = 12%, Standard Deviation = 180%



We now illustrate one application of this analysis in the probability of a call option being in-the-money.

Probability call option in-the-money

Consider an underlying instrument that is lognormally distributed with a normally distributed mean μ and standard deviation σ. The value of the underlying instrument at some future point in time, say T, can be modeled as

$$\tilde{S}_T = S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}\tilde{\varepsilon}}, \tag{3.56}$$

Where we assume $\tilde{\varepsilon}$ is distributed normal with mean zero and standard deviation one. One can easily demonstrate that the mean and standard deviation of $\tilde{S}_T$ is consistent with the lognormal distribution.

The probability of a call option being in-the-money can be expressed as

$$\Pr\left(S_T > X\right) = \Pr\left[S_0 e^{\left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}\tilde{\varepsilon}} > X\right].$$

Rearranging, we note

$$\begin{aligned}
\Pr\left(S_T > X\right) &= \Pr\left[\ln\left(\frac{S_0}{X}\right) + \left(\mu - \frac{\sigma^2}{2}\right)T + \sigma\sqrt{T}\tilde{\varepsilon} > 0\right] \\
&= \Pr\left[\tilde{\varepsilon} > \frac{-\ln\left(\frac{S_0}{X}\right) - \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}\right] = 1 - \mathrm{Prob}\left[\tilde{\varepsilon} < -\frac{\ln\left(\frac{S_0}{X}\right) + \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}\right], \\
&= 1 - N\left(-d_2\right) = N\left(d_2\right)
\end{aligned}$$

where

$$d_2 \equiv \frac{\ln\left(\frac{S_0}{X}\right) + \left(\mu - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}} \text{ and}$$

$$N\left(d\right) \equiv \int_{-\infty}^{d} \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx.$$

We will return to this type of notation in Modules 5.4 and many others.

Multivariate Normal and Lognormal Distribution

Quantitative finance suffers from what is known as the "curse of dimensionality." In physics, often problems are couched in terms of three or four dimensions when modeling space and time. At times, perhaps based on string theory, physicist may model up to 11 or 12 dimensions. In finance, one easily struggles with hundreds of dimensions. For example, the S&P 500 index is technically a 500 dimensional problem. Further, if you consider each stock as having multiple factors or dimensions, one could easily get to thousands of dimensions. Thus, one quest is to aggressively seek to reduce dimensionality. In practice, one rarely gets the dimensions down to only one. Hence, we need the capacity to model multiple dimensional problems.

The probability density function of a multivariate normal distribution, denoted as $\mathbf{x} \sim N\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, can be expressed with matrix notation as

$$n_N\left(\mathbf{x} \middle| \boldsymbol{\mu}, \boldsymbol{\Sigma}\right) = \frac{e^{\left[-\frac{(\mathbf{x}-\boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x}-\boldsymbol{\mu})}{2}\right]}}{\sqrt{\left(2\pi\right)^N \left|\boldsymbol{\Sigma}\right|}}, \tag{3.57}$$

where

$\mathbf{x}$ denotes the random variable vector of size N,

$\boldsymbol{\mu}$ denotes the mean of the distribution, again size N,

$\boldsymbol{\Sigma}$ denotes the covariance of the distribution, matrix size NxN, assumed to be symmetric and positive definite,

$|\;|$ denotes the determinant, and

$(\;)'$ denotes the transpose.

The multivariate lognormal distribution cannot be written as concisely. If, however, $\mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ is a multivariate normal distribution and we define $\mathbf{y} = e^{\mathbf{x}}$, then $\mathbf{y}$ is said to follow a multivariate lognormal distribution, where for each element the mean and covariance can be expressed as

$$E(y_i) = e^{\mu_i + \frac{\Sigma_{ii}}{2}} \text{ and} \tag{3.58}$$

$$\operatorname{cov}(y_i, y_j) = e^{\mu_i + \mu_j + \frac{1}{2}(\Sigma_{ii} + \Sigma_{jj})}\left(e^{\Sigma ij} - 1\right). \tag{3.59}$$

For many applications, a bivariate distribution is adequate. The bivariate PDF and CDF for the normal distribution can be expressed as

$$n_2(x_1, x_2; \rho) \equiv \frac{e^{-\frac{\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x_1-\mu_1}{\sigma_1}\right)\left(\frac{x_2-\mu_2}{\sigma_2}\right) + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2}{2(1-\rho^2)}}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}. \tag{3.60}$$

$$N_2(a, b; \rho) \equiv \int_{-\infty}^{a}\int_{-\infty}^{b} n_2(x_1, x_2; \rho)\, dx_1\, dx_2\,. \tag{3.61}$$

The bivariate PDF and CDF for the lognormal distribution can be expressed as

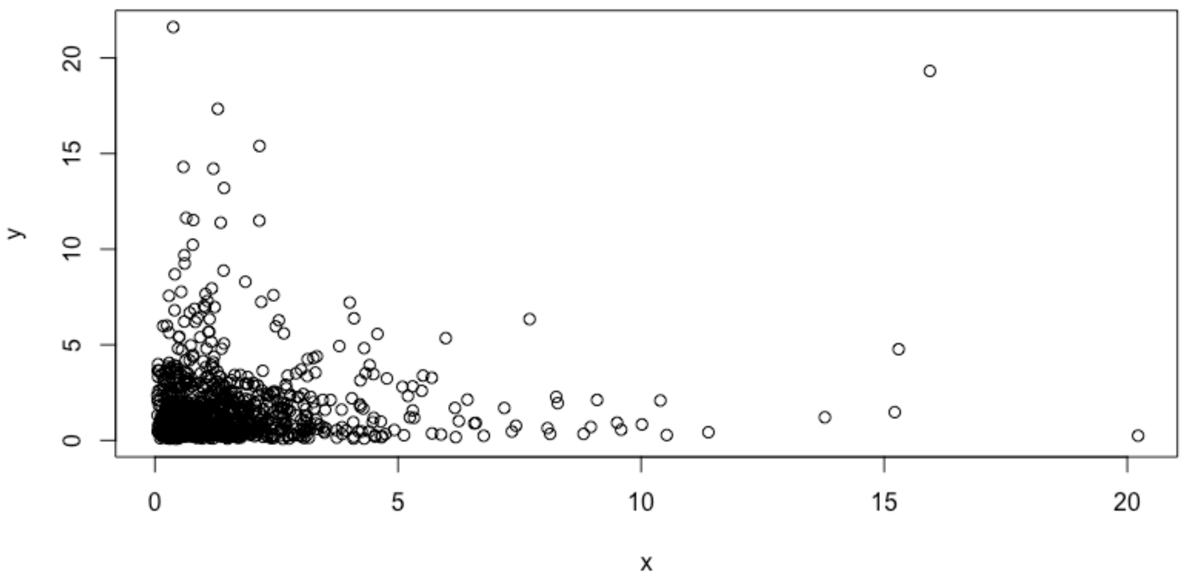
$$\lambda_2(x_1, x_2; \rho) \equiv \frac{e^{-\frac{\left[\frac{\ln(x_1)-\mu_1}{\sigma_1}\right]^2 - 2\rho\left[\frac{\ln(x_1)-\mu_1}{\sigma_1}\right]\left[\frac{\ln(x_2)-\mu_2}{\sigma_2}\right] + \left[\frac{\ln(x_2)-\mu_2}{\sigma_2}\right]^2}{2(1-\rho^2)}}}{2\pi x_1 x_2\sigma_1\sigma_2\sqrt{1-\rho^2}}. \tag{3.62}$$

$$\Lambda_2(a, b; \rho) \equiv \int_{-\infty}^{a}\int_{-\infty}^{b} \lambda_2(x_1, x_2; \rho)\, dx_1\, dx_2\,. \tag{3.63}$$

The relationship between stock price *i*, $S_{i,t}$, and the continuously compounded return, $R_{i,t}$, from time $t-1$ to time *t* is $S_{i,t} = S_{i,t-1}e^{R_{i,t}}$. Note that if $R_{i,t}$ is normally distributed, then we know S_t is lognormally distributed. The value of a portfolio of stocks in this case is the sum of lognormally distributed random variables. Unfortunately, the sum of lognormally distributed variables does not follow any known distribution. The sum of normally distributed random variables, however, is well-known to be normally distributed. Therefore, it is easier to assume the underlying stock prices are normally distributed.

Figure 3.7.5 illustrates a simulation of 1,000 draws from a bivariate lognormal distribution, with normal mean 0, variance 1, and correlation 0. Note that the bivariate lognormal lower bound is zero for both *x* and *y*. Zero is not feasible. Therefore, one advantage of the lognormal distribution is that negative values are not possible. Closely related, one disadvantage of the lognormal distribution is that zero values are also not possible.

Figure 3.7.5 Bivariate Lognormal Simulation



Conditional Normal and Lognormal Distribution

Consider two variables with correlated normal distributions, x_j for j = 1, 2. Thus, $-\infty < x_j < +\infty$ for j = 1, 2. We denote the distribution of these normal variables as

$$n\left(x_j\right) \sim N\left(\mu_j, \sigma_j^2\right); j = 1,2. \qquad (3.64)$$

Recall the univariate normal density function can be expressed as

$$n\left(x_j\right) = \frac{e^{-\frac{\left(x_j - \mu_j\right)^2}{2\sigma_j^2}}}{\sigma_j\sqrt{2\pi}}; j = 1,2. \qquad (3.65)$$

The correlation coefficient between x_1 and x_2 is expressed as ρ_{12} or just ρ. Thus, x_1 and x_2 are normal variates with a bivariate normal joint distribution.

Let

$$z_j = \frac{x_j - \mu_j}{\sigma_j}; j = 1,2. \qquad (3.66)$$

Rewriting the univariate normal density function either as a function of the underlying variable, x_j,

$$n\left(x_j\right) = \frac{e^{-\frac{z_j^2}{2}}}{\sigma_j\sqrt{2\pi}}; j = 1,2 \text{ or} \qquad (3.67)$$

as a function of z_j

$$n\left(z_j\right) = \frac{e^{-\frac{z_j^2}{2}}}{\sqrt{2\pi}}; j = 1,2. \qquad (3.68)$$

Recall the PDF is the result of a CDF. Hence, the transformation based on z_j results in the standard deviation being eliminated from the denominator.

$$\int_{-\infty}^{d} n\left(x_j\right)dx_j = \int_{-\infty}^{d} \frac{e^{-\frac{\left(x_j-\mu_j\right)^2}{2\sigma_j^2}}}{\sigma_j\sqrt{2\pi}}dx_j = \int_{-\infty}^{\frac{d-\mu_j}{\sigma_j}} \frac{e^{-\frac{z_j^2}{2}}}{\sigma_j\sqrt{2\pi}}\sigma_j\,dz_j = \int_{-\infty}^{\frac{d-\mu_j}{\sigma_j}} \frac{e^{-\frac{z_j^2}{2}}}{\sqrt{2\pi}}dz_j = \int_{-\infty}^{\frac{d-\mu_j}{\sigma_j}} n\left(z_j\right)dz_j. \tag{3.69}$$

The bivariate normal density function can be expressed as

$$n_2\left(x_1,x_2\right) = \frac{e^{-\frac{z_1^2-2\rho z_1 z_2+z_2^2}{2\left(1-\rho^2\right)}}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}. \tag{3.70}$$

The conditional normal density function is

$$n\left(x_2 \mid x_1\right) = \frac{e^{-\frac{\left(z_2-\rho z_1\right)^2}{2\left(1-\rho^2\right)}}}{\sigma_2\sqrt{2\pi\left(1-\rho^2\right)}}. \tag{3.71}$$

Proof: From the definition of conditional density function

$$n\left(x_2 \mid x_1\right) = \frac{n_2\left(x_1,x_2\right)}{n\left(x_1\right)} = \frac{\dfrac{e^{-\frac{z_1^2-2\rho z_1 z_2+z_2^2}{2\left(1-\rho^2\right)}}}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}}{\dfrac{e^{-\frac{z_1^2}{2}}}{\sigma_1\sqrt{2\pi}}}. \tag{3.72}$$

Rearranging

$$n\left(x_2 \mid x_1\right) = \frac{e^{-\frac{z_1^2-2\rho z_1 z_2+z_2^2}{2\left(1-\rho^2\right)}}}{\sigma_2\sqrt{2\pi\left(1-\rho^2\right)}\exp\left(-\frac{z_1^2}{2}\right)} = \frac{e^{-\frac{z_1^2-2\rho z_1 z_2+z_2^2}{2\left(1-\rho^2\right)}+\frac{z_1^2}{2}}}{\sigma_2\sqrt{2\pi\left(1-\rho^2\right)}}. \tag{3.73}$$

Further rearranging,

$$n\left(x_2 \mid x_1\right) = \frac{e^{-\frac{z_1^2-2\rho z_1 z_2+z_2^2-z_1^2\left(1-\rho^2\right)}{2\left(1-\rho^2\right)}}}{\sigma_2\sqrt{2\pi\left(1-\rho^2\right)}} = \frac{e^{-\frac{z_1^2\rho^2-2\rho z_1 z_2+z_2^2}{2\left(1-\rho^2\right)}}}{\sigma_2\sqrt{2\pi\left(1-\rho^2\right)}}. \tag{3.74}$$

Finally, we have

$$n\left(x_2 \mid x_1\right) = \frac{e^{-\frac{\left(z_2-z_1\rho\right)^2}{2\left(1-\rho^2\right)}}}{\sigma_2\sqrt{2\pi\left(1-\rho^2\right)}}. \tag{3.75}$$

Substituting and rearranging,

$$n\left(x_2 \mid x_1\right) = \frac{e^{-\frac{\left\{x_2-\left[\mu_2+\rho\frac{\sigma_2}{\sigma_1}\left(x_1-\mu_1\right)\right]\right\}^2}{2\sigma_2^2\left(1-\rho^2\right)}}}{\sigma_2\sqrt{2\pi\left(1-\rho^2\right)}}. \tag{3.76}$$

Therefore,

$$n\left(x_2 \mid x_1\right) \sim N\left(\mu_{2|1},\sigma_{2|1}^2\right), \tag{3.77}$$

where

$$\mu_{2|1} = \mu_2 + \rho \frac{\sigma_2}{\sigma_1}\left(x_1 - \mu_1\right) \text{ and} \tag{3.78}$$

$$\sigma_{2|1}^2 = \sigma_2^2\left(1 - \rho^2\right). \tag{3.79}$$

Thus, we have numerous ways to find quantitative solutions with integration.

We conclude this module by illustrating several aspects of the normal and lognormal distributions. Figure 3.7.6 illustrate the normal distribution PDF with increasing volatility.

Figure 3.7.6 Univariate Normal PDFs (30%, 80%, 130%, 180%)

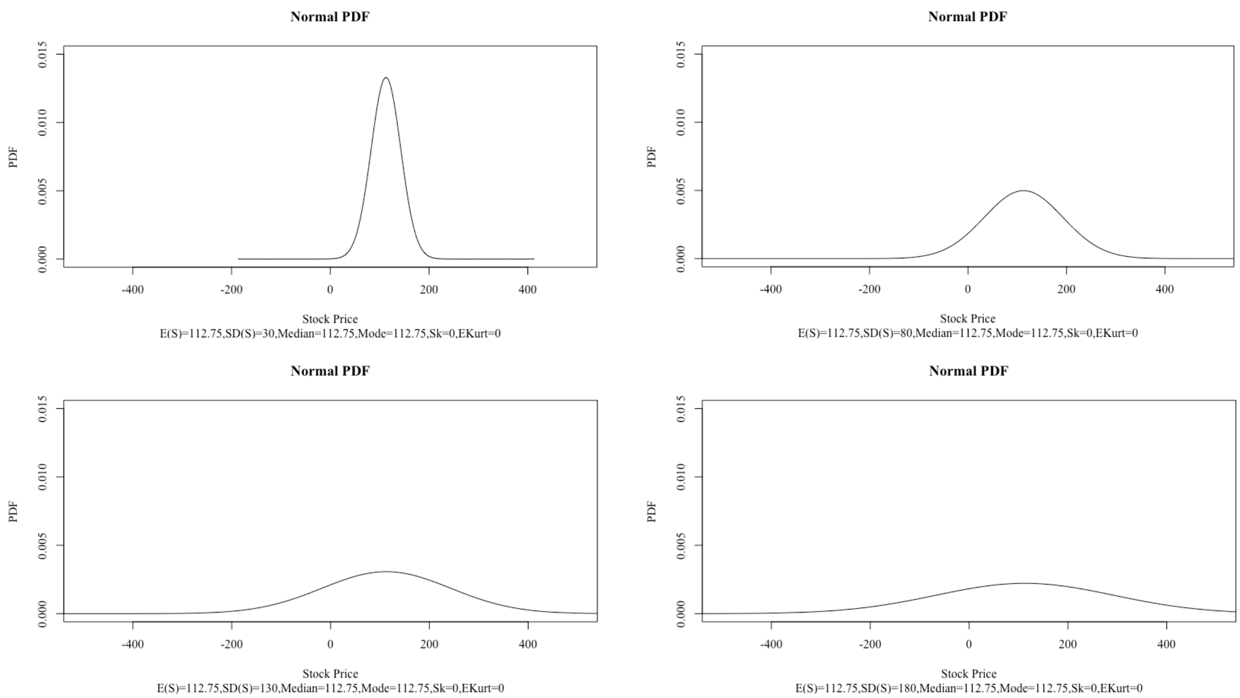
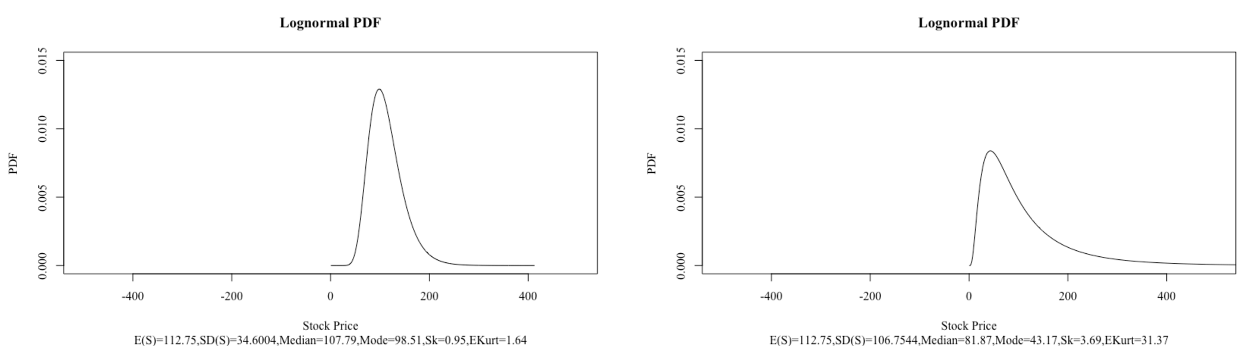
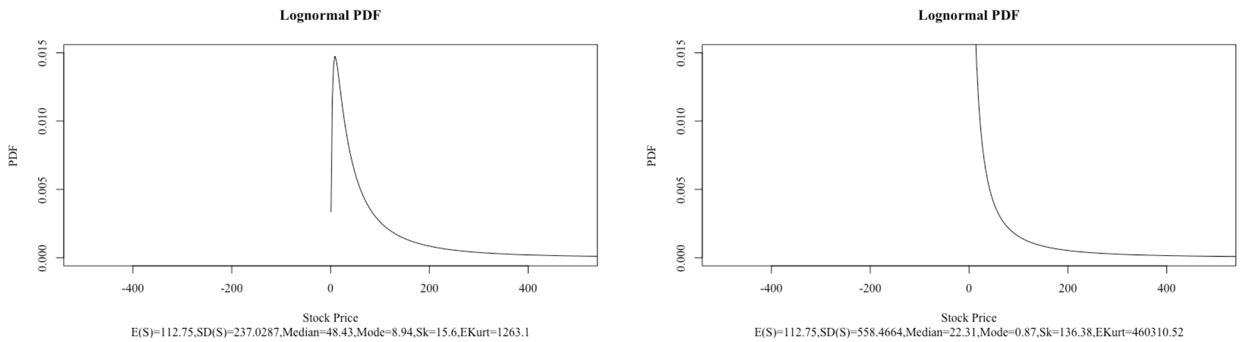


Figure 3.7.7 illustrates the lognormal distribution PDF with increasing volatility. The volatility is calibrated to be roughly equivalent to the previous figure.

Figure 3.7.7 Univariate Lognormal PDFs ($34, $106.75, $237, $558)





We clearly see the existence of negative stock prices with the normal distribution and non-zero stock prices with the lognormal distribution. Figure 3.7.8 illustrate the normal distribution CDF with increasing volatility.

Figure 3.7.8 Univariate Normal CDFs (30%, 80%, 130%, 180%)

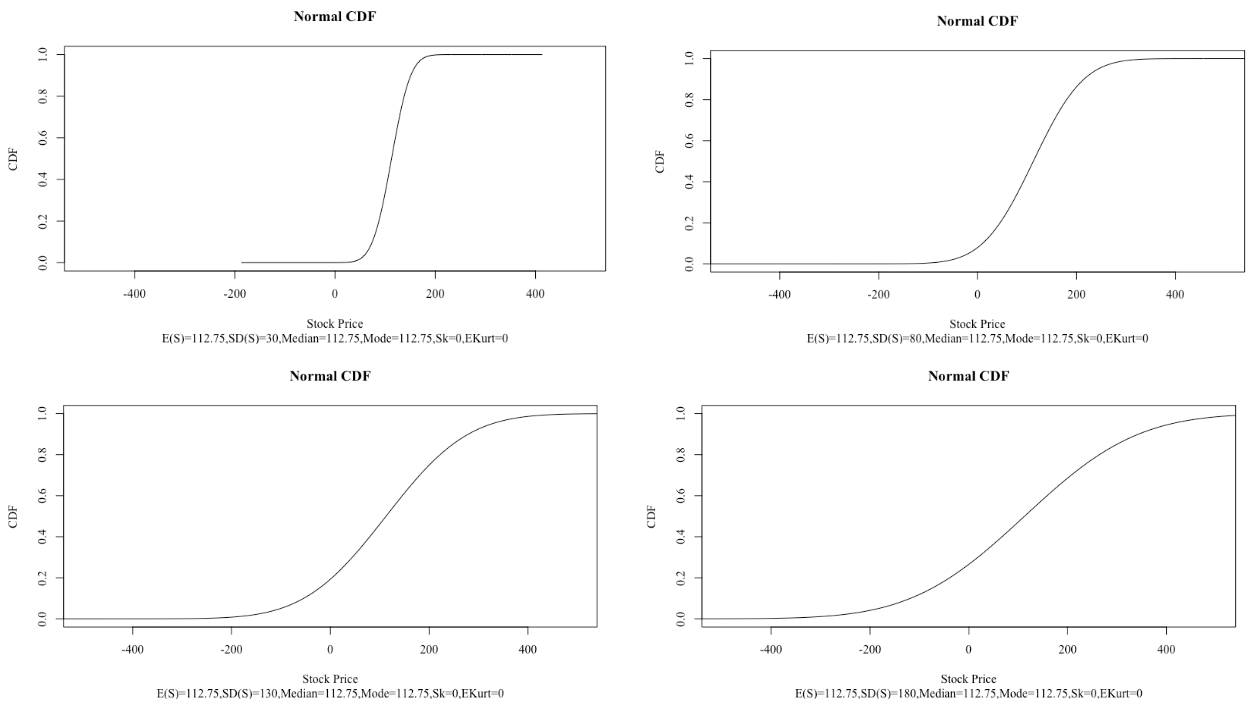
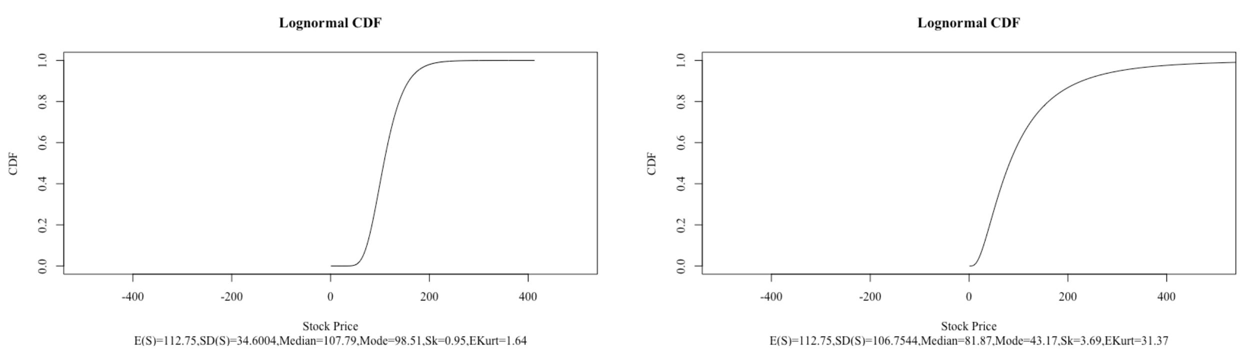
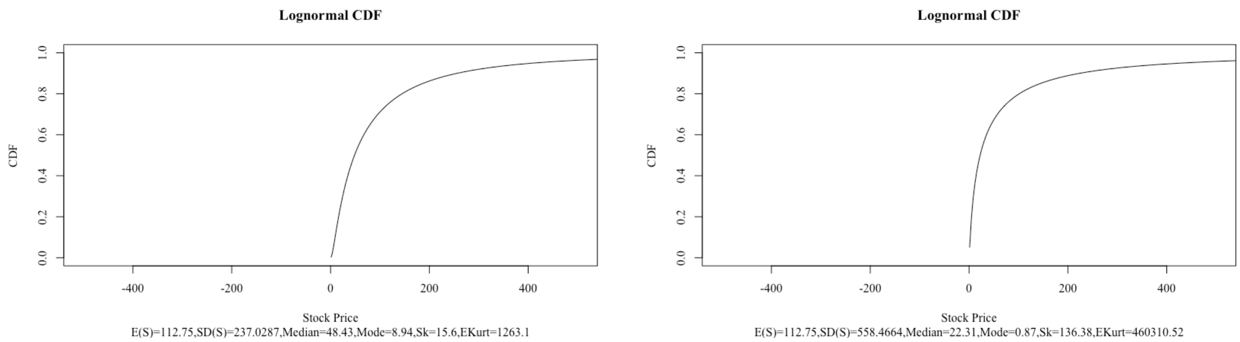


Figure 3.7.9 illustrates the lognormal distribution CDF with increasing volatility.

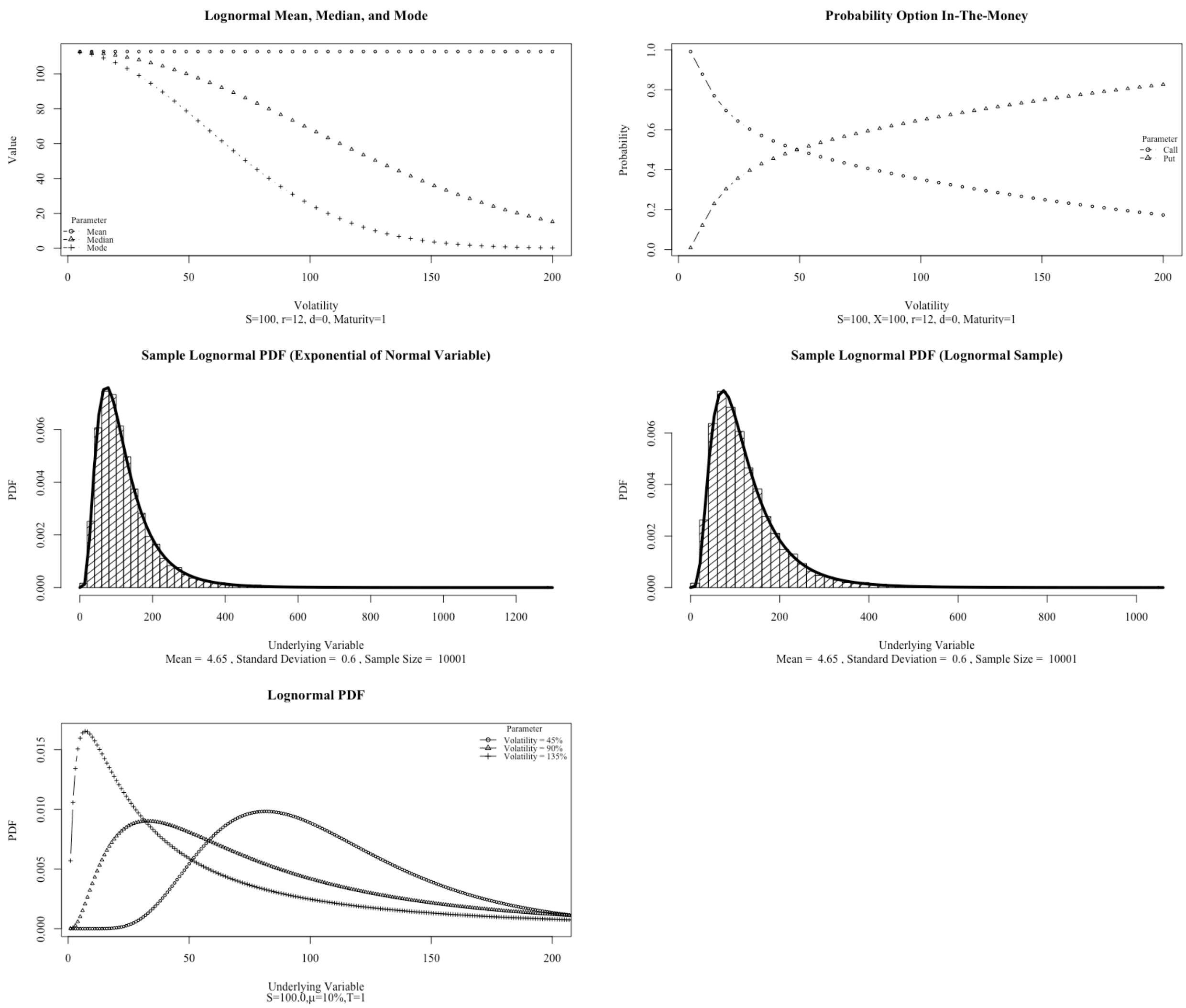
Figure 3.7.9 Univariate Lognormal CDFs ($34, $106.75, $237, $558)





Several other interesting plots are produced in this code shown in Figure 3.7.10.

Figure 3.7.10 Univariate Lognormal PDFs ($34, $106.75, $237, $558)



Summary

In chapter 3, we reviewed several important tools for performing quantitative finance tasks. In the remainder of this book, these tools are deployed to specific tasks.

References

Aitchison, J. and J. A. C. Brown, *The Lognormal Distribution with special reference to its uses in economics*, (Cambridge University Press, 1957).
Barrett, W., T. Gosnell, Jr., and A. Heuson, 1995. Yield curve shifts and the selection of immunization strategies. Journal of Fixed Income September 1995, 53-64.
Crow, E. L. and K. Shimizu, Eds., *Lognormal Distributions: Theory and Applications*, (New York: Dekker, 1988).
Dobson, S., 1978. Estimating term structure equations with individual bond data. Journal of Finance 33, 75-92.
Durand, D., 1942. Basic yields of corporate bonds, 1900-1942. National Bureau of Economic Research, Technical Paper no. 3, Cambridge, MA.
Echols, M. and Elliott, J., 1976. A quantitative yield curve model for estimating the term structure of interest rates. Journal of Finance and Quantitative Analysis 11, 87-114.
Galton, F., *Natural Inheritance* (London: Macmillan, 1889).
Limpert, E., M. Abbt, and W. A. Stahel, *Lognormal Distributions across the sciences – keys and clues* (2000, submitted) www.inf.ethz.ch/~gut/lognormal/.
Litterman, R. and Scheinkman, J., 1991. Common factors affecting bond returns. Journal of Fixed Income, 1 (June 1991), 54-61.
Macaulay, F., 1938. Some theoretical problems suggested by the movement of interest rates, bonds, yields, and stock prices in the United States since 1856. Columbia University Press, New York, NY.
Malkiel, B., 1966. The term structure of interest rates. Princeton University Press, Princeton, N.J.
Mayle, Jan, *Standard Securities Calculation Methods Fixed Income Securities Formulas for Price, Yield, and Accrued Interest* Volume I, Third Edition (New York: Securities Industry Association, 1993).
McAlister, D., *Proceeding of the Royal Society* 29 (1879), p. 367.
McCulloch, J., 1971. Measuring the term structure of interest rates. Journal of Business 34, 19-31.
Nelson, C., and Siegel, A., 1987. Parsimonious modeling of yield curves. Journal of Business 60, 473-489.
Press, William H., Brian P. Flannery, Saul A. Teukolsky, and William T. Vetterling, *Numerical Recipes in C The Art of Scientific Computing* Second Edition, (Cambridge, England: Cambridge University Press, 1989).
Stuart, Alan and J. Keith Ord, *Kendall's Advanced Theory of Statistics Volume 1 Distribution Theory*, Fifth Edition (New York: Oxford University Press, 1987).
Steeley, J., 2008. Testing term structure estimation methods: Evidence from the UK STRIPS Market. Journal of Money, Credit and Banking 40, 1489-1512.
Svensson, L., 1995. Estimating futures interest rates with the extended Nelson and Siegel method. Institute for International Economic Studies, Stockholm University, Reprint Series, No. 543 from Quarterly Review (Swedish Central Bank) No. 3.
Vasicek, O. and Fong, G., 1982. Term structure modeling using exponential splines. Journal of Finance 37, 339-48.
Wichura, Michael J., "Algorithm AS 241: The Percentage Points of the Normal Distribution," *Applied Statistics* 37(3), 1988, 477-484.
Willner, R., 1996. A new tool for portfolio managers: level, slope, and curvature durations. Journal of Fixed Income 6, 48-59.