

# **Retirement Income Analysis with scenario matrices**

**William F. Sharpe**

## **5. Inflation**

### ***Price Indices***

To belabor the obvious, this is a book about retirement income. And income is conventionally measured in terms of some unit of currency, such the United States dollar. But such income is only a means to the end of providing human needs and luxuries – food, housing, entertainment, health care and the myriad of goods and services that make life possible and enjoyable. Clearly we cannot deal with consumption with this level of granularity. But we can and will focus on a broad metric that takes into account some measure, however imperfect, of the overall cost of living.

In the United States, the U.S. Bureau of Labor Statistics computes a number of measures of the cost of purchasing baskets of goods and services. The most cited is the Consumer Price Index for All Urban Consumers, or CPI-U. Each month, the BLS gathers pricing information for a “basket” of goods and services then computes the cost in dollars for the overall list. The cost at a selected starting month is taken as an index of 100, with the relative costs for subsequent months used to compute associated indices. The basket of goods and services is designed to be broadly representative of consumption of those living in urban areas, with changes made from month to month to attempt to take into account changes in habits, responses to changes in relative prices, technological change, and so on. A related index, the CPI-W (for all Urban Wage Earners and Clerical Workers) is used to adjust the benefits paid by the U.S. Social Security Administration.

The choice of the ingredients in baskets of goods and services used to compute a price index and the procedures for changing their weights over time is controversial and subject to considerable debate on both economic and political grounds. The CPI-U and CPI-W are based on prices of over 200 goods and services in almost 40 different areas of the country, with different weights used to reflect the spending habits of the two prototypical groups. Substitution of different goods and services in response to changes in relative prices is taken into account but, according to some critics, insufficiently and/or too infrequently. A number of commissions have examined these issues and proposed alternatives. One, the Chained Consumer Price Index, or C-CPI-U is also computed by the BLS.

At one point, the BLS examined the possibility of a cost of living index that would better reflect the purchases of older Americans, notably giving greater weight to housing and health care expenses. The resulting CPI-E, computed for the period from 1982 through 2011, was found to differ relatively little from the CPI-U. Over the entire period, the CPI-E grew at an average annual rate of 3.1% while the CPI-U rose at a rate of 2.9%. Moreover, the two moved in close concert throughout the period. For good or ill, the CPI-E was abandoned and is no longer computed.

Of course no such index can precisely measure the cost of obtaining a specified level of happiness for any particular consumer or household, let alone thousands or millions of diverse people. But in most cases it will be far better to take into account such economy-wide changes than to ignore them completely. Hence our inclusion of matrices of changes in the cost of living.

## Nominal and Real Incomes

Assume that you obtained an income of \$40,000 last year and the same amount this year. Your *nominal income* was \$40,000 in each year. But if prices are now 5% higher than they were last year you can purchase  $\$40,000/1.05$  as much as last year – your *purchasing power* has fallen. To take this into account, we can express each income in terms of the (estimated) amount in a base year that would have purchased the same goods and services. We call these adjusted amounts *real incomes*. They better represent your attainable standard of living when the *cost of living* changes.

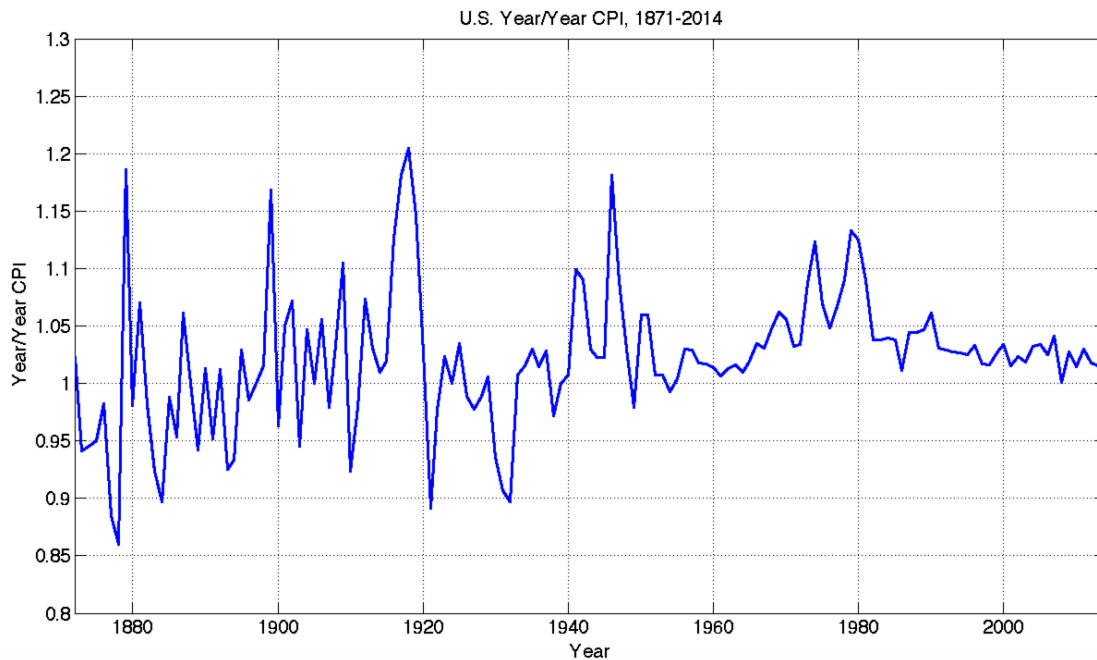
To focus on what matters, we express all our estimates of income in real terms. But to accommodate different sources of income we need to consider changes in the cost of living. While such changes have been negative at times, they are usually positive. Using standard terminology, *deflation* is far less common than *inflation*.

Let  $C(t)$  be the ratio of the cost of living in year  $t$  to the cost in year 1. If  $N(t)$  is the nominal income in year  $t$ , then the real income, expressed in terms of the cost in year 1 is  $N(t)/C(t)$ . We need to generate a matrix of values of  $C(t)$  for different scenarios (rows) and years (columns). Then it will be a simple matter to convert nominal incomes to real incomes, and/or vice-versa.

Note in passing, that such calculations are more precise than the common procedure of subtracting or adding the rate of inflation. For example, assume an investment provides a nominal return of 10% in a year in which inflation is 3%. A simple calculation could conclude that the real value increased 7% (10% - 3%). But in fact the investment changed \$1 into \$1.10, which had a purchasing power of  $\$1.10/1.03$ , or \$1.068. The real return was thus 6.8%. Not a major difference in this case, but potentially substantial for longer periods and more dramatic price changes. For this reason we will express changes in the cost of living in terms of ratios of year-end values.

## Historic Inflation in the United States

The figure below shows year-over-year ratios of measures of the cost of living in the United States from 1871 through 2014, as provided on the website of Robert Shiller at Yale University.



As can be seen, there have been dramatic changes in the cost of living, with prices increasing as much as 20% in one year and falling in a number of years by 10% or, in one case, more. The 1920's and 1930's saw more deflation than inflation, but from 1940 on, deflation has been the exception, with prices increasing in almost every year. From the early 1980's onward, inflation has been relatively mild, varying from year to year between 0 and 5%.

Many factors contribute to changes in overall price levels. A simple adage holds that inflation results from “too much money chasing too few goods.” While this may be true, the reality is far more complex. In the United States, Europe and other major economies, *central banks* are charged with attempting to keep aggregate prices in reasonable ranges by adjusting the money supply, regulating banks and engaging in financial operations.

An overly simplistic view of the argument for modest inflation is that it allows for more efficient utilization of labor and for better contracting between parties in general. For example, if overall prices and wages increase, an employer can reward better workers with raises and leave other workers' wages the same. But in a deflationary economy it will be necessary to cut some or all wages and prices, which may be difficult or impossible due to existing contracts or labor agreements. A little inflation, it is hoped, can lubricate the economic system, allowing it to perform more efficiently and maintain high levels of employment. While this may be true, it does pose a challenge for retirement income planning. If inflation is more likely than not, it is imperative that one concentrate on future *real* income, as we will do.

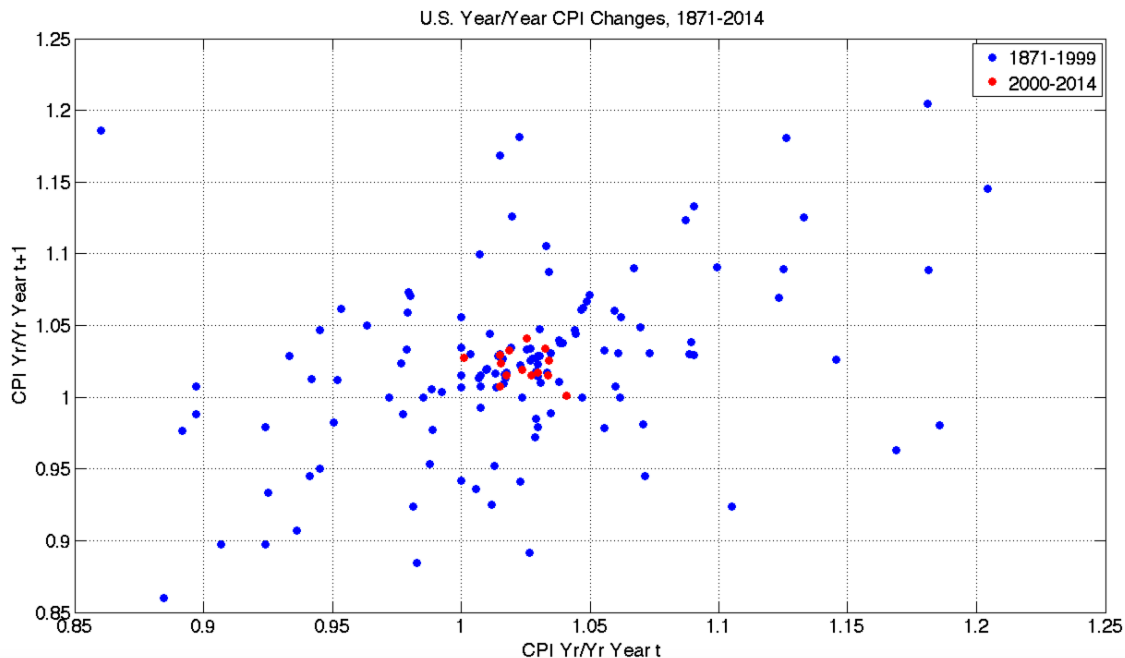
Most countries, have modest inflation as a goal. The United States Federal Reserve Bank is typical, with an explicit goal of 2% inflation. As stated on its website in 2015:

*“Why does the Federal Reserve aim for 2 percent inflation over time? The Federal Open Market Committee (FOMC) judges that inflation at the rate of 2 percent (as measured by the annual change in the price index for personal consumption expenditures, or PCE) is most consistent over the longer run with the Federal Reserve's mandate for price stability and maximum employment. Over time, a higher inflation rate would reduce the public's ability to make accurate longer-term economic and financial decisions. On the other hand, a lower inflation rate would be associated with an elevated probability of falling into deflation, which means prices and perhaps wages, on average, are falling--a phenomenon associated with very weak economic conditions. Having at least a small level of inflation makes it less likely that the economy will experience harmful deflation if economic conditions weaken. The FOMC implements monetary policy to help maintain an inflation rate of 2 percent over the medium term.”*

By luck or design, in the U.S. this goal has been mostly achieved in the current century. From 2000 to 2014 the average change in the cost of living was 2.25%, with a standard deviation of 1.09%.

## Serial Correlation in Cost of Living Changes

Historically, periods of above-average inflation have tended to be followed by periods of inflation that are also above average, and periods of below-average inflation have also tended to persist. This can be seen in the figure below, which shows the inflation in each year on the x-axis and the inflation in the following year on the y-axis.



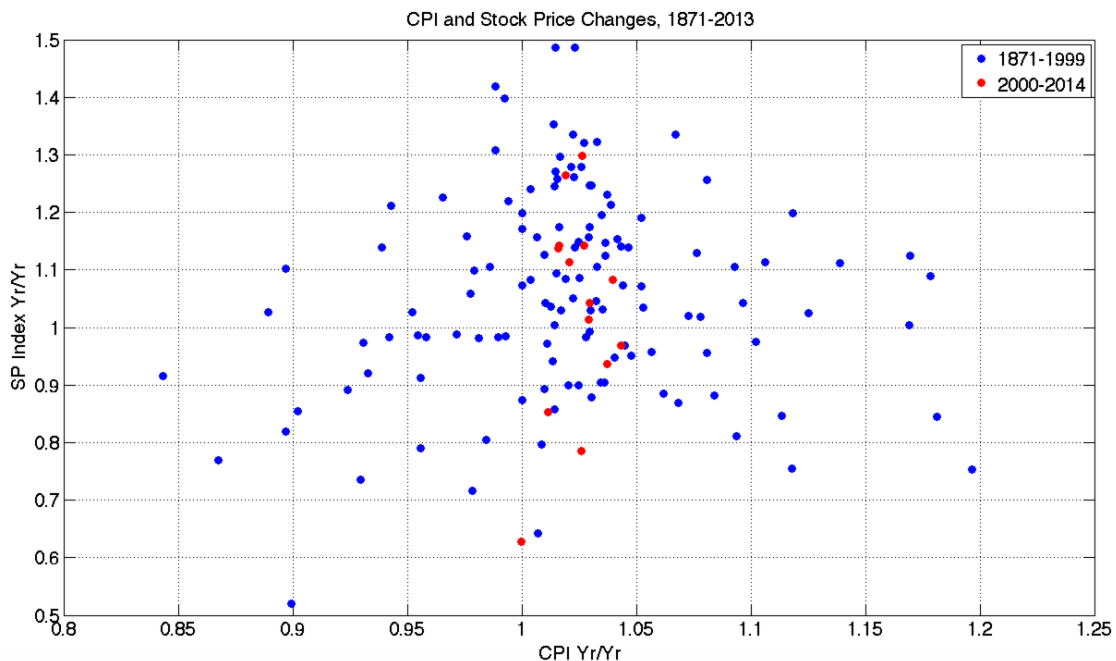
If a straight line were fit to all the points, it would be upward-sloping, indicating that there is a tendency for positive persistence in inflation. And the relationship is statistically significant, with a t-value of 4.98, considerably above the standard 2.0 threshold. This said, the red points representing the 21<sup>st</sup> century show no such tendency. In fact the relationship for these years is slightly negative, although not statistically significant.

Perhaps recent history is an aberration and the future will be characterized by widely varying inflation with the possibility of significant persistence. On the other hand, it is at least possible that central banks will maintain enough control over inflation to produce variations more like those of the early part of the century. As will be seen, for the examples in this book we take the more optimistic outlook, choosing to assume relatively low expected inflation, relatively modest uncertainty and no predictable persistence from year to year.

## Correlations of Changes in the Cost of Living and Stock Returns

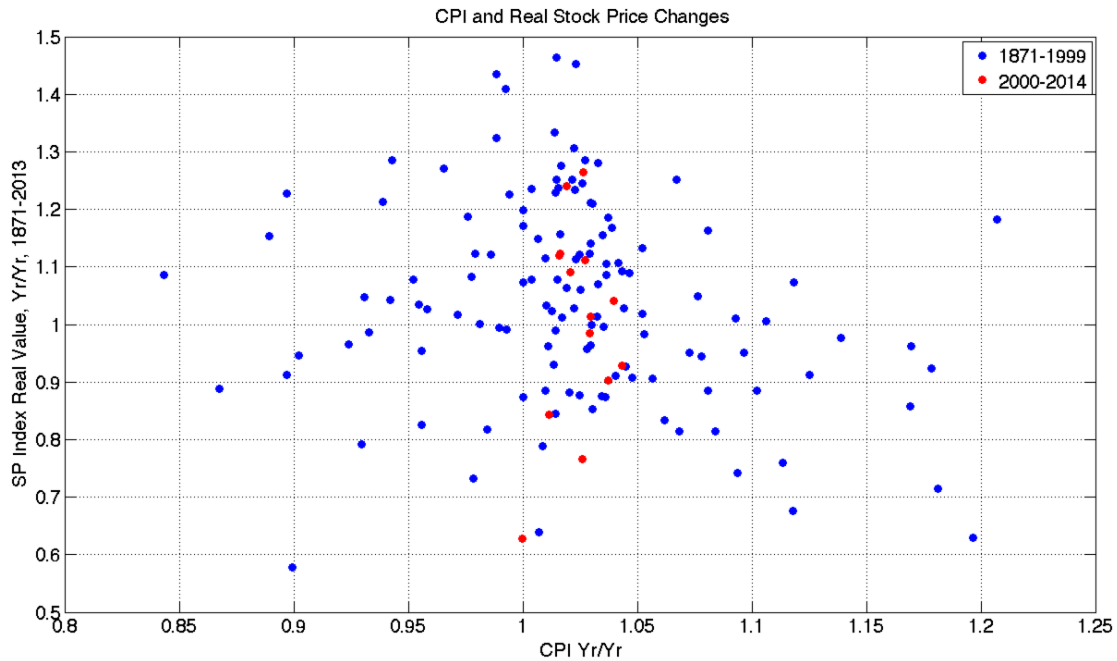
In the next chapter we will discuss returns on investments, including those of a portfolio of bonds and stocks. An important issue is the possible correlation of real and/or nominal returns on such a portfolio with changes in the cost of living. Since most of the variation in the value of such a portfolio comes from changes in the levels of stock markets, it is useful to examine historic relationships between changes in such levels and inflation.

The next figure provides data from Robert Shiller's website. It shows the relationship between changes in the CPI and changes in the *nominal* value of the stock market.



As can be seen, the two variables are virtually unrelated. There is a small positive correlation, but it is statistically insignificant (with a t-value of +1.66).

This said, we will be more concerned with changes in the *real* value of a portfolio of risky investments. And, since each such a change is equal to the ratio of nominal portfolio values over that of the cost of living, there is more likely to be a negative correlation between such changes and changes in the cost of living. The next figure shows that this was the case for the overall period from 1871 through 2013. The t-statistic was statistically significant (-2.21), although not dramatically so. However, during the 21<sup>st</sup> century, there was, if anything, a positive correlation, but it was statistically insignificant, with a t-statistic of +0.80.



Given our focus on the more recent past inflation experience in the United States, we will choose to assume zero correlation between changes in the cost of living and the real returns on investments for the examples in this book. No parameters or computations will be included to accommodate other assumptions, although the task might not prove particularly arduous for those who might wish to do so.



## Purchasing Power Parity

While we will express incomes and many other economic quantities in real terms, it will still be necessary to choose a currency to serve as a unit of account. Given the author's domicile and the importance of the United States in the global economy, we succumb to the temptation to use the U.S. Dollar. Thus nominal values will be in current U.S. dollars and real values will be in units of purchasing power equal to those in U.S. Dollars at the beginning of the analysis (year 1). For parsimony we will use the term “*dollar*” and the symbol \$ when needed.

These decisions may not be as limiting as might first appear. There is a famous concept in international economics that holds that in the absence of trade restrictions, taxes, shipping costs, etc. goods should sell for the same price in different currencies. This if commodity  $M$  sells for \$4.79 U.S. Dollars and it is possible to trade one U.S. Dollar for 0.8631 Euros, purchasing power parity would indicate that its price in Europe should be  $4.79/0.8631$  (4.13) Euros. In fact, these are historic values for the “Big Mac” hamburger sold by McDonalds in most parts of the world. In January 2015, a Big Mac sold for \$4.79 in the U.S. And one could trade one U.S. Dollar for 0.8631 Euros. If indeed there were purchasing power parity, the price of a Big Mac in Europe should have been 4.13 Euros. But it wasn't. At the time the average price of a Big Mac in Europe was 3.68 Euros – a bargain for the hungry U.S. citizen.

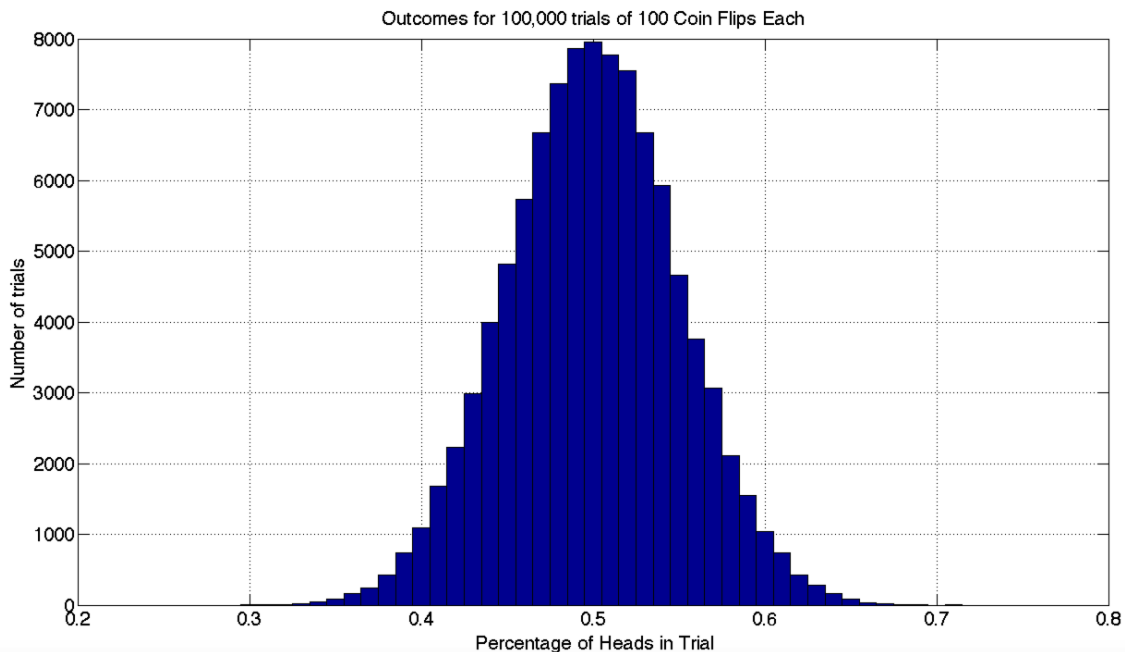
Why didn't purchasing power parity hold for this standardized product? Primarily because you can't ship a ready-to-eat hamburger across the Atlantic Ocean in zero time for zero cost. There are frictions of many kinds in international trade including shipping costs, quality deterioration, tariffs, etc.. That said, there is a tendency for the aggregate prices of large baskets of goods and services across national and currency boundaries when converted to a single currency to converge to at least some extent and for large disparities in such values to diminish over time.

Partly as a lark, for many years the Economist magazine has been calculating purchasing power disparities in the price of a Big Mac across a number of countries. The figures in the prior paragraph are from their database which is entertaining, if not particularly useful as a guide to preferred travel destinations.

While parochially using the U.S. Dollar as our base currency, we will make assumptions consistent with the purchase of securities representing investments around the globe. We will also assume consumption that includes goods from different countries. To be sure, the real returns in any given scenario might differ depending on one's domicile and base currency, but the range of real outcomes across scenarios may still be representative of outcomes for a broader group of retirees. In any event, we will assume a level of abstraction at which the risk associated with the cost of living will suffice to cover uncertainty about changes in the prices of goods and services both at home and abroad.

## Normal Distributions

In the early 18<sup>th</sup> century, a mathematician named Abraham de Moivre, who was born in France but spent much of his life in England, discovered that when one flipped a coin over and over, the proportion of heads began to plot as a symmetric, somewhat bell-shaped curve. Here is a modern version, using the `randi()` function in Matlab, to flip the coins (in this case, with 100,000 trials, each involving the simulated flipping of 100 coins).



In 1778, Pierre-Simon de Laplace, a French mathematician, put forth the *central limit theorem*, which holds that the sum or average of the results from a number of trials, each made independently of each other, will converge towards such a *normal distribution* as the number of trials increases. The actual formula for such a distribution is usually credited to the German mathematician Johann Carl Friedrich Gauss, leading some to term it the Gaussian distribution.

Fortunately for us, Matlab knows the formula for the normal distribution and how to create random samples from it. Here is an example:

```
M = 100;  
N = 20;  
mm = randn(M,N)
```

In this case, *mm* will be a matrix with 100 rows and 20 columns, with each entry a number drawn randomly from a normal distribution with a mean of 0 and a standard deviation of 1. Since the normal distribution is symmetric, the *mean* value (obtained by weighting each element by its probability, then summing) is the same as the *median* value (the value below which lie the same percentage of the elements as lie above it) and the *mode* (the value or midpoint of the range of values with the highest probability).

In investment parlance, the *mean* of a probability distribution of returns is often called the *expected return*. Often people assume that it equals the *median* of possible future returns, so that there is roughly a 50/50 chance of obtaining a higher or lower outcome. This is true for a symmetric distribution, but not in most other cases, including ones of great importance for us.

Now, to a measure of dispersion. The *standard deviation* of a distribution is the square root of a sum of elements, each of which is the square of the deviation of a value from the mean. In a normal distribution, two-thirds of the values lie within plus or minus one standard deviation of the mean.

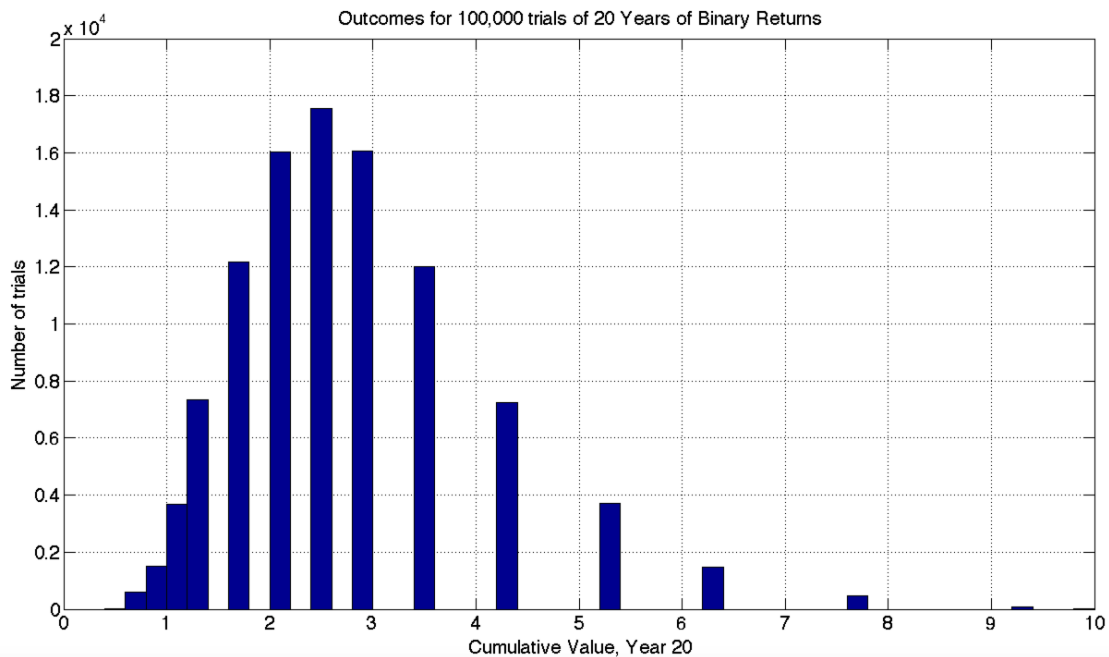
It is a simple matter to create a matrix of random normally-distributed values with a given mean and standard deviation. For example:

```
M = 100;  
N = 20;  
mn = 1.05;  
sd = 0.10;  
mm = mn + sd*randn(M,N)
```

## Lognormal Distributions

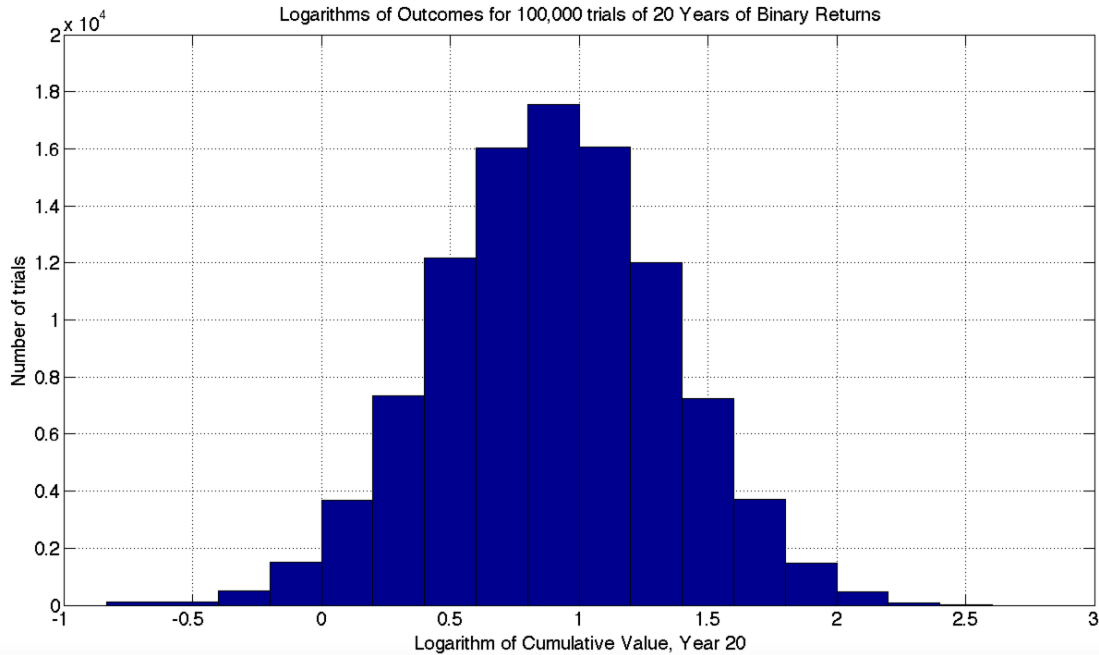
It is time to move at least part way from flipping coins to matters of investment. Consider a game in which you bet \$1.00 to flip a coin. If the coin comes up tails, you receive \$0.95 back; if it comes up heads, you get \$1.15. Clearly, this is an investment with a positive mean return (\$1.05) but some risk (a standard deviation of \$ 0.10). Now, let's assume that you can invest any desired amount in shares (or partial shares) of this investment, each of which costs \$1.00 and returns either \$0.95 or \$1.15, depending on the flip of a coin.

You plan to start with \$1.00, then play the game 20 times, reinvesting the proceeds of each bet every year. How much will you have at the end of 20 plays? The figure below shows the results of a simulation with 100,000 trials.



Notice that this is not a bell-shaped normal distribution. Instead of being symmetric, it is skewed to the right, with a longer tail of large outcomes. Such is the effect of compounding, as we will see.

The figure below is based on the same set of outcomes, but plots the logarithm of each one. Note that it is symmetric and looks very much like a normal distribution. As, in fact, it should.



As you may know, the *natural logarithm* of a number  $y$  is the value  $x$  in the following equation:

$$y = e^x$$

or, in MATLAB:

$$y = e^x$$

where  $e$  is approximately equal to 2.718281828459046 (it has, in fact, been calculated to 869,894,101 decimal places, which is still an approximation). For brevity, we will use the term *log* to signify the natural logarithm of a number. To find the log of a value in Matlab, one can use the **log()** function. Thus:

$$x = \log(y)$$

To reverse the process:

$$y = \exp(x)$$

The value of  $e$ , a crucial ingredient of mathematics theory and practice, was discovered by Jacob Bernoulli, a Swiss mathematician, in the 1600's. Bernoulli was in fact investigating a problem in Finance. Here is a summary (taken from the Wikipedia entry on  $e$ ) :

“An account starts with \$1.00 and pays 100 percent interest per year. If the interest is credited once, at the end of the year, the value of the account at year-end will be \$2.00. What happens if the interest is computed and credited more frequently during the year?

If the interest is credited twice in the year, the interest rate for each 6 months will be 50%, so the initial \$1 is multiplied by 1.5 twice, yielding  $\$1.00 \times 1.5^2 = \$2.25$  at the end of the year. Compounding quarterly yields  $\$1.00 \times 1.25^4 = \$2.4414$ , compounding monthly yields  $\$1.00 \times (1 + 1/12)^{12} = \$2.613035$ . ... If there are  $n$  compounding intervals, the interest for each interval will be  $100\%/n$  and the value at the end of the year will be  $\$1.00 \times (1 + 1/n)^n$ .

The limit as  $n$  grows larger is the number that came to be known as  $e$ ; with continuous compounding, the account value will reach 2.7182818....”

But why the letter  $e$ ? Because of work done by Bernoulli's friend, another Swiss mathematician named Leonard Euler (pronounced “oiler”), who was responsible for many key concepts of calculus, and for the very notion of a *function*.

If the logarithm of a probability distribution of a variable plots as a normal distribution, the variable is said to be *lognormally* distributed, to have a *lognormal distribution*, or simply to be a *lognormal variable*.

Why is this important? Because any variable derived by multiplying a series of values, each of which is drawn from the same distribution will approach a lognormal distribution as the number of draws increases. Consider our assumptions about inflation. We assume that each year's inflation, expressed as a ratio of the cost of living at one point in time divided by the cost of living a year earlier, is drawn from a probability distribution and that the probability distribution is the same each year. Formally, we say that the variable is *independent and identically distributed* – *iid* for short (pronounced “eye eye dee”). Inflation over a long period of many years will thus be very close to lognormally distributed, no matter what the shape of the distribution from which each annual value is drawn.

This result goes back to LaPlace. The log of the product of a series of values will equal the sum of their logs. In principle, this will be exactly true; in our programming languages it might approximately true due to rounding errors, but any difference will be negligible. Here is a simple example from an interactive session:

```
a = 1.05;  
b = 1.10;  
c = exp(log(a) + log(b)) - (a*b)  
c = 0
```

More generally, the central limit theorem holds for the sum of the logarithms of any *iid* process. Since the sum of the logs will be approximately normally distributed, then the product of the original values will be lognormally distributed. Q.E.D. (*Quod Erat Demonstrandum*).

This is a powerful and important result. But we still need to make some sort of assumption about the distribution of one-year inflation. It makes no sense to adapt our coin flipping example by assuming that each year's inflation will take on one of only two possible values. We might assume that each year's inflation is drawn from a normal distribution (which is frequently employed in the financial industry when creating scenarios of possible multi-year outcomes). However, we choose instead to assume that inflation over shorter time periods (say, monthly or even weekly) is *iid* and that as a result, annual inflation is sufficiently close to lognormally distributed that it is reasonable to assume that the distribution is in fact lognormal. Admittedly, this is a bit *ex cathedra*, but as chapter 7 will show, it aligns well with a comparable assumption for the returns on diversified investments, for which there is a stronger rationale.

## The Market Structure

With these preliminaries out of the way, it is time to create some matrices. To house required assumptions and the resulting scenario outcomes, we will use a structure named *market*. As with the client structure, there will be two key functions. The first, *market\_create()* will create a market structure with default parameter values as elements. The second, *market\_process(market,client)*, will use the parameter values in a market structure and the size of the client *pStatesM* matrix to create and add new scenario matrices as elements.

We will build each of these functions in steps. Inflation-related elements will be treated in this chapter, investment-related elements in the next chapter, and valuation elements in the following chapter.

Here is the inflation-related portion of the function used to create a market structure.

```
function market = market_create( )  
    % create a market data structure with default values  
    % cost of living  
    market.eC    = 1.02;    % expected cost of living ratio  
    market.sdC   = 0.01;    % standard deviation of cost of living ratios
```

The first two elements are the parameters indicating the expected value of the year/year cost of living ratio and its standard deviation. The default values conform with the results detailed earlier in this chapter. Of course they can be easily changed in this function or after the structure is created, as in this example:

```
    % create a new market data structure  
    market = market_create( );  
    % reset expected inflation  
    market.eC = 1.03;
```

We will not do this. All the examples in this book will use default assumptions that the expected cost of living ratio is 1.02 and that the standard deviation of the ratio is 0.01.



## Cost of Living Matrices

Now, to make the cost of living matrices. Here is the relevant part of the function that does the job.

```
function market = market_process( market, client )
  % get size for all matrices from client.pStatesM
  [nrows, ncols] = size(client.pStatesM);
  % compute cost of living (inflation) matrix
  u = market.eC;
  v = market.sdC^2;
  b = sqrt(log((v/(u^2)) + 1));
  a = 0.5 *log((u^2)/exp(b^2));
  market.csM = exp( a + b*randn(nrows,ncols) );
  % compute cumulative cost of living (inflation) matrix
  m = cumprod(market.csM,2);
  market.cumCsM = [ones(nrows,1) m(:,1:ncols-1)];
```

Note that the function requires two arguments – a *market* structure and a *client* structure. The first executable statement shows why the latter is needed. The *client.pStateM* matrix has the number of rows (scenarios) and columns (years) needed to cover the life spans of the clients. It is imperative that all other matrices have the same size. As can be seen, the first statement finds the required number of rows (*nrows*) and columns (*ncols*) so that subsequent statements can create the appropriate number of scenarios and annual values.

The next four statements provide year/year ratios of the cost of living for every scenario and year. The first assigns the expected value from the element of the market structure to the variable *u*. The second uses the standard deviation element to compute the variance (standard deviation squared) and assigns it to the variable *v*. The next two statements use standard formulas to compute first the standard deviation (*b*), then the expected value (*a*) of the logarithm of the year/year cost of living ratios. The next line warrants more detailed examination.

Consider first the expression:

```
a + b*randn(nrows,ncols)
```

The *randn* function will produce a matrix with *nrows* rows and *ncols* columns in which each cell contains a value drawn randomly from a normal distribution with a mean of 0 and a standard deviation of 1. Multiplying each value in this matrix by *b* produces a matrix of values with a mean of 0 and a standard deviation of *b*. And adding *a* to each value creates a matrix of values with a mean of *a* and a standard deviation of *b*. At this point, we have a matrix of the logarithms of year/year cost of living ratios. To finish requires only the conversion from logarithms to actual ratios, which is accomplished with the *exp* function. We do this, then put the results in an element containing the scenario matrix *market.csM*:

```
market.csM = exp( a + b*randn(nrows,ncols) );
```

One statement thus produces a great many random samples from the posited lognormal distribution.

Only one task remains. For convenience we would like to have a matrix which shows the ratio of the cost of living at the beginning of each year to that at the beginning of year 1. To start, we multiply all the yearly ratios in a scenario up to and including each year. This cries out for the use of the *cumprod* function. However, its' default mode is to cumulate the products of elements in each column – going down vertically in the matrix. But we need to go horizontally. This could be done by transposing the matrix, using the function, then transposing the result, which would require only two more key strokes. But this would involve needless additional internal manipulations. Happily, the *cumprod* function can be given a second argument indicating the dimension along which it is to operate. If this is omitted or equals 1, the cumulative computations are done column-wise (vertically); if it is set to 2, they are done row-wise, as we need in this case. Thus:

```
m = cumprod( market.csM , 2 );
```

We now have a new matrix of cumulative changes in the cost of living, with one element for each scenario and year. But the results are for the end-year values. To create a matrix of beginning-of-year values we need to start with a column of 1's, then use all but the last of the previously computed values:

```
market.cumCsM = [ ones(nrows,1) m(:,1:ncols-1) ];
```

This completes the procedure for generating possible cost of living changes for our scenarios. Eight lines of code create two matrices, each with millions of values. And they do so quickly. For Bob and Sue's case, with 100,000 scenarios and 57 years, the process took under 0.25 seconds on the author's Macbook Pro!

With inflation in hand, we turn to other elements of the market structure. Investment returns are next.