B6015 – Decision Models
Review Session 4

The primary aims of this review session are:

- To get practice with non-linear problems, as covered in class.
- To understand the basics of least-squares model fitting.
- To solve a first simulation model in Excel, and get used to the simulation features of Risk Solver Platform.
- To review the concept of confidence intervals, as covered in managerial statistics.

■ □ ■

Question 1 (Replicating the S&P 500)

Figure 1 shows the value of the S&P 500 index from January 2011 to November 2011

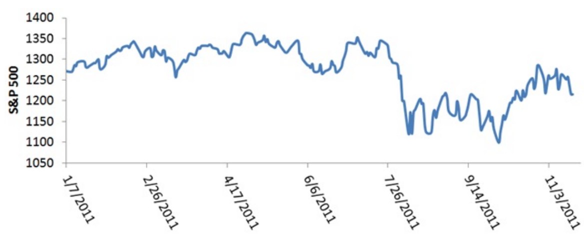

Figure 1: S&P 500 index from January 2011 to November 2011. Daily values at close of market.

Suppose you bought into this index. It might then be of interest to *hedge* the index – in other words, to protect yourself from any drops in the value of your asset. One way to do this is to buy an option – an instrument that gives you the right but not the obligation to sell a given asset at some point in the future at a specified price. If the price of the index were to later drop too low, you'd have the option to sell at the higher price locked-in by the option thus mitigating your losses.

Another approach to hedging is to find a combination of assets (a portfolio) that best replicates the S&P 500 and then to short that combination of assets. Shorting means selling these assets without actually owning them, but then having to 'buy them back' at some point in the future. So for example, if an asset currently has value \$5 and you short it, you get \$5 today as revenues from 'selling' that asset. However, you need the buy the asset back later in the future; if the price of the asset goes *up*, you've lost out because you need to buy it back at a higher price, but if it goes down, you make a profit because the money you'll spend buying it back is less than the money you got for selling it in the first place.

So in this case, shorting assets that replicate the S&P 500 will help you hedge your exposure. Indeed, if the index were to go down (thus losing you money on your S&P 500 holdings), the value of the replicating portfolio would also go down, thus *making* you money on your short sale, offsetting your losses.

In this review session, we'll consider a selection of 10 stocks and then use historical data to find the portfolio based on these 10 stocks that best replicates the S&P 500. (I have chosen ten stocks that paid no dividends over the period considered to reduce complications. As a result, our portfolio will be rather tech-heavy!)

The data we will be using is available in the blank Excel file for this review session. It consists of 11 columns. The first contains the value of the S&P 500 over the historical period considered, and the next 10 contain the value of the 10 stocks under consideration during that period.

Part A

Based on the historical data in the question, find the portfolio containing some or all of the 10 stocks given that best replicates the S&P 500.

To answer this question, you will need some sort of concept of what 'best' replicating the S&P 500 means. We will use a very common method which works as follows

- Consider a given portfolio of these 10 stocks. Find the value of that portfolio at every point in our historical data set.
- Find the difference between the value of that portfolio and the *actual* value of the S&P 500, at every point in our historical data set.
- Square these differences.
- Sum all these differences over every point in our historical data set. This sum is the 'badness' of the combination of stocks – or, more formally, we call it the "squared error loss".

Thus, we want to find the portfolio with the *lowest* squared error loss.

Solution

As usual, let us follow our three steps to solving an optimization problem. First, our decision variables – what decisions do we need to make here? Clearly, all we need to decide is how much of each stock to include in our portfolio. Thus, we have 10 decision variables – x_{GOOG}, x_{GILD}, etc...

Now, to our objective. This problem is unusual in that most of the 'meat' lies in the objective, not in the constraints[1]. Our objective is to minimize the *total* squared error loss. One way to write this is

$$\sum (\text{Value of S\&P 500} - \text{Value of our portfolio})^2$$

The full objective function would take several pages, since it would include a term for every single day in our historical data set. As you'll see, though, it's much easier to express in Excel.

Finally, we move to constraints – and good news, there are none for this problem! Indeed, the problem didn't disallow short-selling of stocks our portfolio, so our decision variables can be negative, and nothing constrains our purchase. We just want the portfolio that *best* replicates the S&P 500 (further investigations could require us to constraint the price or risk or the portfolio, but not here).

[1] This is not so much a function of the problem as a function of our *formulation* of the problem. It turns out that simpler constraints can usually be obtained at the price of a more complex objective, and vice-versa.

■ □ ■

Question 2 (Selling the Mona-Lisa)

As a result of bad publicity from Ban Drown's latest book, the number of visitors to the Louvre in Paris has drastically fallen. To make up for the resulting monetary loss, the Louvre has decided to sell the Mona Lisa. The Louvre's various art experts have valued the painting at \$100 million, and 5 prospective buyers have expressed an interest.

The curator of the Louvre has obtained indications that the prospective buyers would be willing to participate in a sealed-bid auction to determine who will buy the painting and at what price. The museum's consultants suggest the museum should use a Vickrey Auction, in which the highest bidder gets to buy the painting, but pays the price that was bid by the second highest bidder. (So, for example, if buyers 1 through 5 bid \$99m, \$101m, \$102m, \$98m, \$110m, the highest bidder – buyer 5 – would get to buy the painting at a price of \$102m – the second highest bid).

The curator, however, is unsure how much each bidder is going to bid. The Louvre's experts advise that even though most valuations of the painting are around \$100 million, some expects have valued it as low as \$90 million, and some as high as \$105 million. The museum's consultants conclude that each bidder is likely to bid an amount that follows a triangular distribution between \$90m and \$105m, with likeliest value at \$100m. They also conclude that there will not be any correlation between the buyers' bids.

Part A

Set up and run a simulation model to determine an estimate for the expected price the Louvre will receive for its painting.

Solution

This is our first attempt at a *simulation* model. Let's therefore start by spending a few minutes looking at simulation models in general. When we looked at optimization models, our 'road map' was always decision variables first, then the objective, then the constraints. We'd like to develop a similar framework for simulation.

Before we do that, though, what exactly *is* simulation? In a nutshell, it's basically a way to deal with uncertainty in the world. For example, imagine I have a \$5 bet with a friend that my company's stock will be worth at least \$50 tomorrow midday. The money I'll make tomorrow morning is given by

$$\text{Money I'll make} = \begin{cases} \$5 & \text{if the stock price } \geq 50 \\ -\$5 & \text{if the stock price } \leq 50 \end{cases} \tag{1}$$

Unfortunately, I don't know the price of the stock tomorrow midday! However, thanks to my in-depth knowledge of financial engineering, I might know the average/expected price of the stock tomorrow midday (say \$40), and I might know the distribution of that price around the mean[2].

[2] In case you've forgotten from managerial statistics, a quick reminder – the *distribution* of a random variable tells us how likely it is that the random variable will take certain values. For example, the probability distribution might tell us how likely it is the stock price will be roughly \$60 tomorrow.

One option would be to simply ask 'well, what would I win if the stock price came out to be the average'? In this case, the answer would be I'd lose \$5 because the average stock price is less than \$50. This is clearly a pretty awful measure of my expected winnings, though – there are clearly *some* times the stock price will be greater than that and I'll actually win! So to say my 'average' gain is –\$5 seems ridiculous.

The simulation approach to dealing with this issue is simple. All we do is simulate lots and lots of different possible values for the stock price using the probability distribution for the stock price (so if a value of the stock price is particularly likely, we simulate it lots of times; if it's less likely, we simulate it fewer times) and then for each of these values, we look at the resulting money we'd get. We can then look at each resulting value and see how often it occurs – this gives us a *distribution* over resulting values.

All that said, what are the steps to formulating a simulation problem?

Assumptions : The first step to constructing a simulation model is to highlight your *assumptions*. These are the underlying variables in the problem, which ar uncertain. As a rule of thumb, a given quantity is an assumption if other values in the problem depend *on it*, and if it doesn't depend on anything.

So for example – in the problem above, the assumption is the underlying stock price. It doesn't depend on anything else in the problem, and other things (ie: the amount of money I'll be making) depend *on it*.

Assumptions need to be specified together with their distributions. So for example, an assumption cell might have a normal distribution with a given mean and variance, or a binomial distribution with a given value of n and p. (In the problem above, we didn't specify the distribution of the stock price).

Results : The second step is to identify the *results* – these are the quantities that we actually care about and that depend on the assumption cells. So for example, in the problem above, what do we actually care about? The amount of money we make as a result of the bet! Thus, this is our result cell, and it does indeed depend on the assumption cell, in the manner described by equation 1.

Having said that, let's get started on the model in this question.

Assumptions : In this case, what are the underlying sources of uncertainty upon which everything else lies? Answer: the amounts that each of the five bidders is going to bid. *That's* the quantities whose values we're unsure about.

As such, this problem has five assumption variables – the bids that will be entered by each of the five bidders. Each of these variables has a triangular distribution, with a minimum of \$90, a maximum of \$105, and a most likely value of \$100.

Just like in optimization problems, we'll give each of those assumption variables a name – in this case, $x_1, x_2, \ldots, x_5$.

Results : Now in terms of the results – what's the result we actually care about in this case? Answer: the price the winner will pay, which is the second-largest bid entered by the five players. Thus, our result is simply

$$\text{Price winner pays} = \text{Second largest of } (x_1, x_2, x_3, x_4, x_5)$$

And that's it! All that now remains to do is to enter the model into risk premium solver. With 10,000 trials and a random seed of 123, the expected winning bids comes out to be \$100.06m.

■ □ ■