Martingales, Part II, with Exercise Due 9/21

1. Brownian Motion

A process $\{X_t\}$ is a **Brownian Motion** if and only if

i. it is a martingale,
ii. t is a continuous time parameter (i.e. takes on values in the whole real line $\mathbb{R}$ instead of just integer values),
iii. the values of X_t at different times are always jointly normally distributed, and
iv. X has stationary increments (i.e. $\mathrm{Var}(X_{t+s} - X_t)$ has a distribution that depends on s but not t).

Campbell, Lo and MacKinlay (CLM) have a discussion of Brownian motion in their section 9.1.1. We will use just its definition, given above, not the argument CLM give to derive it as a limit. However, the CLM argument may help you understand one fact about martingales and Brownian motion that is important:

> *A martingale with continuous time parameter and stationary increments whose observed time paths with probability one show no discontinuities is a Brownian Motion.*

This give us the surprising conclusion that if changes in prices are not normally distributed, but prices are a martingale with stationary increments, then the time paths of prices must sometimes jump discontinuously.

As CLM show in the first section of assigned reading, price changes on the US stock market are definitely not normally distributed, especially for changes over small time intervals. Price changes are also hard to predict, so that, even though there are detectable differences from martingale behavior in stock prices, a martingale is a pretty good approximation. Put together with the fact about Brownian motion cited above, this means that it is quite likely that we must recognize a non-zero probability of discontinuous jumps in US stock prices.

2. How to Manage a Hedge Fund

The models actually used by Wall Street "rocket scientists" are more complicated than these simple martingale models we are studying, but the models we are studying are similar enough in certain important respects that they can provide some insight into the reasons for the success of their modeling methods in ordinary times and the difficulties that some of them have encountered recently.

Suppose we are managing an investment portfolio and have X dollars of capital to invest. We have identified a single investment opportunity that, we think, has an above-market rate of return. We believe the price of this investment, P_t per share, can be described as a Brownian motion times a non-random growth term that grows at the rate g. That is,

$$P_t = e^{gt} W_t \,, \tag{1}$$

where W is a Brownian motion with variance parameter σ^2. We can borrow money in unlimited amounts at the fixed continuously compounded rate r. That is, if we borrow B dollars at t with a promise to pay it back at $t+s$, we must pay Be^{rs} when the loan comes due.

We could of course just invest our capital X in the risky investment and achieve an expected rate of return of g. But if in addition we borrow an amount B, investing that as well, we can achieve much higher rates of return—at the expense of higher risk. Suppose that we are investing at time $t = 0$ and that the loans we can obtain must be paid back at $t = 1$. We invest our own capital X plus borrowed capital B. The actual net value of what we have at $t = 1$ will be[1]

$$\frac{P_1(X+B)}{P_0} - e^r B = \frac{W_1}{W_0} e^g (X+B) - e^r B \,. \tag{2}$$

The yield of the investment is the ratio of this time-t value to the time-0 investment, less 1, i.e.

$$y = \frac{(W_1/W_0)e^g(X+B) - e^r B}{X} - 1 \,. \tag{3}$$

When we take its expectation conditional on information at $t = 0$, the only random component of the yield (3) is W_1, and that has conditional expectation W_0, so we arrive at

$$E_0 y = \left(e^g + \frac{B}{X}(e^g - e^r) \right) \,. \tag{4}$$

So long as $g > r$, it is clear from (4) that we can make the expected yield as large as we like by making B/X arbitrarily large.

Why would we not then just borrow as much as lenders would lend us, to get astronomically large expected returns? Because we may not want to face a large probability that we could lose our shirts. Portfolio managers these days often use a measure known as "value at risk" (or VAR[2]). We might, for example, decide that we want to face no more than a 5% probability that we will lose more than 40% of our initial investment. This choice of a probability and an amount of loss, followed by an analysis of what investment portfolio will achieve these targets, is VAR analysis.

To calculate what level of B is consistent with our 5% risk of losing 40% of X, our initial capital, we have to calculate not only the expected yield of our investment,

[1] Be sure that you understand how each of the equations and expressions in the remainder of this paragraph was arrived at. Ask questions in class about points you don't understand.

[2] The initials VAR are used by economists to refer to a kind of statistical model (a "vector autoregression"). Financial market participants, and some economists specializing in finance, use VAR mainly to refer to value at risk, as we do here. But vector autoregressions and value at risk are completely unrelated concepts. The overlapping usage can be confusing if you are not forewarned.

Figure 1. **Normal Distribution Preserved Under Linear Combinations**

If $X \sim N(\mu, \sigma^2)$ and a and b are real numbers, then

$$aX + b \sim N(a\mu + b, a^2\sigma^2) \,.$$

If X is an $n \times 1$ vector, with $X \sim N(\mu, \Sigma)$, a is an $n \times 1$ vector of real numbers, and b is a single real number, then

$$a'X + b \sim N\left(a'\mu + b, a'\Sigma a\right) \,.$$

but the entire distribution of the yield, conditional on information at t. Now $W_1 = W_0 + (W_1 - W_0)$, and conditional on information at 0 W_0 is non-random. Furthermore $W_1 - W_0 \sim N(0, \sigma^2)$, independent of the value of W_0, so this is its conditional as well as its unconditional distribution. From (3) we see that the yield is a linear function of the normally distributed W_1 and hence itself normal. (See Figure 1.) To be precise, we have

$$y | \{\text{ info. at } t = 0\} \sim N\left(E_0 y, \frac{\sigma^2}{P_0^2} e^{2g}\left(1 + \frac{B}{X}\right)^2\right) \,. \tag{5}$$

From (5) we have the information we need to do our VAR calculations. To do them, we have to consult a computer or a book that can tell us $P[y < -.4]$, that is, the probability that the yield will be lower than -40%, our target maximum loss. If you have access to a computer with a spreadsheet program it probably can give you the answer directly, once you have given it the mean and standard deviation of the normal distribution for y. In Excel, for example, there is a function `NORMDIST(a,mean,std_dev,cumulative)`, which, when `cumulative` is set to `TRUE` returns $P[y < a]$ for a normally distributed y with the specified mean and standard deviation. (When `cumulative` is set to `FALSE`, it instead returns the value of the normal pdf at $y = a$.)

Some other statistical or spreadsheet programs, statistics textbooks, or even fancy calculators, will give only the cdf (cumulative distribution function) for the "standard" normal distribution, i.e. for a random variable $Z \sim N(0, 1)$. That is, they give you $P[Z < a]$ as a function of a for such a standard-normal Z. In that case you have to use the fact that if $y \sim N(\mu, \sigma^2)$, then $z = (y - \mu)/\sigma \sim N(0, 1)$, so that $P[y < a] = P[z < (a - \mu)/\sigma]$.

Suppose in our example that

$$\frac{\sigma}{P_0} = .1 \qquad g = .12 \qquad r = .08 \,.$$

A plot of the probability of $y < -.4$ as a function of B/X appears in Figure 2. From the lower plot it is clear that if we want to contemplate no more than a 5% probability of losing 40% or more of our capital, we cannot borrow more than about 2.4 times our

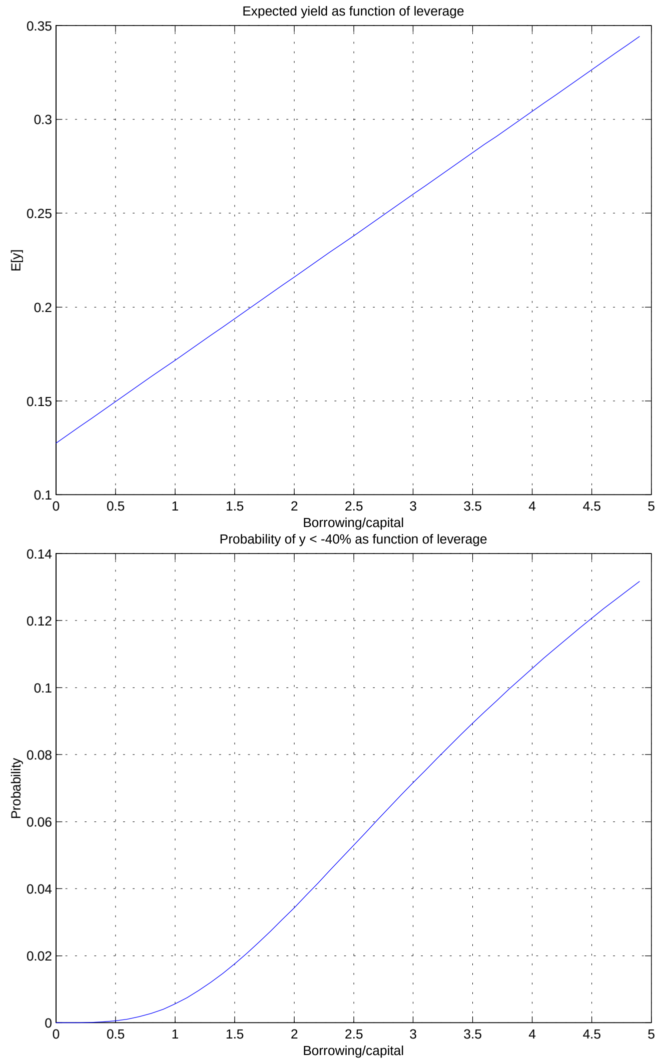

FIGURE 2. VAR Analysis for the Example

capital. Therefore, from the upper plot, we are limited to an expected yield of no more than about 23%.

3. Pitfalls of Risk Management

In order to apply the ideas of the previous section to a real-world problem, one would have to come up with numbers for P_0, r, g, and σ. The first two of these are market data. The latter two have to be estimated. Estimating the true expected rate of return g is difficult and essential to successful investing. However, VAR analysis is only modestly sensitive to the accuracy of this estimate. Much more important is the accuracy of the estimate of σ, which is known as the **volatility** of P. It is one of the

Figure 3. Independence

Two vectors of random variables X and Y are **independent** if and only if the distribution of $X|Y$ does not depend on Y.

If they have a joint pdf $p(x, y)$, they are independent if and only if $p(x, y) = g(x)h(y)$, in which case g and h are both the marginal and the conditional pdf's of x and y.

If they are jointly normally distributed they are independent if and only if their covariance is zero, i.e. $E[(X - E[X])(Y - E[Y])'] = 0$. Saying that they have zero covariance is the same as saying they are **uncorrelated**.

If X and Y are independent, they are always uncorrelated. If they are uncorrelated, this generally does not mean they are independent, unless we know also that they are jointly normally distributed.

remarkable features of the Brownian motion model that it implies that volatility can be estimated with as much precision as may be desired from data over an arbitrarily short time interval—so long as we can sample the data over that interval with arbitrarily high frequency.

The reason is that if P_t is a Brownian motion with variance parameter σ^2, then $(P_{t+\delta} - P_t)\Big/\sqrt{\delta}$ is distributed as $N(0, \sigma^2)$, no matter what the value of δ, and changes in P over non-overlapping intervals are **independent** (see Figure 3). If we are given data on P over an interval (a, b), we can break the interval up into as many non-overlapping pieces as we like, for example n subintervals of length $\delta = (a - b)/n$. If for each of these subintervals we form $(P_{t+\delta} - P_t)^2/\delta$, we have n independent random variables, all with the same expected value σ^2. By the **law of large numbers** (See Figure 4), if we take the average of these random variables we should get something very close to σ^2 if n is large enough. So even if we have only 20 minutes of data, if we can split the time up finely enough, we can estimate volatility as accurately as we like.

This fact is to some extent a reason for the success of quantitative methods of portfolio analysis. Data on asset prices can be matched fairly well by models that are more complicated than Brownian motion, but share with it the property that they imply lots of variation in P at a very fine time scale, and that this variation can be used to estimate the behavior of P at larger time scales. But analysis in which the recent history of an asset price at a fine time scale is used to predict the future, is also an Achilles' heel of quantitative modeling.

The sum of two martingales is also a martingale. Suppose the truth were that P is the sum of a Brownian motion and a Poisson martingale, which you will recall is constant except for randomly timed, discontinuous, jumps. This would be a martingale.

Figure 4. Law of Large Numbers

There are many versions of the law of large numbers, all of which state that the average of n random variables $\{X_i\}$ that all have the same expectation μ tends to get close to μ when n is large enough. A version that covers what is needed in the text can be stated precisely as

> If for each n $\{X_{in}, i = 1, \ldots, n\}$ is a collection of uncorrelated random variables, with $\mathrm{Var}(X_{in}) = \nu^2 < \infty$ and $E[X_{in}] = \mu$ for all i and n, then
>
> $$\frac{1}{n}\sum_{i=1}^{n} X_{in} \xrightarrow[n\to\infty]{\text{q.m.}} \mu\ .$$

If the frequency of jumps in the Poisson component is low, so that the probability of a jump in the time period covered by the data we have available is low, then we might easily have data in which there is no variation reflecting the Poisson component—i.e. no discontinuous jumps. A pure Brownian motion model would explain the data perfectly, even though the truth were that there is some probability of a discontinuous jump in every period of time.

Consider our example of VAR analysis above. Imagine that the true asset price process is, instead of $P_t = e^{gt}W_t$ as we originally assumed, $P_t = e^{gt}(W_t + \theta J_t)$, where J_t is a Poisson jump process with parameter λ, independent of W. In our example we contemplate investment with a loan that has to be paid off after one period. Suppose that we had three periods of data, for $t \in (-3, 0)$, on which to base the model of price behavior we use at $t = 0$ to do VAR analysis. If $\lambda = \frac{1}{3}$, the probability of no jump during an interval of length three is $e^{-1} = .368$. So there is a reasonable probability that we would find no jump in our data. Suppose that $\theta = .2P_0$, so that the size of a jump if it occurs is 20% of the initial asset price. How big a mistake would this imply that we would make by carrying out VAR analysis with the pure Brownian motion model?

Recognizing the Poisson component would have no effect on the expected return, because the Poisson component does not change the martingale property of the random component of the return. It therefore would not affect the upper plot in Figure 2. But the lower plot would be strongly affected. To calculate the full effect exactly is quite a bit of work, because there could be many jumps in the time interval $0, 1$. But the probability of no jumps in the interval is $e^{-1/3} = .717$, and the probability of more than one jump is fairly small. So let's make the approximate assumption that there will be just one jump in $(0, 1)$, of size $\pm .2P_0$, with probability .283. Then there is probability .1417 of a downward jump, probability .1417 of an upward jump, and probability .717 of no jump. We can analyze each these three cases separately just as we did the

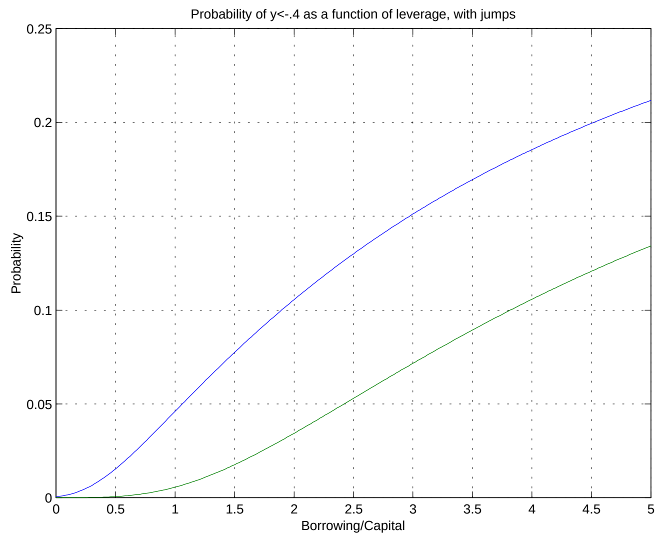

FIGURE 5. VAR Analysis for the Example, with and without Jump

original no-jump situation, then add up the probabilities of $y < -.4$ from the three cases, weighting by the probabilities of the cases themselves.

For example for the case where a downward jump occurs, the mean of y conditional on the downward jump is

$$E[y|\,\{\text{info at } t = 0 \textit{ and } \text{down jump}\}] = .8e^g + \frac{B}{X}(.8e^g - e^r)\,. \tag{6}$$

Comparing this to (4), we see that the term e^g has simply been replaced by $.9e^g$. This then replaces the mean on the right-hand-side of (5), while the variance remains unchanged. (The overall variance of the yield has gone up, but *conditional* on the size of the jump, the only remaining randomness is in the Brownian motion part, which is unchanged.) We make a similar calculation for the expected yield conditional on an upward jump, and of course in the case of no jump the conditional distribution is just what we had in the pure Brownian motion model. Combining the VAR calculations from the three cases, with the appropriate probability weights, gives us the graph in Figure 5. Clearly our notion of the maximum safe amount of borrowing would be quite different if we recognized the true probability of a jump. To achieve the same probability .05 of a 40% loss as in the no-jump case, we would have to limit B/X to around 1, instead of 2.4.

4. EXERCISE

Suppose that an asset price P satisfies the model

$$\log(P_t) = p_t = gt + W_t + \theta J_t\,, \tag{7}$$

where W is a Brownian motion with $\sigma = .2$ and J is a jump process that will make a single jump of size ± 1, with mean zero, and with the cdf of the time until the jump given by $1 - e^{\lambda t}$, with $\lambda = .2$. [In the example worked out in the text, the jump process is assumed to be a Poisson process, but the approximate calculations are performed as if the process were the simpler type we are assuming here.] Plot the probability of losing 20% or more of initial capital over one year as a function of borrowing/capital for both the case $\theta = .1$ and the case $\theta = 0$ (i.e., no jumps). Assume $g = .12$ and assume that the borrowing interest rate is $r = .08$. The plots can be sketches, based on four or five points, if they give a reasonably good idea of where the probability hits .05. If you have constructed plots with a spreadsheet or other computer software before, though, making the plots with a computer is probably faster than computing points one at a time and "sketching".