Investments Notes

John H. Cochrane[1]

November 16, 2006

Contents

Chapter 1

Preface

These notes compactly summarize some of the theory and background for investments classes. They are not a complete treatment, but focus on areas in which I feel the textbooks and other materials leave gaps. I also tried to write them as compactly as possible rather than develop things slowly as in textbooks.

Please let me know about any typos or other comments. The latest version of this document is always available on the class website.

Chapter 2

Notation and return definitions

Don't get hung up on notation. Understanding the concepts is the important thing here. Don't memorize these formulas. If you understand the concepts, you can invent your own notation and get it right. If you don't understand the concepts, no amount of staring at superscripts and subscripts is going to help.

Time notation Where necessary, I'll use time subscripts to denote when things happen. For example, the price at the end of 1996 is P_{1996}. More generally, the price at time t is P_t, the interest rate at time t is R_t etc. When I want to be excruciatingly clear about the timing of a return, I'll denote the rate of return from time t to time $t+1$ as $R_{t \to t+1}$. "Today" is often "time 0."

Returns I use capital R to denote a *gross return*, e.g.

$$R = \frac{\$\text{back}}{\$\text{paid}}.$$

For a stock that pays a dividend D, the gross return is

$$R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t} = \frac{\$\text{back}_{t+1}}{\$\text{paid}_t} \quad \text{(for example, 1.10)}$$

R is a number like 1.10 for a 10% return.

Several other units for returns are convenient. The *net* return is

$$r_{t+1} = R_{t+1} - 1 \text{ (For example, 0.10).}$$

The *percent* return is

$$100 \times r_{t+1} \text{ (For example, 10\%)}$$

The *log* or *continuously compounded* return is

$$r_t = \ln R_t \text{ (For example, } \ln(1.10) = 0.09531 \text{ or } 9.531\%)$$

The *real return* corrects for inflation,

$$R_{t+1}^{\text{real}} = \frac{\text{Goods back}_{t+1}}{\text{Goods paid}_t}.$$

The consumer price index is defined as

$$CPI_t \equiv \frac{\$_t}{\text{Goods}_t}; \;\; \Pi_{t+1} \equiv \frac{CPI_{t+1}}{CPI_t}$$

Thus, we can use CPI data to find real returns as follows.

$$R_{t+1}^{\text{real}} = \frac{\$_{t+1} \times \frac{\text{Goods}_{t+1}}{\$_{t+1}}}{\$_t \times \frac{\text{Goods}_t}{\$_t}} = \frac{\$_{t+1} \frac{1}{CPI_{t+1}}}{\$_t \frac{1}{CPI_t}} = R_{t+1}^{\text{nomial}} \frac{CPI_t}{CPI_{t+1}} = \frac{R_{t+1}^{\text{nominal}}}{\Pi_{t+1}}.$$

I.e., divide the gross nominal return by the gross inflation rate to get the gross real return.

You're probably used to *subtracting* inflation from nominal returns. This is exactly true for *log* returns. Since

$$\ln(A/B) = \ln A - \ln B,$$

we have

$$\ln R_{t+1}^{\text{real}} = \ln R_{t+1}^{\text{nomial}} - \ln \Pi_{t+1}.$$

For example, 10%-5% = 5%. It is approximately true that you can subtract net returns this way,

$$\frac{R_{t+1}^{\text{nominal}}}{\Pi_{t+1}} = \frac{\left(1 + r^{nomial}\right)}{1 + \pi} \approx 1 + r^{nom} - \pi.$$

The approximation is ok for low inflation (10%) or less, but really bad for 100% or more inflation.

Using the same idea as for real returns, you can find *dollar* returns of international securities. Suppose you have a German security, that pays a gross Euro return

$$R_{t+1}^{DM} = \frac{E \text{ back}_{t+1}}{E \text{ paid}_t}$$

Then change the units to dollar returns just like you did for real returns. The exchange rate is defined as

$$e_t^{\$/DM} = \frac{\$_t}{E_t}.$$

Thus,

$$R_{t+1}^{\$} = \frac{\$_{t+1}}{\$_t} = \frac{E_{t+1}}{E_t} \times \frac{\$_{t+1}/E_{t+1}}{\$_t/E_t} = R_{t+1}^{E} \times \frac{e_{t+1}^{\$/E}}{e_t^{\$/E}}.$$

Compound returns Suppose you hold an instrument that pays 10% per year for 10 years. What do you get for a \$1 investment? The answer is not \$2, since you get "interest on the interest." The right answer is the *compound return.* Denote

$$V_t = \text{value at time t}$$

Then

$$V_1 = RV_0 = (1+r)\,V_0$$

$$V_2 = R \times (RV_0) = R^2 V_0$$

$$V_T = R^T V_0$$

Thus, R^T is the *compound return.*

As you can see, it's not obvious what the answer to 10 years at 10% is. Here is why log returns are so convenient. Logs have the property that

$$\ln(ab) = \ln a + \ln b;\ \ln\left(a^2\right) = 2\ln a.$$

Thus

$$\ln V_1 = \ln R + \ln V_0$$

$$\ln V_T = T \ln R + \ln V_0$$

Thus the compound *log* return is T times the one-period log return.

More generally, log returns are really handy for multi-period problems. The T period return is

$$R_1 R_2 ... R_T$$

while the T period log return is

$$\ln(R_1 R_2 ... R_T) = \ln(R_1) + \ln(R_2) + ... \ln(R_T)$$

Continuously compounded or log returns are also convenient because you can subtract rather than divide to get exact real returns or exchange rate conversions.

$$R^{\text{real}} = \frac{R^{\text{nominal}}}{\Pi} \Rightarrow \ln(R^{\text{real}}) = \ln R^{\text{nominal}} - \ln \Pi.$$

Within period compounding This is best explained by example. Suppose a bond that pays 10% is compounded semiannually, i.e. two payments of 5% are made at 6 month intervals. Then the total annual gross return is

$$\text{compounded semi-annually:} \quad (1.05)\,(1.05) = 1.1025 = 10.25\%$$

What if it is compounded quarterly? Then you get

$$\text{compounded quarterly:} \quad (1.025)^4 = 1.1038 = 10.38\%$$

Continuing this way,

$$\text{compounded N times:} \left(1 + \frac{r}{N}\right)^N$$

What if you go all the way and compound *continuously?* Then you get

$$\lim_{N\to\infty} \left(1 + \frac{r}{N}\right)^N = 1 + r + \frac{1}{2}r^2 + \frac{1}{3\times 2}r^3 ... = e^r.$$

Well, if the gross return is $R = e^r$, then we can find the *continuously compounded* or *log return* as $r = \ln R$. For example a stated rate of 10%, continuously compounded, is really a gross return of $e^{0.10} = 1.1051709 = 10.517\%$. Conversely, given a gross return of 10.517%, you can express it as a continuously compounded return of 10%.

Both kinds of compounding What is the three year return of a security that pays a stated rate R, compounded semiannually? Well, again with $r = R - 1$, it must be

$$\left(1 + \frac{r}{2}\right)^{2\times 3}.$$

Similarly, the continuously compounded T year return is

$$e^{rT}.$$

Chapter 3

Probability and statistics

3.1 Probability

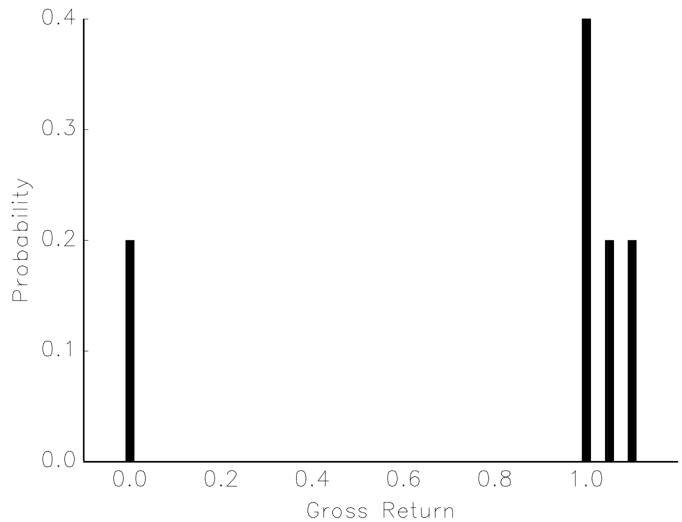
3.1.1 Random variables

We model stock returns as *random variables.* A random variable can take on one of many values, with an associated probability. For example, the gross return on a stock might be one of four values.

$$R = \begin{array}{cc} \text{Value} & \text{Probability} \\ 1.1 & 1/5 \\ 1.05 & 1/5 \\ 1.00 & 2/5 \\ 0.00 & 1/5 \end{array}$$

Each value is a possible *realization* of the random variable. Of course, stock returns can typically take on a much wider range of values, but the idea is the same. Many finance texts distinguish the *random variable* from its *realization* by using $\tilde{R}$ for the random variable and R for the realization. I don't.

The *distribution* of the random variable is a listing of the values it can take on along with their probabilities. For example, the distribution of return in the above example is



(Real statisticians call this the *density* and reserve the word *distribution* for the *cumulative distribution*, a plot of values vs. the probability that the random variable is at or below that value.)

A deeper way to think of a random variable is a *function.* It maps "states of the world" into real numbers. The above example might really be

	Value	State of the world	Probability
	1.1	New product works, competitor burns down	1/5
$R =$	1.05	New product works, competitor ok.	1/5
	1.00	Only old products work.	2/5
	0.00	Factory burns down, no insurance.	1/5

The probability really describes the external events that define the state of the world. However, we usually can't name those events, so we just think about the probability that the stock return takes on various values.

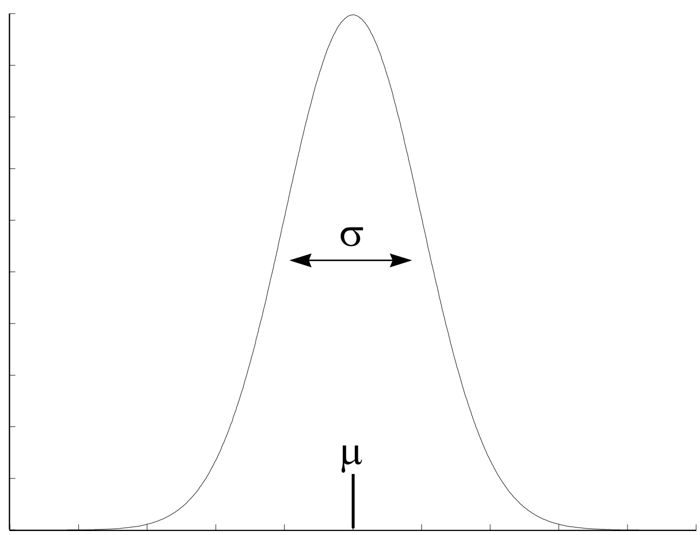
In the end, all random variables have a discrete number of values, as in this example. Stock prices are only listed to 1/8 dollar, all payments are rounded to the nearest cent, computers can't distinguish numbers less than

10^{-300} or so apart. However, we often think of *continuous* random variables, that can be any real number. Corresponding to the discrete probabilities above, we now have a continuous probability *density*, usually denoted $f(R)$. The density tells you the probability per unit of R; $f(R_0)\,\Delta R$ tells you the probability that the random variable R lies between R_0 and $R_0 + \Delta R$.

A common assumption is that returns (or log returns) are *normally distributed.* This means that the density is given by a specific function,

$$f(R) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(R-\mu)^2}{2\sigma^2}\right].$$

The graph of this function looks like this:



About 30% (really 31.73%) of the probability of a normal distribution is more than one standard deviation from the mean and about 5% is more than two standard deviations from the mean (really 4.55%, the 5% probability line is at 1.96 standard deviations). That means that there is only one chance in 20 of seeing a value more than two standard deviations from the mean of a normal distribution. Stock returns have "fat tails" in that they are slightly more likely to take on extreme values than the normal distribution would predict.

3.1.2 Moments

Rather than plot whole distributions, we usually summarize the behavior of a random variable by a few *moments* such as the mean and variance.

I'll denote the values that R can take on as R_i with associated probabilities π_i. Then the *mean* is defined as

$$\text{Mean: } E(R) = \sum_{\text{possible values i}} \pi_i R_i.$$

The mean is a *measure of central tendency*, it tells you where R is "on average." A high mean stock return is obviously a good thing!

The *variance* is defined as

$$\text{Variance: } \sigma^2(R) = E\left[(R - E(R))^2\right] = \sum_i \pi_i \left[R_i - E(R)\right]^2$$

Since squares of negative as well as positive numbers are positive, variance tells you how far away from the mean R typically is. It measures the spread of the distribution. High variance is not a good thing; it will be one of our measures of *risk.*

The *covariance* is

$$\begin{aligned}\text{Covariance: } cov\left(R^a, R^b\right) &= E\left[\left(R^a - E(R^a)\right)\left(R^b - E\left(R^b\right)\right)\right] \\ &= \sum_i \pi_i \left[R_i^a - E(R^a)\right]\left[R_i^b - E\left(R^b\right)\right]\end{aligned}$$

It measures the tendency of two returns to move together. It's positive if they typically move in the same direction, negative if one tends to go down when the other goes up, and zero if there is no tendency for one to be high or low when the other is high.

The size of the covariance depends on the units of measurement. For example, if we measure one return in cents, the covariance goes up by a factor of 100, even though the tendency of the two returns to move together hasn't changed. The *correlation coefficient* resolves this problem.

$$\text{Correlation: } corr\left(R^a, R^b\right) = \rho = \frac{cov\left(R^a, R^b\right)}{\sigma(R^a)\,\sigma\left(R^b\right)}.$$

The correlation coefficient is always between -1 and 1.

For continuously valued random variables, the sums become integrals. For example, the mean is

$$E(R) = \int R\, f(R)\, dR.$$

The normal distribution defined above has the property that the mean equals the parameter μ, and the variance equals the parameter σ^2. (To show this, you have to do the integral.)

3.1.3 Moments of combinations

We will soon have to do a lot of manipulation of random variables. For example, we soon will want to know what is the mean and standard deviation of a *portfolio* of two returns. The basic results are

1) Constants come out of expectations and expectations of sums are equal to sums of expectations. If c and d are numbers,

$$E\left(cR^a\right) = cE\left(R^a\right)$$

$$E\left(R^a + R^b\right) = E\left(R^a\right) + E\left(R^b\right)$$

or, more generally,

$$E\left(cR^a + dR^b\right) = cE\left(R^a\right) + dE\left(R^b\right).$$

2) Variance of sums works like taking a square,

$$var\left(cR^a + dR^b\right) = c^2 var\left(R^a\right) + d^2 var\left(R^b\right) + 2cd\, cov\left(R^a, R^b\right).$$

3) Covariances work linearly

$$cov\left(cR^a, dR^b\right) = cd\, cov\left(R^a, R^b\right)$$

To derive any of these or related rules, just go back to the definitions. For example,

$$E\left(cR^a\right) = \sum_i \pi_i c R_i^a = c \sum_i \pi_i R_i^a = cE\left(R^a\right).$$

3.1.4 Normal distributions.

Normal distributions have an extra property. Linear combinations of normally distributed random variables are again normally distributed. Precisely, if R^a and R^b are normally distributed, and

$$R^p = cR^a + dR^b$$

then, R^p is also normally distributed with the mean and variance given above.

3.1.5 Lognormal distributions

A variable R is *lognormally* distributed if $r \equiv \ln(R)$ is normally distributed. This is a nice model for stock and bond returns since you can never lose more than all your money; we can never see $R < 0$. A lognormal captures that fact. A normal distribution always includes events in which $R < 0$. Lognormal returns are like log returns, useful for handling multiperiod problems.

Since $R = e^{\ln R} = e^r$ by definition, wouldn't it be nice if $E(R) = e^{E(r)}$? Of course, that isn't true because $E[f(x)] \neq f[E(x)]$. But something close to it is true. By working out the integral definition of mean and variance, you can show that

$$E(R) = e^{E(r)+\sigma^2(r)/2}.$$

The variance is a little trickier. $R^2 = e^{2r}$ so this is also lognormally distributed. Then

$$\sigma^2(R) = e^{2E(r)+\sigma^2(r)} \left[e^{\sigma^2(r)} - 1 \right].$$

(To show this,

$$\sigma^2(R) = E(R^2) - E(R)^2 = e^{2E(r)+2\sigma^2(r)} - e^{2E(r)+\sigma^2(r)}.)$$

As a linear combination of normals is normal, a product of lognormals (raised to powers) is lognormal. For example,

$$R_1 R_2 = e^{r_1+r_2};$$

since r_1 and r_2 are normal so is $r_1 + r_2$, and therefore $R_1 R_2$ is lognormal.

3.2 Statistics

3.2.1 Sample mean and variance

What if you don't know the probabilities? Then you have to *estimate* them from a *sample*. Similarly, if you don't know the mean, variance, regression coefficient, etc., you have to estimate them as well. That's what *statistics* is all about.

The *average* or *sample mean* is

$$\bar{R} = \frac{1}{T} \sum_{t=1}^{T} R_t$$

where $\{R_0, R_1, ...R_t, ...R_T\}$ is a *sample* of data on a stock return. Just to be confusing, many people use μ for sample as well as population mean. Sometimes people use hats, $\hat{\mu}$ to distinguish estimates or sample quantities from true population quantities.

Keep the *sample mean* and the true, or *population* mean separate in your head. For example, the true probabilities that a coin will land heads or tails is 1/2, so the mean of a bet on a coin toss ($1 for heads, -$1 for tails) is 0. A *sample* of coin tosses might be {H,T,T,H,H}. In that *sample*, the frequency of heads is 3/5 and tails 2/5, so the *sample mean* of a coin toss bet is 1/5.

Obviously, as the sample gets bigger and bigger, the *sample mean* will get closer and closer to the true or *population mean.* That property of the sample mean (*consistency)* makes it a good estimator. But the sample and population mean are not the same thing for any finite sample!

Also, sample means approach population means *only* if you are repeatedly doing the same thing, such as tossing the same coin. This may not be true for stocks. If there are days when expected returns are high and days when they are low, then the average return will not necessarily recover either expected return.

The *sample variance* is

$$s^2 = \hat{\sigma}^2 = \frac{1}{T-1} \sum_{t=1}^{T} \left[R_t - \bar{R}\right]^2 .$$

Sample values of the other moments are defined similarly, as obvious analogs of their population definitions.

3.2.2 Variation of sample moments

The sample mean and sample variance vary from sample to sample. Let's flip a coin, with Heads = +1, Tails = -1. If I get {H,T,T,H,H}, the sample mean is 1/5, but if I happened to get {T,T,H,T,T}, the sample mean would be -3/5. The true, population mean, is zero of course. Thus the sample mean, standard deviation, and other statistics are also *random variables*; they vary from sample to sample. They are random variables that depend on the whole sample, not just what happened one day, but they are random variables nonetheless. The population mean and variance, by contrast are just numbers.

We can then ask, "how much does the sample mean (or other statistic) vary from sample to sample?" This is an interesting question. If a mutual fund manager tells you "my mean return for the last five years was 20% and the S&P500 was 10%" you want to know if that was just due to chance, or if it means that his true, population mean, which you are likely to earn in the *next* 5 years, is also 10% more than the S&P500. In other words, was the *realization* of the random variable called "my estimate of manager A's mean return" near the mean of the true or population mean of the random variable "manager A's return?"

Figuring out the variation of the sample mean is a good use of our formulas for means and variances of sums. The sample mean is

$$\bar{R} = \frac{1}{T}\sum_{t=1}^{T} R_t.$$

Therefore,

$$E\left(\bar{R}\right) = \frac{1}{T}\sum_{t=1}^{T} E\left(R_t\right) = E\left(R\right)$$

assuming all the $R_t's$ are drawn from the same distribution (a crucially important assumption). This verifies that the sample mean is *unbiased.* On average, across many samples, the sample mean will reveal the true mean.

The variance of the sample mean is

$$\sigma^2\left(\bar{R}\right) = \sigma^2\left(\frac{1}{T}\sum_{t=1}^{T} R_t\right) = \frac{1}{T^2}\sum_{t=1}^{T}\sigma^2\left(R_t\right) + (\text{covariance terms})$$

If we assume that all the covariances are zero, we get the familiar formula

$$\sigma^2\left(\bar{R}\right) = \frac{\sigma^2\left(R\right)}{T}$$

or

$$\sigma\left(\bar{R}\right) = \frac{\sigma\left(R\right)}{\sqrt{T}}.$$

For stock returns, $cov\left(R_t, R_{t+1}\right) = 0$ is a pretty good assumption. It's a great assumption for coin tosses: seeing heads this time makes it no more likely that you'll see heads next time. For other variables, it isn't such a good assumption, so you shouldn't use this formula.

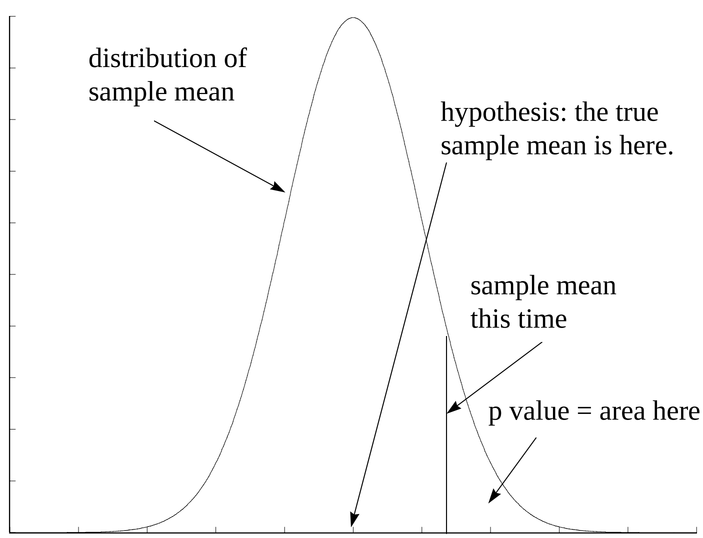
You don't know σ. Well, you can *estimate* the sampling variation of the sample mean by using your estimate of σ, namely the sample standard deviation. Using hats to denote estimates,

$$\hat{\sigma}\left(\bar{R}\right) = \frac{\hat{\sigma}\left(R\right)}{\sqrt{T}}.$$

The classic use of this formula is to give a standard error or measure of uncertainty of the sample mean, and to test whether the sample mean is equal to some value, usually zero.

The test is usually based on a *confidence interval*. Assuming normal distributions, the confidence interval for the mean is the sample mean plus or minus 2 (well, 1.96) standard errors. The meaning of this interval is that if the true mean was outside the interval, there would be less than a 5% chance of seeing a sample mean as high (or low) as the one we actually see.

Now that we have computers, there is an easier method. We can just calculate the probability that the sample mean comes out at its actual value (or larger) given the null hypothesis, i.e., calculate the area under the distribution of the sample mean past the sample mean we happen to see, given an assumption (hypothesis) about what the true mean is. This is called the *p-value* or *probability value.*



Usually, tests are run using the *t-distribution*. When you take account of sampling variation in $\hat{\sigma}$, you can show that the ratio

$$\sqrt{T}\frac{\bar{R}-E\left(R\right)}{\hat{\sigma}}$$

is not a normal distribution with mean zero and variance 1, but a *t distri-*

bution.

3.3 Regressions

We will run regressions, for example of a return on the market return,

$$R_t = \alpha + \beta R_{m,t} + \epsilon_t;\ t = 1, 2...T$$

and sometimes multiple regressions of returns on the returns of several portfolios

$$R_t = \alpha + \beta R_{m,t} + \gamma R_{p,t} + \epsilon_t;\ t = 1, 2...T.$$

The generic form is usually written

$$y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + ... + \epsilon_t;\ \ t = 1, 2, ...T$$

Both textbooks and regression packages give standard formulas for estimates of the regression coefficients β_i and standard errors with which you can construct hypothesis tests. All of these numbers are based on assumptions, most of which are wrong for any given regression. Hence, it's important to know what the assumptions are and which complications you have to correct for.

Several important facts about regressions:

1) The population value of a single regression coefficient is[1]

$$\beta = \frac{cov\,(y, x)}{var\,(x)}.$$

2) The regression recovers the true β (precisely, the estimate of β is *unbiased*) only if the error term is uncorrelated with the right hand variables. For example, suppose you run a regression

$$\text{sales} = \alpha + \beta \text{ advertising expenses} + \epsilon.$$

[1]If you forgot why, start with

$$y_t = \alpha + \beta x_t + \epsilon_t$$

and the usual assumption that errors are uncorrelated with right hand variables $E(\varepsilon_t) = 0, E(\varepsilon_t x_t) = 0$. Multiply both sides by $x_t - E\,(x_t)$ and take expectations, which gives you

$$cov\,(x_t, y_t) = \beta var\,(x_t)\,.$$

Discounts also help sales, so discounts are part of the error term. If advertising campaigns happen at the same time as discounts, then the coefficient on advertising will pick up the effects of discounts on sales.

3) In a multiple regression, β_1 captures the effect on y of only movements in x_1 that are not correlated with movements in x_2. If you run a regression of price of shoes on sales of right shoes and left shoes, the coefficient on right shoes only captures what happens to price when right shoe sales go up and left shoe sales don't. I.e., it doesn't mean much.

3.3.1 Regression formulas with matrices

Here are the standard regression formulas in matrix notation.

The linear regression model is

$$Y = X\beta + \varepsilon$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \\ \vdots & \vdots \\ x_{T1} & x_{T2} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

If there is a constant in the regression, the first column of X is all 1s.

OLS regressions

The OLS estimate is

$$\hat{\beta} = (X'X)^{-1}X'Y$$

(It is usually a bad practice to program it this way, i.e. beta = inv(X'*X)*X'*Y, since inversion is not that stable. In matlab, the command beta = X\Y does the same thing but it is better numerically.)

Standard errors measure the variability of $\hat{\beta}$ over samples, i.e., if you redraw all the data and run the experiment over and over again. The standard formula is

$$\sigma^2(\hat{\beta}) = (X'X)^{-1}\sigma_\varepsilon^2 \tag{3.1}$$

This formula holds *IF* the errors all have the same variance and are uncorrelated with each other.

We usually estimate this quantity by first finding the errors

$$e = Y - X\hat{\beta}$$

and forming

$$\begin{aligned} s^2 &= var(e) \\ \sigma^2(\hat{\beta}) &= (X'X)^{-1}s^2 \end{aligned}$$

(Some books like to use $1/(T-K)$ where K is the number of right hand variables to form s^2.) With the errors, we can also calculate the

$$R^2 = \frac{var(X\beta)}{var(Y)} = 1 - \frac{var(e)}{var(Y)}.$$

What appends if the errors have different variances or are correlated with each other? Remember how we derive the mean and variance of $\hat{\beta}$

$$\begin{aligned} \sigma^2(\hat{\beta}) &= \sigma^2\left[(X'X)^{-1}X'Y\right] \\ &= \sigma^2\left[(X'X)^{-1}X'\left(X\beta+\varepsilon\right)\right] \\ &= \sigma^2\left[(X'X)^{-1}X'\varepsilon\right] \\ &= E\left[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right] \end{aligned}$$

The last line uses the assumption $E(\varepsilon) = 0$ and the formula $var(Ax) = Acov(x, x')A'$. Let

$$\Omega = E(\varepsilon\varepsilon').$$

Then

$$\sigma^2(\hat{\beta}) = (X'X)^{-1}X'\Omega X(X'X)^{-1} \tag{3.2}$$

IF $\sigma^2(\varepsilon_t)$ are the same for all t and $\sigma(\varepsilon_t\varepsilon_s) = 0$ uncorrelated, then Ω is diagonal

$$E(\varepsilon\varepsilon') = \Omega = \sigma_\varepsilon^2 I$$

Then the standard error formula is our usual friend (3.1) *Otherwise* use the more general formula (3.2).

GLS regressions

GLS estimate. If $\Omega \neq \sigma^2 I$, the following "GLS estimate" is more "efficient."

$$\hat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y$$

The GLS procedure is: 1) run OLS to find ε 2) Use ε to estimate Ω 3) run GLS.

The standard errors of the GLS estimate are

$$\sigma\left(\hat{\beta}_{GLS}\right) = (X'\Omega^{-1}X)^{-1}$$

Why?

$$\begin{aligned}
\sigma^2(\hat{\beta}_{GLS}) &= \sigma^2\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}Y\right] \\
&= \sigma^2\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\left(X\beta+\varepsilon\right)\right] \\
&= \sigma^2\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon\right] \\
&= E\left[(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}\varepsilon\varepsilon'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\right] \quad (3.3)
\end{aligned}$$

OLS vs. GLS

When the errors don't satisfy the OLS assumption $\Omega = \sigma_\varepsilon^2 I$, GLS is advocated by econometrics text books, because it's more "efficient." This means that if all goes well, $\sigma(\hat{\beta}_{GLS}) < \sigma(\hat{\beta}_{OLS})$ for large T – the estimate is closer to the true value.

However, in practice all may not go well. *It is quite common in finance to use OLS estimates anyway.* OLS is still *unbiased*[2] $E(\hat{\beta}_{OLS}) = \beta$. It is potentially "inefficient," but often squeezing the last drop of efficiency out of the data is not that vital. On the other side, experience has proved that putting an Ω^{-1} matrix in the middle can really screw things up, especially if (as always) the model is pretty good but not perfect.

Just because we use OLS *estimates* does not mean that we use the OLS *standard error formula*. If $\Omega \neq \sigma_\varepsilon^2 I$, the OLS standard error formula is biased, and often is too optimistic by a factor of 5-10 in typical finance data sets. Thus, if you use OLS *estimates*, it's vital to use the general formula (3.2) or some equivalent procedure (Fama-MacBeth).

A typical example is a "panel data" regression taken over N companies and T time periods. If the data are returns, errors are usually decently uncorrelated over time, but if company i is unusually high at time t, company j is also likely to be unusually high. Thus, the errors are correlated across companies. In corporate finance data sets, when the data might be investment, cashflows, etc., the errors are likely to be correlated across *time* as well.

---

[2]

$$\begin{aligned}
\hat{\beta} &= \left(X'X\right)^{-1}X'Y \\
&= \left(X'X\right)^{-1}X'\left(X\beta+\varepsilon\right)
\end{aligned}$$

$$\begin{aligned}
E(\hat{\beta}) &= E\left[(X'X)^{-1}\left(X\beta+\varepsilon\right)\right] \\
&= \beta + E\left[(X'X)^{-1}X'\varepsilon\right]
\end{aligned}$$

Thus if the error is uncorrelated with the right hand variables $E(X\varepsilon) = 0$ we see that OLS is unbiased (consistent when X is stochastic.)

Don't confuse *OLS standard errors when* $\Omega \neq \sigma_\varepsilon^2 I$, (3.2), with *Standard error of GLS estimate* (3.3) The first gives the (larger) standard errors of the OLS estimate, the second gives the standard errors of the GLS estimate.

3.4 Time series

Time series is the name for the set of statistical tools and models we use to think about asset prices, returns and so forth. A time series is a set of repeated observations of a random variable. A repeated coin toss, or the temperature in Chicago are other examples worth thinking about. We denote a time series $x_1, x_2 ... x_t, ...$ meaning the observation at time 1, time 2, a representative time t, and so on.

Unconditional and Conditional Mean and Variance

Time series have a mean and variance, like any other random variable, denoted $E(x_t)$ and $\sigma^2(x_t)$. For example, the mean annual stock return is about 8% with variance about 16%; the mean of a bet on a coin flip is zero.

Some time series move slowly over time; if they are high today it's a good bet they are high tomorrow; they decay slowly back after a shock. Temperature, price/earnings ratios, and interest rates act this way. Other time series are less predictable; the fact that they are high today doesn't give much information about whether they'll be high tomorrow. Stock returns and the series of coin flips are good examples. Other properties may persist over time as well. The *level* of stock returns may not persist much, but the *variance* does; if this week was very volatile, next week is likely to be volatile as well.

These properties are captured by the *conditional mean and variance.* Using all information at time t, what is the mean of x_{t+1}? We denote this $E_t(x_{t+1})$, or if you really want to be clear, $E(x_{t+1}|I_t)$ where I_t represents all information available at time t. We call the regular mean and variance the *unconditional* mean and variance when we want to clearly distinguish which mean we're talking about. We can think about conditional means two three or more steps ahead, $E_t(x_{t+j})$. The persistent time series, like temperature, price/earnings ratio, or interest rate, have the property that the sequence of conditional means $E_t(x_{t+j})$ falls slowly back to the *unconditional mean* $E(x_t)$. For a coin flip, the conditional mean is always the same and equal to the unconditional mean – today's flip gives you no information about tomorrow's flip.

White noise; MA models

Time series models capture this different behavior of conditional means and variances. The basic building block of time series models is the *white noise process* usually denoted ε_t. This is like the coin flip, completely unpredictable over time. In addition, it's convenient to start with a mean zero, so $E_t(\varepsilon_{t+1}) = E(\varepsilon_{t+1}) = 0$. We usually specify a constant conditional variance $\sigma_t^2(\varepsilon_{t+1}) = \sigma^2(\varepsilon_{t+1}) = \sigma_\varepsilon^2$ as well. Since they are unpredictable from any information, they are also unpredictable from their own past. Thus the *autocorrelation* of the white noise process is zero, $corr(\varepsilon_t, \varepsilon_{t+j}) = 0; corr(\varepsilon_t, \varepsilon_{t-j}) = 0$, and since the mean is zero we can also write this as $E(\varepsilon_t \varepsilon_{t-j}) = 0$, $E(\varepsilon_t \varepsilon_{t+j}) = 0$.

We then build up models with more interesting dynamics from the white noise process. The most basic example are the MA processes. MA stands for "moving average." The MA(1) process is

$$x_t = \varepsilon_t + \theta\varepsilon_{t-1}$$

(1) means how many terms. The $MA(2)$ process is

$$x_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}.$$

The MA(1) process captures the sense in which the conditional mean might be different from the unconditional mean – it has some *persistence.*

$$\begin{aligned} E_t(x_{t+1}) &= E_t(\varepsilon_{t+1} + \theta\varepsilon_t) = \theta\varepsilon_t \\ E_t(x_{t+2}) &= E_t(\varepsilon_{t+2} + \theta\varepsilon_{t+1}) = 0. \end{aligned}$$

Similarly, the MA(2) process remembers shocks for two periods,

$$\begin{aligned} E_t(x_{t+1}) &= E_t(\varepsilon_{t+1} + \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1}) = \theta_1\varepsilon_t + \theta_2\varepsilon_{t-1} \\ E_t(x_{t+2}) &= E_t(\varepsilon_{t+2} + \theta_1\varepsilon_{t+1} + \theta_2\varepsilon_t) = \theta_2\varepsilon_t \\ E_t(x_{t+3}) &= 0 \end{aligned}$$

Notice the rule for working these out: $E_t(\varepsilon_{t+j}) = 0$. Things with index less than or equal to t are known at time t, so they are numbers, not random variables. They stay in the conditional mean formula. Things with index greater than t are random variables, and we assumed the conditional mean of the unknown ε are zero.

The conditional variances of the MA(1) process are

$$\begin{aligned} \sigma_t^2(x_{t+1}) &= \sigma_t^2(\varepsilon_{t+1} + \theta\varepsilon_t) = \sigma_\varepsilon^2 \\ \sigma_t^2(x_{t+2}) &= \sigma_t^2(\varepsilon_{t+2} + \theta\varepsilon_{t+1}) = \left(1 + \theta^2\right)\sigma_\varepsilon^2 \\ \sigma_t^2(x_{t+3}) &= \sigma_t^2(\varepsilon_{t+3} + \theta\varepsilon_{t+2}) = \left(1 + \theta^2\right)\sigma_\varepsilon^2 \\ \sigma_t^2(x_{t+j}) &= \sigma_t^2(\varepsilon_{t+j} + \theta\varepsilon_{t+j-1}) = \left(1 + \theta^2\right)\sigma_\varepsilon^2;\ j \geq 3 \end{aligned}$$

Notice the rules for working these out. The *variance* of the known ε_t with index t or less is zero. Things with index greater than t are random variables at time t, so enter variance formulas. The ε_t are uncorrelated over time; that's how I was able to eliminate the last term in the second line,

$$\sigma_t^2(x_{t+2}) = \sigma_t^2(\varepsilon_{t+2})+\theta^2\sigma_t^2(\varepsilon_{t+1})+2\theta cov_t(\varepsilon_{t+1},\varepsilon_{t+2}) = \sigma_t^2(\varepsilon_{t+2})+\theta^2\sigma_t^2(\varepsilon_{t+1})$$

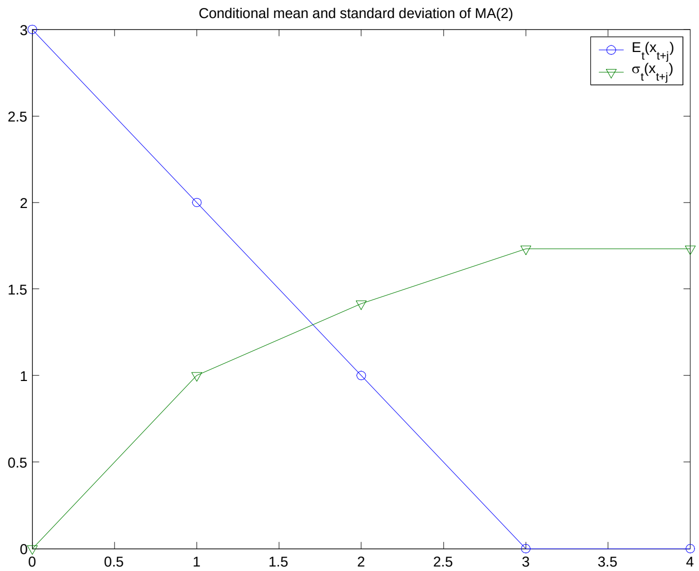
By assumption the ε_t random variable has the same conditional variance at all horizons – like a coin flip, it's the same thing over and over again. That's how I was able to collapse the last term

$$\sigma_t^2(\varepsilon_{t+2}) + \theta^2\sigma_t^2(\varepsilon_{t+1}) = \left[1+\theta^2\right]\sigma_\varepsilon^2.$$

If you got these rules, you should be able to work out the conditional variance of the MA(2),

$$\begin{aligned}
\sigma_t^2(x_{t+1}) &= \sigma_\varepsilon^2 \\
\sigma_t^2(x_{t+2}) &= \left(1+\theta_1^2\right)\sigma_\varepsilon^2 \\
\sigma_t^2(x_{t+j}) &= \left(1+\theta_1^2+\theta_2^2\right)\sigma_\varepsilon^2;\ j \geq 3
\end{aligned}$$

From looking at the formulas, you should see that the MA(k) process has a k period memory. After k periods it forgets where it came from. All of the effects of conditioning information die out after k periods. To show this, I graph below the conditional mean and variance of a MA(2). To make the graph I specify $\theta_1 = \theta_2 = 1$ and $\varepsilon_t = \varepsilon_{t-1} = \varepsilon_{t-2} = 1$. The conditional mean and variance are different of course for different histories; different values of $\varepsilon_t, \varepsilon_{t-1}$. Graphing this for other values of $\varepsilon_t, \varepsilon_{t-1}$ is a great exercise.



Since I assumed $\varepsilon_t = \varepsilon_{t-1} = \varepsilon_{t-2} = 1$, this series has just had a string of good luck. You can see the persistence of the series in the $E_t(x_{t+j})$ line; the expected future values of x_t are also large, and tail off to the unconditional mean $E(x_t) = 0$ after 3 periods. We know a bit about where x_{t+1} will be at time t, and less about where x_{t+2} will be. This is reflected in the slow rise of the standard deviation graph. It too becomes the unconditional standard deviation after 3 periods. Smaller values of θ would give quicker reversion of these lines to their unconditional values.

Means and trends

The models I have described so far all have unconditional mean zero. Most of our series don't do that; they move around some nonzero mean. It's easy enough to fix this by adding constants to the model, for example

$$x_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Now you can see everything shifts up by μ. Some series like income trend up over time, and people sometimes model this by adding a trend,

$$x_t = a + b \times t + \varepsilon_t + \theta_1 \varepsilon_{t-1}.$$

You might want to model temperature in Chicago by adding a deterministic seasonal pattern on top of the random vagaries of the weather,

$$x_t = \mu + b \sin(t) + \varepsilon_t + \theta \varepsilon_{t-1}.$$

This is so easy to do that I, like most writers, routinely suppress the constants and trends, leaving you to put them in where appropriate.

AR Models

The other way to make more complex processes out of the simple white noise process is the *autoregressive process*, denoted AR. The AR(1) is

$$x_t = \rho x_{t-1} + \varepsilon_t.$$

The AR(2) is

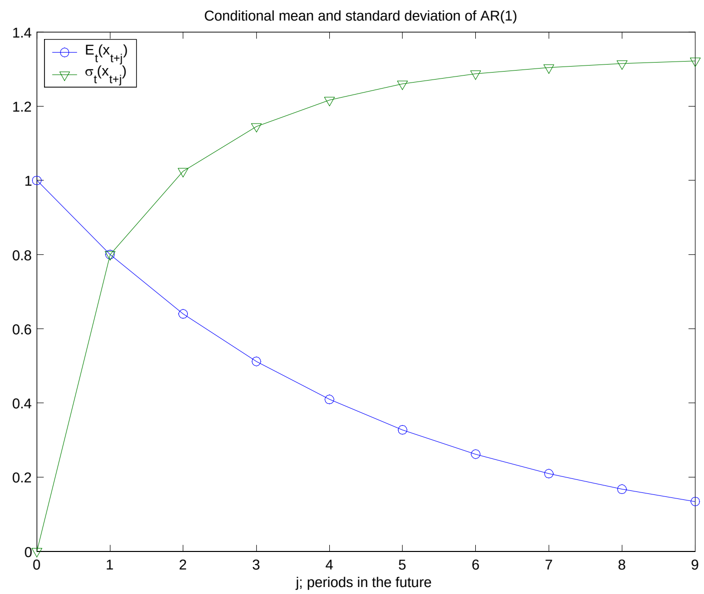
$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \varepsilon_t,$$

and so forth. A problem at the end of the section asks you to work out the conditional mean and variance of the AR(1), using the tricks above. The first two terms are

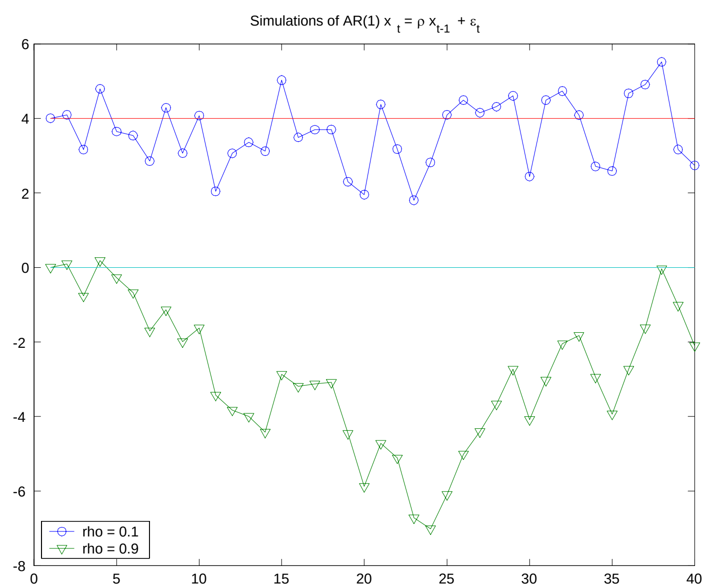
$$\begin{aligned} E_t(x_{t+1}) &= \rho x_t \\ E_t(x_{t+2}) &= \rho^2 x_t \end{aligned}$$

$$\begin{aligned} \sigma_t(x_{t+1}) &= \sigma_\varepsilon^2 \\ \sigma_t^2(x_{t+2}) &= (1+\rho^2)\sigma_\varepsilon^2. \end{aligned}$$

The next figure presents the results. As you can see, the AR(1) is a much nicer process. The conditional mean geometrically decays back to the unconditional mean, rather than lose all its information in exactly k periods as the MA(k) process does. The conditional standard deviation slowly grows, approaching the unconditional standard deviation, as today's information loses its value for predicting the future. It's also nice that the conditional means depend on x_t, the value of the time series itself, rather than the more nebulous shocks ε_t.



Here are plots of an AR(1) with $\rho = 0.1$ and $\rho = 0.9$. You can see how higher ρ induces more persistence in the time series. That's just what we want to model.



MA and AR processes are really not distinct; you can represent each AR as an MA and vice versa. To do this, just start substituting recursively,

$$\begin{aligned} x_t &= \rho x_{t-1} + \varepsilon_t \\ x_t &= \rho\left(\rho x_{t-2} + \varepsilon_{t-1}\right) + \varepsilon_t = \rho^2 x_{t-2} + \varepsilon_t + \rho\varepsilon_{t-1} \\ x_t &= \rho^2\left(\rho x_{t-3} + \varepsilon_{t-2}\right) + \varepsilon_t + \rho\varepsilon_{t-1} = \rho^3 x_{t-3} + \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} \end{aligned}$$

Continuing this way, and if $\|\rho\| < 1$, we see that the AR(1) is the same as an MA(∞),

$$x_t = \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2\varepsilon_{t-2} + \rho^3\varepsilon_{t-3} + ... = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}.$$

This fact can be useful. For example, you saw that it was pretty easy to calculate conditional means and variances of MA processes, so expressing an AR as an MA might be a good idea for that purpose. On the other hand, it was nice to capture the history of the process with the x variable rather than the ε in the AR process, so you might express an MA as an AR for that purpose.

Fitting a model

How would you go about fitting a time series process, i.e., figuring out which AR or MA process best describes a time series like stock price/earnings ratios, returns, interest rates, etc.? The AR processes are particularly convenient because you can *fit the parameters by simple regressions*. Look again at the AR(1) model,

$$x_t = \rho x_{t-1} + \varepsilon_t.$$

The error term is, by assumption, unpredictable by and hence uncorrelated with any information at time $t-1$, including x_{t-1}. The central requirement to run a regression is that the right hand variable x_{t-1} is uncorrelated with the error term ε_t. So, we can obtain a consistent estimate of ρ by just running a regression! In practice, you would include the constant or trend too, i.e. run

$$x_t = \mu + \rho x_{t-1} + \varepsilon_t.$$

It takes a while to get used to this kind of regression. There is no sense in which x_{t-1} "causes" x_t, and we will see cases in which the causality in fact runs exactly the opposite way, (return forecasting regressions in particular). Yet all OLS requires is that the error is uncorrelated with the right hand variable, and *forecast errors* are by definition uncorrelated with information at time t.

The other parameter is the variance of the error term σ_ε^2, and you get this of course from the variance of the residual of the regression. What could be simpler?

More complex models

Once you see the picture, you can see lots of other interesting ways to build more complex and interesting time series models from the simple white noise building blocks, in order to capture the many interesting things we see in the data.

The future of one variable, x_t might be related to the past history of other variables y_t. For example, it seems we can forecast stock returns by the D/P ratio as well as by past stock returns. This suggests a many-variable AR(1),

$$\begin{aligned} x_t &= \mu_x + \rho_{xx} x_{t-1} + \rho_{xy} y_{t-1} + \varepsilon_t^x \\ y_t &= \mu_y + \rho_{yx} x_{t-1} + \rho_{yy} y_{t-1} + \varepsilon_t^y . \end{aligned}$$

The natural way to represent this is to throw both variables in together and think of a *vector autoregression*

$$\begin{aligned} z_t &= \mu + A z_{t-1} + \varepsilon_t \\ z_t &= \begin{bmatrix} x_t \\ y_t \end{bmatrix}; \ \mu = \begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}; \ A = \begin{bmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{bmatrix}; \ \varepsilon_t = \begin{bmatrix} \varepsilon_t^x \\ \varepsilon_t^y \end{bmatrix} \end{aligned}$$

You can fit it by simply running *two* OLS regressions, one of x on past x and past y; and one of y on past x and past y. Everything I have done so far can just be reinterpreted with vectors and matrices to handle many variables and cross-forecasting properties.

Financial markets often show persistence in *volatility* as well as persistence in *means.* The models I have written down don't have that, $\sigma_t(x_{t+1})$ is always the same. The *ARCH and GARCH* models for which Rob Engel got the Nobel prize use these same ideas for variances in place of the actual variables. Interest rates are more volatile when interest rates are higher, and this is often represented in *term structure models* with equations like

$$x_{t+1} = \mu + \rho x_t + \sqrt{x_t}\varepsilon_{t+1}.$$

In this model, when the *level* x_t is higher, the *conditional variance* $\sigma_t^2(x_{t+1}) = x_t \sigma_\varepsilon^2$ is also higher. (This is called, no surprise, the *square root model.)*

Stationarity.

The models I have written down all have some properties that are worth pointing out.

I have blithely used the same symbol for the variance of different elements of the time series; I have assumed for example that $\sigma^2(x_{10}) = \sigma^2(x_{20})$, and used the symbol $\sigma^2(x_t)$ to denote the common value of both. I have also assumed that correlations depend on horizon, not on the date, $E(x_t x_{t+10}) = E(x_{t+20} x_{t+30})$ for example. These are properties of the AR and MA models I have written down, built up by stable functions of the white noise coin flip. (With some restrictions, for example $\|\rho\| < 1$ for the AR(1) model.)

This property, that unconditional means and variances exist and that they depend on horizon rather than date, is called *stationarity.* Intuitively, a stationary series is one that looks the same, from a statistical point of view at any moment in time. A repeated coin toss is stationary (so long as the coin doesn't wear out). The first through tenth observations look just like the hundred and first through hundred and tenth observations from a statistical point of view. Stock and bond returns are remarkably stationary, despite the large changes in the economy and trading mechanism over time.

In my examples, the *conditional* mean and variance got closer and closer to the *unconditional* mean and variance as the horizon got longer, $\lim_{j\to\infty} E_t(x_{t+j}) = E(x_t)$. Similarly, as we look back in time, people had less and less information about what today would be like, $\lim_{j\to\infty} E_{t-j}(x_t) = E(x_t)$. These are also properties of stationary time series.

A coin toss has a much stronger property, it's *independent* over time. The next coin flip is completely unpredictable. This means that the conditional mean is the same as the unconditional mean, $E_t(x_{t+1}) = E(x_{t+1})$. In addition, the conditional variance and the whole conditional distribution $f(x_{t+1}|I_t)$ is the same over time and equal to the unconditional distribution $f(x_t)$.

Independence is stronger than stationarity. Price/earnings ratios and bond yields are stationary (at least I hope so), but not independent. If today has a high P/E or a high interest rate, tomorrow is also likely to have large values of these variables. However, over long time periods these swings even out, so we have really no idea whether P/E or interest rates will be high or low in, say 2100. Stock returns are much closer to coin flips, though not exactly so. The AR(1) with $\rho = 0.9$ is *stationary*, but it is not *independent* over time.

For a lot of reasons, we want always to work with stationary time series.

This usually requires no more than a clever choice of units. For example, stock *prices* are not stationary, since they rise over time. If the Dow is 200, you know you're looking at the 1920s, if it's 10,000, you're looking at the 1990s, and the two sets of numbers are not comparable. But price / dividend or price earnings ratios are much more likely to be stationary, and returns even more so. That's one good reason why we work with these transformations of the variables.

More information

I wrote a more comprehensive set of time series notes titled "Time series for macroeconomics and finance" which you can get off my webpage. James Hamilton's book *Time Series Analysis* is the standard textbook for this kind of (discrete-time) time series analysis.

Chapter 4

Maximization

The heart of Finance is saving/investment/portfolio problems. How much should an investor save vs. consume and which assets should he or she buy? We solve these problems by maximizing an objective (utility) subject to the constraint that you only have so much money.

In this way, finance is really just the application of apples and oranges economics to financial markets. The standard microeconomics problem is, maximize utility subject to a budget constraint. In finance, using different letters (c_t and c_{t+1}) this becomes the optimal portfolio/investment-saving problem.

Here I review how to do constrained optimization, using the standard micro problem as an example.

The consumer wants to maximize utility of two goods, apples X and oranges Y (or pizza and beer, or whatever example your micro teacher used) subject to the budget constraint

$$\max_{\{X,Y\}} U(X,Y) \text{ s.t. } P_X X + P_Y Y = W$$

For example, using one popular utility function,

$$\max_{\{X,Y\}} [\log(X) + a\log(Y)] \quad \text{s.t. } P_X X + P_Y Y = W$$

Solving by substitution

In this simple sort of problem you can get by using the constraint to substitute out one of the goods:

$$\max_{\{X,Y\}} U\left(X, \frac{W - P_X X}{P_Y}\right)$$
$$\max_{\{X\}} \log(X) + a\log\left(\frac{W - P_X X}{P_Y}\right)$$

You find any maximum by setting the derivative of the objective with respect to the choice variable to zero:

$$\begin{aligned}
\frac{d}{dX}U\left(X, \frac{W - P_X X}{P_Y}\right) &= 0 \\
\frac{\partial U}{\partial X} - \frac{P_X}{P_Y}\frac{\partial U}{\partial Y} &= 0.
\end{aligned}$$

In our example,

$$\begin{aligned}
\frac{d}{dX}\left\{\log(X) + a\log\left(\frac{W - P_X X}{P_Y}\right)\right\} &= 0 \\
\frac{1}{X} - a\frac{P_X}{P_Y}\frac{P_Y}{W - P_X X} &= 0 \\
\frac{1}{X} &= \frac{aP_X}{W - P_X X} \\
W - P_X X &= aXP_X \\
W &= (a+1)\,P_X X \\
X &= \frac{W}{(1+a)P_X}
\end{aligned}$$

Notice the downward sloping demand curve. The fact that P_Y does not enter is a peculiarity of the log utility function; usually demand depends on P_X/P_Y. Then we find Y from the constraint,

$$Y = \frac{W - P_X X}{P_Y} = \frac{W - P_X\left(\frac{W}{(1+a)P_X}\right)}{P_Y} = \frac{a}{(1+a)}\frac{W}{P_Y}.$$

Solving by Lagrangian

This gets the answer but it clearly is not going to be a pretty way to attack the problem when we have a lot of goods (stocks) to choose from. In that case, it's better to solve the problem by a "Lagrangian." Here's the trick: add (or subtract) λ times the constraint to the problem. Then differentiate with respect to X, Y and the constraint λ, and solve the resulting system of three equations. (The derivative with respect to λ just restates the constraint.)

$$\max_{\{X,Y,\lambda\}} U(X,Y) - \lambda(P_X X + P_Y Y - W)$$

$$\begin{aligned}
X &: \frac{\partial U}{\partial X} = \lambda P_X \\
Y &: \frac{\partial U}{\partial Y} = \lambda P_Y \\
\lambda &: P_X X + P_Y Y = W
\end{aligned}$$

The first two equations give "marginal rate of substitution equals price ratio"

$$\frac{\partial U/\partial X}{\partial U/\partial Y} = \frac{P_X}{P_Y}.$$

In our example,

$$\begin{aligned}
\frac{1}{X} &= \lambda P_X \to X = \frac{1}{\lambda P_X} \\
\frac{a}{Y} &= \lambda P_Y \to Y = \frac{a}{\lambda P_Y} \\
P_X X + P_Y Y &= W
\end{aligned}$$

Use the X and Y equations in the constraint, and solve for λ

$$\begin{aligned}
P_X\left(\frac{1}{\lambda P_X}\right) + P_Y\left(\frac{a}{\lambda P_Y}\right) &= W \\
\left(\frac{1}{\lambda}\right) + \left(\frac{a}{\lambda}\right) &= W \\
\lambda &= \frac{1+a}{W}
\end{aligned}$$

Now use λ (the "shadow price of the constraint") to find X and Y

$$\begin{aligned}
X &= \frac{1}{\lambda P_X} = \frac{W}{1+a}\frac{1}{P_X} \\
Y &= \frac{a}{\lambda P_Y} = \frac{aW}{1+a}\frac{1}{P_Y}
\end{aligned}$$

Same answer, but I hope you can see that the method is much prettier, treats X and Y symmetrically, and that this will be the way to go when we have many things to choose, as in a portfolios problem

Chapter 5

Matrix algebra

The only way to keep track of 25 portfolios, 3 factors, 600 months of data and so forth and stay sane is to organize the data in matrices. This is a quick refresher of all you need to know about matrices for this course.

A *matrix* is just a rectangular set of numbers, i.e. a cell range in excel.

$$A = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \\ j & k & l \end{bmatrix}$$

A frequent special case is a *vector* which only has one column,

$$x = \begin{bmatrix} a \\ d \\ g \\ j \end{bmatrix}$$

We often use small letters for vectors and big letters for matrices.

You add matrices just by adding their elements together

$$\begin{aligned} A &= \begin{bmatrix} a & b \\ c & d \end{bmatrix}, B = \begin{bmatrix} e & f \\ g & h \end{bmatrix} \\ A+B &= \begin{bmatrix} a & b \\ c & d \end{bmatrix} + \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a+e & b+f \\ c+g & d+h \end{bmatrix} \end{aligned}$$

To add matrices, they must have the same shape.

You can multiply matrices together. The rule is that you take the columns of the matrix on the right, put them on the rows of the matrix on the left, multiply the elements and add them up.

$$\begin{aligned} A &= \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}; B = \begin{bmatrix} g & h \\ i & j \\ k & l \end{bmatrix} \\ AB &= \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \begin{bmatrix} g & h \\ i & j \\ k & l \end{bmatrix} = \begin{bmatrix} ag+bi+ck & ah+bj+cl \\ dg+ei+fk & dh+ej+fl \end{bmatrix} \end{aligned}$$

You can only do this if the matrices have the right size – if the number of columns of the first matrix equals the number of rows of the second matrix. If not, you can't multiply them. Matrix multiplication works like regular multiplication in a lot of ways. For example

$$(A+B)C = AC + BC$$

The big difference is that, in general,

$$AB \neq BA$$

For example, in the above multiplication you can't even do BA. Thus, the order of multiplication matters.

The *identity matrix* is the matrix equivalent of 1.

$$I = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Note $AI = IA = A$ for any A.

In matlab, you can also do an *element by element* multiplication,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} .* \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} ae & bf \\ cg & dh \end{bmatrix}$$

This is a different thing than matrix multiplication as defined above. Just keep track of which one you're doing. You can also do *element by element division*

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} ./ \begin{bmatrix} e & f \\ g & h \end{bmatrix} = \begin{bmatrix} a/e & b/f \\ c/g & d/h \end{bmatrix}$$

Another fun thing we often do is *transpose* a matrix, using the prime ' notation.

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \rightarrow A' = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$

A matrix is *symmetric* if the off diagnoal elements are the same, so the matrix equals its transpose

$$A = \begin{bmatrix} a & d & e \\ d & b & f \\ e & f & c \end{bmatrix};\ A = A'$$

The transpose of a vector is a *row vector*

$$x = \begin{bmatrix} a \\ b \end{bmatrix} \to x' = \begin{bmatrix} a & b \end{bmatrix}$$

It's good to know the *general shape of things.* We often multiply a square matrix by a vector, and this produces another vector,

$$Ax = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \begin{bmatrix} bf + ae \\ df + ce \end{bmatrix} = y$$

A row vector times a matrix produces another row vector

$$x'A = \begin{bmatrix} e & f \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} cf + ae & df + be \end{bmatrix} = y'$$

A row vector times a column vector equals a number. This is sometimes called an *inner product,*

$$x'y = \begin{bmatrix} a & b & c \end{bmatrix} \begin{bmatrix} d \\ e \\ f \end{bmatrix} = [ad + cf + be]$$

The other way around, a column vector times a row vector produces a matrix. This is sometimes called an *outer product.*

$$xy' = \begin{bmatrix} a \\ b \\ c \end{bmatrix} \begin{bmatrix} d & e & f \end{bmatrix} = \begin{bmatrix} ad & ae & af \\ bd & be & bf \\ cd & ce & cf \end{bmatrix}$$

One of the most fun things to do is called a *quadratic form.* This also produces a number. We use it most commonly with a symmetric matrix in the middle,

$$x'Ax = \begin{bmatrix} e & f \end{bmatrix} \begin{bmatrix} a & b \\ b & d \end{bmatrix} \begin{bmatrix} e \\ f \end{bmatrix} = \left[ae^2 + df^2 + 2bef\right]$$

The last thing we do with matrices is *invert* them. For example, if you have an equation

$$Ax = b$$

with A a matrix and b, x vectors, it would be nice to solve this as

$$x = A^{-1}b$$

The inverse has the property that

$$AA^{-1} = A^{-1}A = I$$

You can't always invert a matrix. It has to be square, and it has to have "full rank"; the equivalent of "you can't divide by zero" for matrices. The computer will tell you when this is the case. There is a formula for inverse in the 2x2 case,

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

but you almost never invert matrices by hand. You can see we need $ad - bc \neq 0$ for this to work.

Examples: when we form a portfolio R^p of an underlying set of assets $R^1, R^2, , , R^n$, with weights $w_1, w_2, ...w_N$ we might write

$$R^p = w_1R^1 + w_2R^2 + ... + w_NR^N = \sum_{i=1}^{N} w_iR^i$$

We could also write it more compactly with matrices as

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix};\ R = \begin{bmatrix} R^2 \\ R^2 \\ \vdots \\ R^N \end{bmatrix}; R^p = w'R$$

This is a nice use of *inner product.* We also represent multiple regressions this way,

$$y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_t = \beta' x_t + \varepsilon_t$$

We use the *outer product* for covariance matrices. We keep track of the variance of two random variables x and y and their covariance in a *covariance matrix*,

$$\begin{bmatrix} \sigma^2(x) & cov(x,y) \\ cov(x,y) & \sigma^2(y) \end{bmatrix}$$

If we call x and y elements of a vector z,

$$z = \begin{bmatrix} x \\ y \end{bmatrix}$$

Then we can think of "variance of z" as

$$E(zz') = E\begin{bmatrix} x \\ y \end{bmatrix}\begin{bmatrix} x & y \end{bmatrix} = \begin{bmatrix} E(x^2) & E(xy) \\ E(xy) & E(y^2) \end{bmatrix}$$

If they have mean zero, this is the covariance matrix. If not, the natural vector version of the standard formula $\sigma^2(z) = E(z^2) - E(z)^2$ works,

$$cov(z, z') = E(zz') - E(z)E(z')$$

We use a *quadratic form* to find the variance of a portfolio given the variance of the underlying returns. With mean zero R

$$var(w'R) = E((w'R)(R'w)) = w'E(RR')w$$

(w are numbers, R are random so the numbers come out of E). The formula for the Sharpe ratio in the APT looks like this, as does the central part of the GRS test, $\hat{\alpha}'\Sigma^{-1}\hat{\alpha}$

Chapter 6

Stocks

6.1 The "fallacy of time-diversification."

Many people suggest you should hold stocks "for the long run" since their returns are "more stable over long horizons." This isn't true. This is a nice example that uses the formulas for mean and variance of a sum.

The two period gross return is

$$R_{0\to 2} = R_{0\to 1} R_{1\to 2}$$

so

$$\ln R_{0\to 2} = \ln R_{0\to 1} + \ln R_{1\to 2}$$

Let's look at the mean and standard deviation of the two period return, assuming that mean returns are the same every year and that returns are independent over time.

$$E\left(\ln R_{0\to 2}\right) = E\left(\ln R_{0\to 1}\right) + E\left(\ln R_{1\to 2}\right) = 2E\left(\ln R\right)$$

$$\sigma^2\left(\ln R_{0\to 2}\right) = \sigma^2\left(\ln R_{0\to 1}\right) + \sigma^2\left(\ln R_{1\to 2}\right) + cov.. = 2\sigma^2\left(\ln R\right)$$

Look. The *ratio of mean return to variance of return is independent of horizon* (again, if the mean is the same every year and returns are independent over time).

Where did the fallacy come from? Let's look at *annualized* returns.

$$R^{ann}_{0\to 2} = \left(R_{0\to 1} R_{1\to 2}\right)^{\frac{1}{2}}$$

$$\ln R_{0\to 2}^{ann} = \frac{1}{2}\left(\ln R_{0\to 1} + \ln R_{1\to 2}\right)$$

$$E\ln R_{0\to 2}^{ann} = \frac{1}{2}\left(E\ln R_{0\to 1} + E\ln R_{1\to 2}\right) = E\ln R$$

$$\sigma^2 \ln R_{0\to 2}^{ann} = \sigma^2\left[\frac{1}{2}\left(\ln R_{0\to 1} + \ln R_{1\to 2}\right)\right] = \frac{1}{4}2\sigma^2\left(\ln R\right) = \frac{1}{2}\sigma^2\left(\ln R\right).$$

The mean of annualized returns is the same as the horizon gets longer, but the variance of annualized returns goes down as the horizon gets longer.

But who cares if the variance of annualized returns gets smaller? You care about the total return, which is the annualized return raised to the power of the horizon. The explosive effect of compounding exactly undoes the stabilizing effects of longer horizon.

6.2 Mean-variance frontier for N risky assets

Every textbook does the case for two assets, two assets plus risk free rate, and then the tangency portfolio. What's the real answer – how do we compute the MVF for N risky assets? Here we go. The tools are the means and variances of sums as above, plus matrix manipulations.

6.2.1 Summary

Let

$$\mathbf{E} = \text{Vector of mean returns}$$

$$\mathbf{V} = \text{Variance-covariance matrix}$$

$$\mathbf{1} = \text{Vector of 1's}$$

$$A = \mathbf{E}'\mathbf{V}^{-1}\mathbf{E};\;\; B = \mathbf{E}'\mathbf{V}^{-1}\mathbf{1};\;\; C = \mathbf{1}'\mathbf{V}^{-1}\mathbf{1}.$$

Then, for a given mean portfolio return $\mu = E\left(R^p\right)$, the minimum variance portfolio has variance

$$var\left(R^p\right) = \frac{C\mu^2 - 2B\mu + A}{AC - B^2}$$

and is formed by portfolio weights

$$\mathbf{w} = \mathbf{V}^{-1}\frac{\mathbf{E}\left(C\mu - B\right) + \mathbf{1}\left(A - B\mu\right)}{\left(AC - B^2\right)}.$$

6.2.2 Derivation

The problem is, minimize the variance of a portfolio given a value for the portfolio mean.

$$\min var\left(R^p\right) \text{ s.t. } E\left(R^p\right)=\mu$$

The portfolio return R^p is a combination of N individual asset returns

$$R^p=\sum_{i=1}^{N} w_i R^i; \quad \sum_{i=1}^{N} w_i = 1$$

The w's are *portfolio weights*; they express what fraction of your wealth goes into each asset. Thus, they sum to 1. Thus, you choose portfolio weights to do the minimization.

It's much easier to do all this with matrix notation. Let

$$\mathbf{w}=\left[\begin{array}{c} w_1 \\ w_2 \\ \vdots \\ w_N \end{array}\right]; \ \mathbf{R}=\left[\begin{array}{c} R^1 \\ R^2 \\ \vdots \\ R^N \end{array}\right]; \ \mathbf{1}=\left[\begin{array}{c} 1 \\ 1 \\ \vdots \\ 1 \end{array}\right].$$

Then the portfolio return is

$$R^p=\mathbf{w}'\mathbf{R} \ ;$$

the condition that the weights add up to 1 is

$$1=\mathbf{1}'\mathbf{w}.$$

The mean of the portfolio return is

$$E\left(R^p\right)=E\left(\mathbf{w}'\mathbf{R}\right)=\mathbf{w}'E\left(\mathbf{R}\right)=\mathbf{w}'\mathbf{E}.$$

The last equality just simplifies notation. The vector of mean returns shows up so much that I'll call it $\mathbf{E}$ instead of carrying $E\left(\mathbf{R}\right)$ around.

The variance of the portfolio return is

$$var\left(R^p\right)=E\left[\left(R^p-E\left(R^p\right)\right)^2\right]=E\left[\left(\mathbf{w}'\mathbf{R}-\mathbf{w}'\mathbf{E}\right)^2\right]=E\left[\left(\mathbf{w}'\left(\mathbf{R}-\mathbf{E}\right)\right)^2\right]=$$

$$=E\left[\mathbf{w}'\left(\mathbf{R}-\mathbf{E}\right)\left(\mathbf{R}-\mathbf{E}\right)'\mathbf{w}\right]=\mathbf{w}'\mathbf{V}\mathbf{w}$$

where

$$\mathbf{V}=E\left[\left(\mathbf{R}-\mathbf{E}\right)\left(\mathbf{R}-\mathbf{E}\right)'\right]$$

is the *variance-covariance* matrix of returns.

Now, let's restate our problem with this notation. Minimizing the variance is the same as minimizing $1/2$ the variance, so we want to choose weights to

$$\min_{\mathbf{w}} \frac{1}{2}\mathbf{w}'\mathbf{V}\mathbf{w} \;\; \text{s.t.} \quad \mathbf{w}'\mathbf{E} = \mu; \; \mathbf{w}'\mathbf{1} = 1.$$

To do a constrained minimization, you form the *Lagrangian,*

$$\mathcal{L} = \frac{1}{2}\mathbf{w}'\mathbf{V}\mathbf{w} - \lambda\left(\mathbf{w}'\mathbf{E} - \mu\right) - \delta\left(\mathbf{w}'\mathbf{1} - 1\right).$$

Then, you take derivatives of the Lagrangian with respect to the choice variables and the multipliers λ and δ. Taking derivatives of matrices with respect to vectors works just as you'd think it does, so we have

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} : \quad \mathbf{V}\mathbf{w} - \lambda\mathbf{E} - \delta\mathbf{1} = \mathbf{0}$$

and the two constraints. We want the weights $\mathbf{w}$, so

$$\mathbf{w} = \mathbf{V}^{-1}\left(\mathbf{E}\lambda + \mathbf{1}\delta\right)$$

What about λ and δ? We determine these so that the constraints are satisfied. Plugging this value of $\mathbf{w}$ into the constraint equations, we get

$$\mathbf{w}'\mathbf{E} = \mathbf{E}'\mathbf{w} = \mu \to \mathbf{E}'\mathbf{V}^{-1}\left(\mathbf{E}\lambda + \mathbf{1}\delta\right) = \mu$$

$$\mathbf{w}'\mathbf{1} = \mathbf{1}'\mathbf{w} = 1 \to \mathbf{1}'\mathbf{V}^{-1}\left(\mathbf{E}\lambda + \mathbf{1}\delta\right) = 1.$$

We want to solve these two equations in the two unknowns λ, δ. So write them as

$$\mathbf{E}'\mathbf{V}^{-1}\mathbf{E}\lambda + \mathbf{E}'\mathbf{V}^{-1}\mathbf{1}\delta = \mu$$

$$\mathbf{1}'\mathbf{V}^{-1}\mathbf{E}\lambda + \mathbf{1}'\mathbf{V}^{-1}\mathbf{1}\delta = 1$$

or

$$\begin{bmatrix} \mathbf{E}'\mathbf{V}^{-1}\mathbf{E} & \mathbf{E}'\mathbf{V}^{-1}\mathbf{1} \\ \mathbf{1}'\mathbf{V}^{-1}\mathbf{E} & \mathbf{1}'\mathbf{V}^{-1}\mathbf{1} \end{bmatrix} \begin{bmatrix} \lambda \\ \delta \end{bmatrix} = \begin{bmatrix} \mu \\ 1 \end{bmatrix}.$$

Now we can solve,

$$\begin{bmatrix} \lambda \\ \delta \end{bmatrix} = \begin{bmatrix} \mathbf{E}'\mathbf{V}^{-1}\mathbf{E} & \mathbf{E}'\mathbf{V}^{-1}\mathbf{1} \\ \mathbf{1}'\mathbf{V}^{-1}\mathbf{E} & \mathbf{1}'\mathbf{V}^{-1}\mathbf{1} \end{bmatrix}^{-1} \begin{bmatrix} \mu \\ 1 \end{bmatrix}.$$

$$\begin{bmatrix} \lambda \\ \delta \end{bmatrix} = \left(\begin{bmatrix} \mathbf{E}' \\ \mathbf{1}' \end{bmatrix} \mathbf{V}^{-1} \begin{bmatrix} \mathbf{E} & \mathbf{1} \end{bmatrix} \right)^{-1} \begin{bmatrix} \mu \\ 1 \end{bmatrix}.$$

Now that we know what the λ and δ multipliers are, we can go back and find the weights,

$$\mathbf{w} = \mathbf{V}^{-1}\left(\lambda\mathbf{E}+\delta\mathbf{1}\right) = \mathbf{V}^{-1}\left[\begin{array}{cc}\mathbf{E} & \mathbf{1}\end{array}\right]\left[\begin{array}{c}\lambda \\ \delta\end{array}\right]$$

$$= \mathbf{V}^{-1}\left[\begin{array}{cc}\mathbf{E} & \mathbf{1}\end{array}\right]\left(\left[\begin{array}{c}\mathbf{E}' \\ \mathbf{1}'\end{array}\right]\mathbf{V}^{-1}\left[\begin{array}{cc}\mathbf{E} & \mathbf{1}\end{array}\right]\right)^{-1}\left[\begin{array}{c}\mu \\ 1\end{array}\right]$$

Now, use these weights to find the portfolio variance (did you forget? That's what we're after!)

$$var\left(R^p\right) = \mathbf{w}'\mathbf{V}\mathbf{w} =$$

$$\left[\begin{array}{cc}\mu & 1\end{array}\right]\left(\left[\begin{array}{c}\mathbf{E}' \\ \mathbf{1}'\end{array}\right]\mathbf{V}^{-1}\left[\begin{array}{cc}\mathbf{E} & \mathbf{1}\end{array}\right]\right)^{-1}\left[\begin{array}{c}\mu \\ 1\end{array}\right].$$

This isn't as bad as it seems. To make it look prettier, give names to the elements of the matrix that got inverted,

$$\left(\left[\begin{array}{c}\mathbf{E}' \\ \mathbf{1}'\end{array}\right]\mathbf{V}^{-1}\left[\begin{array}{cc}\mathbf{E} & \mathbf{1}\end{array}\right]\right) = \left[\begin{array}{cc}A & B \\ B & C\end{array}\right]$$

Then,

$$var\left(R^p\right) = \left[\begin{array}{cc}\mu & 1\end{array}\right]\left[\begin{array}{cc}A & B \\ B & C\end{array}\right]^{-1}\left[\begin{array}{c}\mu \\ 1\end{array}\right]$$

$$= \frac{1}{AC-B^2}\left[\begin{array}{cc}\mu & 1\end{array}\right]\left[\begin{array}{cc}C & -B \\ -B & A\end{array}\right]\left[\begin{array}{c}\mu \\ 1\end{array}\right]$$

$$var\left(R^p\right) = \frac{C\mu^2 - 2B\mu + A}{AC-B^2}$$

The variance is a quadratic function of the mean. (We used the notation $\mu = E\left(R^p\right)$.) That's why we draw bow-shaped frontiers all the time.

We can also write out the portfolio weights in this notation,

$$\mathbf{w} = \mathbf{V}^{-1}\left[\begin{array}{cc}\mathbf{E} & \mathbf{1}\end{array}\right]\left[\begin{array}{cc}\mathbf{E}'\mathbf{V}^{-1}\mathbf{E} & \mathbf{E}'\mathbf{V}^{-1}\mathbf{1} \\ \mathbf{1}'\mathbf{V}^{-1}\mathbf{E} & \mathbf{1}'\mathbf{V}^{-1}\mathbf{1}\end{array}\right]^{-1}\left[\begin{array}{c}\mu \\ 1\end{array}\right]$$

$$= \mathbf{V}^{-1}\left[\begin{array}{cc}\mathbf{E} & \mathbf{1}\end{array}\right]\left[\begin{array}{cc}C & -B \\ -B & A\end{array}\right]\left[\begin{array}{c}\mu \\ 1\end{array}\right] / \left(AC-B^2\right)$$

$$\mathbf{w} = \mathbf{V}^{-1}\frac{\mathbf{E}\left(C\mu-B\right)+\mathbf{1}\left(A-B\mu\right)}{\left(AC-B^2\right)}.$$

Chapter 7

Fixed income (bonds)

7.1 Notation

We need to distinguish bonds of varying maturity. I'll use a superscript in parentheses–$P^{(3)}$ is the price of a three year zero-coupon bond, etc. All logs are natural logs—base e.

7.2 Present value

We start by ignoring uncertainty. Then, all future cash flows are known (no default) and all future interest rates are known. Obviously, we'll later put in expectations of things that happen in the future, and patch up the formulas a bit.

The central trick to all of bond pricing with no uncertainty is to repackage the same things in different guises. A set of zero coupon bonds, a set of coupon bonds, and a current and promised future interest rate are all the same thing. To derive the price of any of these items in terms of the prices of any others, you just figure out how to repackage them.

Any bond is a claim to a sequence of cash flows $\{CF_1,\ CF_2, ...CF_N\}$. We can write its value as

$$P = \sum_{j=1}^{N} \frac{CF_j}{R_0 R_1 R_2 ... R_{j-1}}$$

where R_0 is the interest rate from 0 to 1, etc.

The issue in using the present value formula is: where do you get interest rates R_j?

A) If you know what interest rates banks will charge, and the borrowing and lending rates are equal, then these are the rates to use, by arbitrage. This only happens in textbooks.

B) More plausibly, you can use the price of zero coupon bonds as quoted in the *Wall Street Journal*. They should be

$$P^{(N)} = \frac{1}{R_0 R_1 ... R_{N-1}}.$$

Then, we can get rid of the fictitious R's and write

$$P = \sum_{j=1}^{N} P^{(j)} CF_j.$$

You can get to the same formula directly, if you want. View the cash flows as zero-coupon bonds, and then realize that the bond is just a combination of the zeros.

C) You can go backwards, and infer zero prices from the prices of coupon bonds, and then use those zero prices to price other coupon bonds.

7.3 Yield

First, we need to define the yield:

> **Yield**: The yield (to maturity) is defined as that fictional, constant, known, annual, interest rate that justifies the quoted price of a bond, assuming that the bond does not default.

From this definition, the yield of a zero coupon bond is the number $Y^{(N)}$ that satisfies

$$P^{(N)} = \frac{1}{\left[Y^{(N)}\right]^N}.$$

Hence

$$Y^{(N)} = \frac{1}{\left[P^{(N)}\right]^{\frac{1}{N}}}; \quad \ln Y^{(N)} = -\frac{1}{N} \ln P^{(N)}.$$

The yield of any stream of cash flows is the number Y that satisfies

$$P = \sum_{j=1}^{N} \frac{CF_j}{Y^j}.$$

In general, you have to search for the value Y that solves this equation, given the cash flows and the price. So long as all cash flows are positive, this is fairly easy to do.

As you can see, *the yield is just a convenient way to quote the price.* In using yields we make *no* assumptions. We do not assume that actual interest rates are known or constant; we do not assume the actual bond is default-free.

7.4 Forward rates

From the prices of zero coupon bonds given above, you can find an implied future interest rate. From

$$P^{(N)} = \frac{1}{R_0 R_1 .. R_{N-1}}$$

it follows that

$$R_N = \frac{P^{(N)}}{P^{(N+1)}}.$$

What do these rates mean? These are *forward rates*

> **Forward rate**: The forward rate of interest is the rate at which you can contract *today* to borrow or lend money starting at period N, to be paid back at period $N+1$.

Here's the neat thing, implicit in the above math. You can *synthesize* a forward contract from a spectrum of zero coupon bonds. Here's how. Suppose you bought one N period zero and simultaneously sold x $N+1$ period zero coupon bonds. Let's track your cash flow at every date.

	Buy N-Period zero	Sell x N+1 Period zeros	Net cash flow
Today 0:	$-P^{(N)}$	$+xP^{(N+1)}$	$xP^{(N+1)} - P^{(N)}$
Time N:	1		1
Time N+1:		-x	-x

Now, choose x so that today's cash flow is zero:

$$x = \frac{P^{(N)}}{P^{(N+1)}}$$

Look at what you have. You pay or get nothing today, you get \$1.00 at N , and you pay $P^{(N)}/P^{(N+1)}$ at $N+1$. You have synthesized a contract signed today for a loan from N to $N+1$—a forward rate! Thus,

$$F_N = \text{Forward rate } N \rightarrow N+1 = \frac{P^{(N)}}{P^{(N+1)}}$$

and of course

$$\ln F_N = \ln P^{(N)} - \ln P^{(N+1)}.$$

Forward rates are useful when you have to plan today for a project, but you will want to borrow money a few years in the future when construction really gets going. It also allows you to put your money where your mouth is if you think you know where interest rates are going.

7.5 Holding period returns

If you buy an N period bond and then sell it—it has now become an $N-1$ period bond—you achieve a return of

$$HPR^{(N)}_{t+1} = \frac{\$\text{back}}{\$\text{paid}} = \frac{P^{(N-1)}_{t+1}}{P^{(N)}_t}$$

or, of course,

$$\ln HPR^{(N)}_{t+1} = \ln P^{(N-1)}_{t+1} - \ln P^{(N)}_t.$$

We date this return (from t to $t+1$) as $t+1$ because that is when you find out its value. Except for one period bonds, you don't know for sure what this return will be. (If this is confusing, take the time to write returns as $HPR_{t\rightarrow t+1}$ and then you'll never get lost.)

7.6 Yield curve

The *yield curve* is a plot of yields of zero coupon bonds as a function of their maturity. What forces determine the prices or yields of bonds of various maturities?

7.6.1 Yield curve with no uncertainty

Suppose we *know* where interest rates, or future yields on one period bonds, are going. Then, the present value formula is

$$P_0^{(N)} = \left(\frac{1}{R_0}\frac{1}{R_1}...\frac{1}{R_{N-1}}\right) = \left(\frac{1}{Y_0^{(1)}}\frac{1}{Y_1^{(1)}}...\frac{1}{Y_{N-1}^{(1)}}\right)$$

The definition of yield is

$$P^{(N)} = \frac{1}{\left[Y^{(N)}\right]^N}.$$

Substituting,

$$Y_0^{(N)} = \left(Y_0^{(1)}Y_1^{(1)}Y_2^{(1)}...Y_{N-1}^{(1)}\right)^{\frac{1}{N}}$$

or, *the yield on an N-Period zero is the geometric average of future interest rates.*

As usual logs are prettier,

$$\ln Y_0^{(N)} = \frac{1}{N}\left(\ln Y_0^{(1)} + \ln Y_1^{(1)} + \ln Y_2^{(1)}... + \ln Y_{N-1}^{(1)}\right)$$

the log yield on an N-Period zero is the arithmetic average of future interest rates.

The right hand side of the yield curve formula expresses one way of getting a dollar from now to N periods from now—roll over one-period bonds. The left hand side expresses another way of getting a dollar from now to N periods from now—buy a 30 year zero coupon bond. With no uncertainty, the two must give the same return, by arbitrage.

7.6.2 Yield curve with uncertainty

What if we don't know where interest rates are going? Well, if traders are risk-neutral enough, then they will either buy N-period zeros or plan to roll over one period bonds if either strategy promises to do better *on average*. This isn't as clean as pure arbitrage, but is at least a workable approximation. And we add a risk premium fudge factor to soak up any errors. Thus, the actual expression for the yield curve relation is

> *The N-period yield is the average of* **expected future** *one-period yields, perhaps plus a risk premium.*

$$Y_0^{(N)} = E_0 \left[\left(Y_0^{(1)} Y_1^{(1)} Y_2^{(1)} ... Y_{N-1}^{(1)} \right)^{\frac{1}{N}} \right] \text{ (+risk premium)}$$

or

$$\ln Y_0^{(N)} = \frac{1}{N} E_0 \left(\ln Y_0^{(1)} + \ln Y_1^{(1)} + \ln Y_2^{(1)} ... + \ln Y_{N-1}^{(1)} \right) \text{ (+risk premium).}$$

Here "E_0" means "conditional expectation as of time 0."

Unless you say something about the risk premium, either of these equations is a tautology–a definition of the risk premium. The *expectations hypothesis* states that the risk premium isn't present. It is a good approximation when the risk premium is small and does not vary much over time. Often, this is a good approximation. More complex *term structure models* are all about quantifying the size and movement over time in the risk premium.

As a technicality, the level and log statements of the expectations hypothesis (no risk premium) are *not* equivalent. ($\ln[E(x)] \neq E[\ln(x)]$.) But since either is a hypothesis, and they say approximately the same thing, we won't make a big deal about the difference between the two statements.

7.6.3 Forward yield curve.

Suppose we *knew* where interest rates were going. Then, arbitrage requires that

$$\textit{Forward rate} = \textit{Future spot rate}$$

$$F^{(N)} = R_{N \to N+1}.$$

Why? Arbitrage. If the forward rate is lower than the future spot rate, people will arrange today to borrow at the forward rate, wait around, and then lend at the spot rate when the time comes, making a certain profit.

Forward rate = future spot rate implies the yield curve. To see this look at one step ahead:

$$F^{(1)} = R_{1 \to 2}.$$

Substituting in the forward rate formula

$$\frac{P^{(1)}}{P^{(2)}} = R_{1 \to 2}.$$

Using $R_{0\to 1} = 1/P^{(1)}$, and $Y^{(2)} = 1/\sqrt{P^{(2)}}$,

$$\left[Y^{(2)}\right]^2 = R_{0\to 1} R_{1\to 2}$$

or

$$Y^{(2)} = [R_0 R_1]^{\frac{1}{2}}$$

our old friend.

This shouldn't be a surprise. If two ways of getting money from Monday to Tuesday, Tuesday to Wednesday, Wednesday to Thursday, etc. each have to be the same—forward rate = future spot rate—it is no surprise that two ways of getting money from Monday to Thursday have to be the same—the yield curve.

With uncertainty, we add expectations and a risk premium the same way

Forward rate = Expected future spot rate (+ risk premium).

This just states that fairly risk neutral traders will take either side of lock in a loan vs. wait until the *expected* gain from either strategy is about the same.

7.6.4 Holding period return yield curve

Consider two ways of getting money from today to tomorrow: Hold an N-period zero coupon bond for a period, selling it as an N-1 period zero coupon bond, or hold a one-period zero coupon bond for a period. Again, if our bond traders are fairly risk-neutral, we expect that the expected holding period returns from one strategy should be the same as for the other, with perhaps a small risk premium. If you like equations,

$$E_0\left(HPR_{t+1}^{(N)}\right) = E_0\left(HPR_{t+1}^{(M)}\right) \quad (+\text{ risk premium})$$

Again, this is the same as yield curve. In the certainty case,

$$HPR_1^{(2)} = HPR_1^{(1)}$$

$$\frac{P_1^{(1)}}{P_0^{(2)}} = \frac{1}{P_t^{(1)}}$$

$$\frac{\left[Y_0^{(2)}\right]^2}{Y_1^{(1)}} = Y_0^{(1)}$$

$$Y_0^{(2)} = \left[Y_0^{(1)} Y_1^{(1)} \right]^{\frac{1}{2}}.$$

Again, no surprise when you think about it. Getting money from one time to another in two different ways has to give the same result.

7.7 Duration

Here's the big picture. We have figured out how to find the value of bonds. Now we can find out how that value changes if something happens. That something is interest rates, and *duration* is the main measure of a bond's sensitivity to interest rate changes. Then, we can figure out how to structure your portfolio so that you are insured or "immunized" against interest rate changes.

Duration is the sensitivity of prices to yield, or, in equations,

$$D = -\frac{\%\text{ Change in } P}{\%\text{ Change in } Y} = -\frac{Y}{P}\frac{dP}{dY} = -\frac{d\ln P}{d\ln Y}.$$

From the definition, we can easily find the duration of zero-coupon bonds.

$$P^{(N)} = \frac{1}{Y^N}$$

$$-\frac{Y}{P}\frac{dP}{dY} = \frac{Y}{P} N \frac{1}{Y^{N+1}} = N$$

For zeros, duration = maturity.

For other bonds,

$$P = \sum_{j=1}^{N} \frac{CF_j}{Y^j}$$

$$-\frac{Y}{P}\frac{dP}{dY} = \frac{Y}{P}\sum_{j=1}^{N} j\frac{CF_j}{Y^{j+1}} = \frac{1}{P}\sum_{j=1}^{N} j\frac{CF_j}{Y^j} = \sum_{j=1}^{N} j\frac{CF_j/Y^j}{\sum_{j=1}^{N} CF_j/Y^j}$$

$$\text{duration} = \sum_{\text{cash flows}} \text{duration of cash flow} \times \frac{\text{value of cash flow}}{\text{total value}}$$

The duration of any bond = value-weighted average of durations of its individual cash flows

The duration of coupon bonds is less than their maturity. As an extreme example, a perpetuity has duration[1]

$$D = -\frac{1}{P}\sum_{j=1}^{\infty} j\frac{C}{Y^j} = \frac{Y}{Y-1}.$$

At a yield of 10%, the *infinite* maturity perpetuity has an eleven year duration.

Duration tells you that a coupon bond is "like" a zero-coupon bond of maturity equal to that duration, where "like" means "has the same interest-rate sensitivity as."

Sometimes it's convenient to quote *modified duration*, the percent change in price for a one *percentage point* change in yield, rather than a one *percent* change in yield. You don't have to take any more derivatives, since,

$$\text{Mod. duration } \equiv -\frac{\%\text{ ch. Price}}{\text{ch. Yield}} = -\frac{1}{P}\frac{dP}{dY} = \frac{1}{Y}\left(-\frac{Y}{P}\frac{dP}{dY}\right) = \frac{1}{Y}\times\text{duration.}$$

7.8 Immunization

Now, how can we structure a portfolio so that it is not sensitive to interest rate changes? There is always one conceptually easy way,

Dedicated portfolio. For each cash flow of assets or liability, buy or sell a corresponding zero-coupon bond. Then, *no matter what happens to interest*

---

[1]

$$D = -\frac{1}{P}\sum_{j=1}^{\infty} j\frac{C}{Y^j} = -\frac{C}{P}\sum_{j=1}^{\infty} j\frac{1}{Y^j} = \frac{Y}{Y-1}$$

For those of you who like to see all the steps, I use the fact that

$$\sum_{j=1}^{\infty} jz^j = \frac{z}{(1-z)^2}.$$

Then, using the fact that

$$P = \sum_{j=1}^{\infty}\frac{C}{Y^j} = \frac{C/Y}{1-1/Y} = \frac{C}{Y-1},$$

we get

$$D = \frac{C}{P}\frac{1/Y}{(1-1/Y)^2} = (Y-1)\frac{Y}{(Y-1)^2} = \frac{Y}{Y-1}.$$

rates, the cash flow will be covered by the maturing zero-coupon bond. Of course, since you can synthesize zeros from coupon bonds, you can do the same thing with artful portfolios of coupon bonds. This is "expensive"—it requires lots of buying and selling.

Duration-matching. Suppose instead you only change two assets or liabilities, so that 1) the present value of assets = the present value of liabilities and 2) the duration of assets = the duration of liabilities. Then, your total position will be insensitive to interest rate changes.

Chapter 8

Options

8.1 Arbitrage, and two applications

Options are priced by *arbitrage*. Rather than find the fundamental determinants of value, we show how the value of the option can be expressed as a function of other, observed securities.

A *payoff* is how much a security is worth at some future date. The payoff is unknown today, and could take on many values, depending on how things come out. The *price* or *value* is how much it is worth today.

There are two fundamental arbitrage facts we use.

> *The Law of One Price:* If two securities have the same *payoff* they must have the same price.

Note "same payoff" here means same payoff *no matter what happens*. Not the same expected payoff.

> *No Arbitrage:* If payoff A is always greater than (or equal to) payoff B then the price of A must be greater than (or equal to) the price of B.
>
> Again, "greater than" means no matter what happens, not greater on average, etc. These statements seem perfectly obvious, but watch what nice and not-so-obvious implications they have.

8.1.1 Put-Call parity

Suppose you hold a call and simultaneously write a put with the same strike price. The payoff is the same as you would get by just holding the stock and promising to pay an amount X for sure! (Draw yourself a payoff diagram to prove it.) If the payoffs are the same, the prices must be the same (*Law of one price*). I.e.,

$$\text{payoff: } C_T - P_T = S_T - X$$

implies

$$\text{price: } C - P = S - PV(X)$$

or

$$C - P = S - X/R.$$

This is the *put-call parity formula.* The *principle* of the law of one price looked trivially obvious, but I bet this instance wasn't obvious!

This fact is useful for two reasons. 1) It illustrates the fundamental principle we use to price options before expiration, 2) It shows you how to find the price of a put given that of a call and vice versa. Thus, we just have to figure out how to value calls, and use put-call parity to value the put. (Strictly speaking, this is valued for European options on stocks that do not pay dividends, but the extensions are pretty easy.)

8.1.2 Arbitrage bounds and early exercise

What does the no-arbitrage principle say directly about the value of a call *before* expiration? We can say several things.

1) $C \geq 0$. The price of a zero payoff is zero. The call payoff is always more than zero. Hence the call price must be greater than zero.

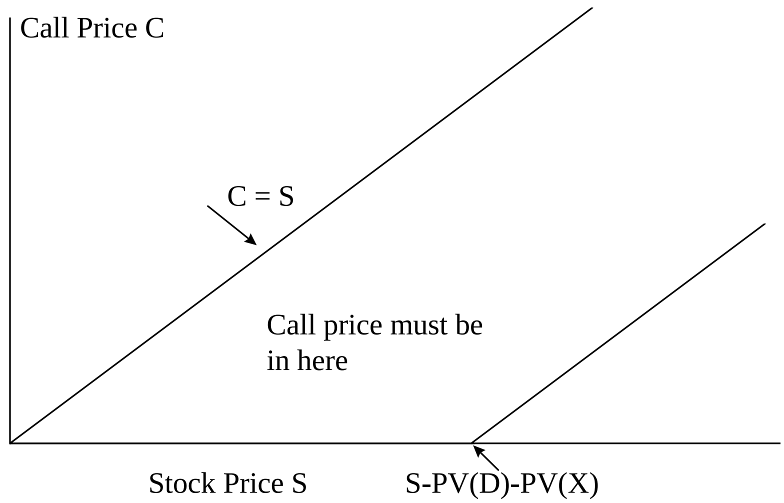
2) $C \leq S$. The call payoff is always less than that of the stock (for positive strike price). Hence the call price must be less than the stock price.

3) $C \geq S - PV(D) - PV(X)$. The call payoff is

$$C_T = \max(S_T - X, 0) \geq S_T - X = S_T + D - D - X$$

Taking prices, (the price of $S_T + D$ is S) we get the relation.

These relations are usually summarized in a graph,



This last inequality has a nice implication. *If interest rates are not zero, it does not pay to exercise an option on a stock that pays no dividends before the expiration date.* Why? With no dividends, you have $C \geq S - PV(X) > S - X$. The right hand side is what you get by exercising, the left by selling. Moral: sell options, don't exercise them.

These are nice illustrations of the logic behind arbitrage bounds. If you don't know the price of something, but you do know the price of something else whose payoff is always larger or smaller, then you can at least get an upper or lower bound.

8.2 Binomial option pricing 1: one step ahead.

Let's find the value of a call option *before* expiration. Again, we'll do European calls on stocks with no dividends, to keep things simple. Since we know the value of a call option *at* expiration,

$$C_T = \max(S_T - X, 0),$$

let's start by finding the value of the call option one period before expiration.

The stock price right now is S. Suppose the stock can do one of two things tomorrow, when it expires: it can go up to $S_T = uS$ or down to

$S_T = dS$. (I'll show you later how to do a more realistic example. As usual, understand the simple version and the more complex one will be clearer.) The call can then take on one of two values, $C_T = C_u = \max(uS - X, 0)$ or $C = C_d = \max(dS - X, 0)$. As usual, denote the value of the stock *today* by S and the value of the option (which we don't know yet) C. I.e., the trees look like

$$\begin{array}{ccccccc} & & uS & & & & C_u = \max(uS - X, 0) \\ & \nearrow & & & & \nearrow & \\ S & & & ; & C & & \\ & \searrow & & & & \searrow & \\ & & dS & & & & C_d = \max(dS - X, 0) \end{array}.$$

We *know* u, d, S, X and we want to find C.

Consider a portfolio consisting of H shares of stock and B face value bonds. The *payoff* of this portfolio is $HuS + B$ if the stock goes up and $HdS + B$ if the stock goes down. Now, arrange the number of shares and the number of bonds so that the *payoff of the stock + bond portfolio is exactly the same as the payoff of the call option.* This means you need to choose H and B so that

$$HuS + B = C_u \text{ and } HdS + B = C_d$$

Two equations in two unknowns. Use your favorite method for solving two equations in two unknowns to check

$$H = \frac{C_u - C_d}{uS - dS} \quad B = \frac{uSC_d - dSC_u}{uS - dS}$$

H is known as the *hedge ratio.* It's the number of shares you hold so that the stock portfolio exactly matches the call option for one period. It is also the change in the option value per change in the stock price. (As S changes from dS to uS, the call value changes from HdS to HuS.) If we make a graph of option value as a function of stock price value, H is the *slope* of that line.

We have two portfolios with exactly the same payoff. By the *law of one price* they must have exactly the same price. In equations, the price or value *today* must satisfy

$$C = HS + B/R.$$

To find the call option price C all we have to do is find the right hand side in terms of primitives. Substituting for H and B,

$$C = \frac{C_u - C_d}{uS - dS} S + \frac{uC_d - C_u d}{u - d} / R$$

$$C = \frac{C_u - C_d}{u - d} + \frac{uC_d - C_u d}{u - d}/R$$

Now, we need to make this formula prettier. Define

$$p = \frac{R - d}{u - d}$$

so

$$1 - p = \frac{u - R}{u - d}$$

In terms of p the formula for C simplifies, as follows

$$C = \frac{C_u}{u - d} - \frac{C_d}{u - d} + \frac{uC_d}{u - d}/R - \frac{C_u d}{u - d}/R$$

$$C = \left(\frac{1}{u - d}\left(1 - \frac{d}{R}\right)\right) C_u + \left(\left(\frac{u}{R} - 1\right)\frac{1}{u - d}\right) C_d$$

$$C = \frac{1}{R}\left[\left(\frac{R - d}{u - d}\right) C_u + \left(\frac{u - R}{u - d}\right) C_d\right]$$

Thus, we have the *Answer:*

$$C = \frac{1}{R}\left[pC_u + (1 - p)\, C_d\right]$$

where, as a reminder,

$$C_u = \max\left(uS - X, 0\right) \quad \text{and} \quad C_d = \max(dS - X, 0).$$

There are three really pretty things to see in this formula.

1) The *probabilities* of the states do not enter! The entire argument followed from the law of one price; if the option does not follow this relation to the stock price S, there is a completely risk-free arbitrage opportunity (hold $HS + B$ and short option or vice versa). Essentially, all we need to know about whether the stock price will go up or down is already included in the stock price S. If the + state gets more likely, the stock price rises.

2) Risk aversion, risk premia, aggregate risk factors, etc. do not enter in the formula. For the same reason: they are already reflected in S. The rest is arbitrage. This works just like pricing a coupon bond *given* the price of zeros.

3) p looks like a probability. It's between 0 and 1. It's called a *risk-neutral probability* for the following reason. Suppose people were in fact

risk neutral and the probability that the stock goes up to uS were in fact p. Then the option price would be its expected discounted value

$$C = \frac{1}{R} E\left(C_{\text{tomorrow}}\right) = \frac{1}{R}\left[pC_u + (1-p)\, C_d\right].$$

In fact, you can use this logic to derive the option value in the first place. Start by realizing that you can synthesize the option with a leveraged stock position—$HS + B$ gives the same payoff as the option. Then, risk aversion must not matter—the price of the option, given the stock price, is the same whether everyone is risk averse or not, since only arbitrage is involved. Now, suppose (counterfactually) that people were risk neutral. What probabilities would give rise to the current stock price? They must be

$$S = \frac{1}{R}\left[puS + (1-p)\, dS\right]$$

solving for p,

$$p = \frac{R-d}{u-d}$$

as before. Again, if people were risk neutral, the option value would be

$$C = \frac{1}{R}\left[pC_u + (1-p)\, C_d\right].$$

That must be the actual option value, since risk aversion can't matter to an arbitrage argument. (This is in fact how much actual option pricing is done. Find a set of risk-neutral probabilities that explain current stock prices, and then use those probabilities to value options.)

Don't confuse risk-neutral probabilities with real probabilities! The real probabilities do not matter for option pricing. Perhaps more accurately, all you need to know about them is reflected in today's stock price. (If you wanted to try to price an option *without* knowledge of today's stock price, you'd be back to beta, and probabilities, means, risk premia, etc. would all matter.)

8.3 Binomial option pricing 2: Two steps ahead

Now, let's try two periods until expiration. Suppose the stock can rise or decline by u or d each period. Then denote the option prices by C_u C_{uu}, etc. so the stock and option prices follow

$$
\begin{array}{ccccc}
 & & u^2S & & \\
 & \nearrow & & & \\
 & uS & & & \\
\nearrow & & \searrow & & \\
S & & udS & & \\
\searrow & & \nearrow & & \\
 & dS & & & \\
 & & \searrow & & \\
 & & d^2S & &
\end{array}
\qquad
\begin{array}{ccccc}
 & & & & C_{uu} = \max\left(u^2S - X, 0\right) \\
 & & & \nearrow & \\
 & & C_u & & \\
 & \nearrow & & \searrow & \\
C & & & & C_{ud} = \max\left(udS - X, 0\right) \\
 & \searrow & & \nearrow & \\
 & & C_d & & \\
 & & & \searrow & \\
 & & & & C_{dd} = \max\left(d^2S - X, 0\right)
\end{array}
$$

To find the option prices, just work back from the end:

$$C_u = \frac{1}{R}\left[pC_{uu} + (1-p)\,C_{ud}\right]$$

$$C_d = \frac{1}{R}\left[pC_{ud} + (1-p)\,C_{dd}\right]$$

$$C = \frac{1}{R}\left[pC_u + (1-p)\,C_d\right].$$

If you prefer, you can substitute in to reexpress the answer as

$$C = \frac{1}{R^2}\left[p^2C_{uu} + 2p\,(1-p)\,C_{ud} + (1-p)^2\,C_{dd}\right]$$

or

$$
\begin{aligned}
C \quad = \quad & \frac{1}{R^2}\left[p^2 \max\left(u^2S - X, 0\right) + \right. \\
& \left. +2p\,(1-p)\max\left(udS - X, 0\right) + (1-p)^2 \max\left(d^2S - X, 0\right)\right].
\end{aligned}
$$

Note:

1) All the things an option price should depend on is there. The stock price S, the strike price X, the volatility u, d the interest rate R and the time–number of periods to expiration.

2) This is a perfectly practical, real-world way to find option prices. If you let $u =$ rise 1/16 and $d =$ decline 1/16, in two periods, you have the possibilities rise 1/8, stay the same, or decline 1/8. With more periods, you can allow virtually any amount of stock price movement. This is in fact how much option pricing is really done, when you want answers more accurate than the simple Black-Scholes formula given below.

8.4 To Black-Scholes

Suppose you increase the number of steps in the binomial model, making each step a smaller size. When you take limits the right way, a beautiful formula emerges:

Black-Scholes Formula:

$$C = S\mathcal{N}(d_1) - Xe^{-rT}\mathcal{N}(d_2)$$

where

$$d_1 \equiv \frac{\ln\left(\frac{S}{X}\right) + \left(r + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}; \;\; d_2 = d_1 - \sigma\sqrt{T}$$

$$\mathcal{N}(x) = \text{ area under normal distribution up to } x$$

$$r = \text{ continuously compounded interest rate } (e.g.\ 0.02)$$

$$\sigma = \text{ standard deviation of stock returns (e.g., 0.17)}$$

This formula looks mysterious initially, but it is derived *exactly* as our binomial formula was. Note a few things.

1) If the option is way in the money, $S >> X$, $\mathcal{N}(\infty) = 1$ so $C \to S - Xe^{-rT}$.

2) If the option is way out of the money, $S << X$, $\mathcal{N}(-\infty) = 0$ so $C \to 0$.

3) The option price is again a *deterministic* function of the stock price, with r, T, σ, X as parameters. Again, the option is priced by arbitrage.

4) σ is not observable. It is the conditional volatility, what traders think the stock's volatility will be. (In the binomial model, we had to take a stand on u, d, or how much traders thought that the price could go up or down.) The Black-Scholes formula is often used the way the definition of yield is used–not to find the correct price of an option, but to find a common basis on which to quote prices of options with widely differing strike prices and times to maturity. Thus, the *implied volatility* is the value of σ that makes a given option satisfy the Black Scholes formula perfectly. Options are often quoted by their volatilities, not their actual prices.

5) Intuition. The present value of receiving the stock in the future is S, today's value. Thus, the first term looks like the present value of the stock times the probability that it will end up in the money, i.e. that you'll get it. The second term looks a lot like the present value of the strike price you have to pay for the stock, again times a probability that you'll get it. These

aren't the same, and they aren't exactly probabilities. They are the *risk-neutral* probabilities. Thus, the Black-Scholes formula can be interpreted as a risk-neutral probability formula, just like the binomial formula.

5) The *Hedge Ratio* is the slope of the call option price, or the derivative of the Black-Scholes formula. An interesting fact is that

$$\text{Hedge ratio} = \frac{\partial C}{\partial S} = \mathcal{N}(d_1)$$

Note that this slope varies as the stock price varies and as time passes. As with duration, to truly synthesize an option, you have to use a *dynamic* trading strategy in which the portfolio position is adjusted all the time.

The hedge ratio is also useful in another context. If you have to hold a lot of stock (say, for a client, or because you are underwriting the offering), the hedge ratios tell you how many options to write to offset the risk of price changes of the stock.