INTRODUCTION TO THE MARKOWITZ MODEL (FIRST DRAFT, LIKELY TO CHANGE)

JOEL FRIEDMAN

Contents

1. Introduction

The point of this article is to describe some aspects of the *Markowitz model* that are relevant to Math 441. We briefly review some basic probability theory to motivate the model, and discuss the quadratic programming problems related to it. In 2017, we referred to the textbook by Vanderbei (see the course website), Chapter 24, for the model, and to the article I wrote "Linear Complementarity and Mathematical (Non-Linear) Programming" (posted on the course website).

2. Models

In financial mathematics, many models to predict the future behaviour of a system are fundamentally different than many such models in physics.

In physics, many models assume $\vec{F} = m\vec{a}$, and then assume that the force $\vec{F}$ comes from one or two sources. For example, the "n-body problem" assumes you have n bodies (e.g., planets or stars), modeled by point masses, where the force is due to a single source, namely gravitation from the other $n-1$ bodies. This gives a

Research supported in part by an NSERC grant.

stunning explanation of why we observe roughly elliptical orbits of planets around our sun. This model, with three or more bodies and with certain initial conditions, can yield trajectories that seem very chaotic. Yet the model is formulated on a precise understanding of one very simple driving force.

In financial mathematics, systems usually have too many driving forces to (realistically) model. Models typically extract a few simple quantities from a complicated system, and write down a simple formula—such as a utility function to be maximized—based on these quantities. Such formulas don't individually model each driving force of a complex system. Instead, the test of such a formula or model is how well it predicts something we want to know. The *Markowitz model* is a quintessential example of such financial models.

3. Probability: Expected Value and Variance

In this section we review some standard notion and ideas of probability.

Probability theory can be used to model aspects of a future event. Say that you predict that the rainfall in Vancouver tomorrow will be between 20 to 25 centimetres, rounded to the nearest centimetre, according to the following probabilities:

Event number	1	2	3	4	5	6
Rainfall in cm	20	21	22	23	24	25
Probability percent	10	15	50	10	3	12
Probability, p	0.10	0.15	0.50	0.10	0.03	0.12

Let X denote the *random variable* of rainfall tomorrow in Vancouver. The above table says that the *probability space* is divided into six possible *events*; event number 1 occurs with probability $p_1 = 0.10$, and in this event X takes the value $x_1 = 20$; similarly $p_2 = 0.15$ for the second event, where X takes the value $x_2 = 21$; the other possible values of X are

$$x_3 = 22,\ x_4 = 23,\ x_5 = 24, x_6 = 25,$$

which occur with respective probabilities

$$p_3 = 0.50,\ p_4 = 0.10,\ p_5 = 0.03,\ p_6 = 0.12.$$

One could imagine a more refined version of X, where we have 51 events that specify X rounded to the nearest millimetre, between 200 mm and 250 mm; this would be a much larger table.

Two useful measurements of a random variable are its *expected value* and its *variance* The expected value of X, denoted $\overline{X}$ (also denoted $E[X]$ in the literature), is defined to be

$$\overline{X} \stackrel{\text{def}}{=} \sum_{i=1}^{6} p_i x_i = p_1 x_1 + \cdots + p_6 x_6,$$

$$= (0.10)20 + (0.15)21 + (0.50)22 + (0.10)23 + (0.03)24 + (0.12)25 = 22.17\ .$$

Its varaince, denoted $\mathrm{Var}(X)$, is defined to be

$$\mathrm{Var}(X) \stackrel{\text{def}}{=} \sum_{i=1}^{6} p_i \bigl(x_i - \overline{X}\bigr)^2 = p_1\bigl(x_1 - \overline{X}\bigr)^2 + \cdots + p_6\bigl(x_6 - \overline{X}\bigr)^2$$

$$= (0.10)(20 - 22.17)^2 + (0.15)(21 - 22.17)^2 + \cdots + (0.12)(25 - 22.17)^2 = 1.8211$$

Of course, if there were n events instead of 6 for X, then the summations above would range over i from 1 to n instead of to 6.

The expected value and variance of X tell us something about X, but hardly the whole story. The variance is always non-negative, and is zero iff the random variable takes on the same value (over all events).

4. Fundamental Identities, Part I

Here are some identities we will use to make computations.

(1) If X, Y are random variables, then

$$\overline{X+Y} = \overline{X} + \overline{Y} \ . \tag{1}$$

(2) If X is a random variable, and $a \in \mathbb{R}$ a real number, then

$$\overline{aX} = a\overline{X}.$$

(3) These first two identities imply that for $a, b \in \mathbb{R}$ and random variables X, Y we have

$$\overline{aX + bY} = a\overline{X} + b\overline{Y}$$

[i.e., "expected value" is a *linear map* from (real-valued) random variables to $\mathbb{R}$].

(4) In the above formula, and the special case where $Y = 1$, i.e., Y is the random variable whose value 1 on all events, we have

$$\overline{aX + b} = a\overline{X} + b \ . \tag{2}$$

(5) If X is a random variable, then

$$\mathrm{Var}(X) = \overline{\left(X - \overline{X}\right)^2};$$

in other words, the variance is the expected value of the random variable $(X - \overline{X})^2$.

(6) The first three identities imply that

$$\mathrm{Var}(X) = \overline{X^2 - 2X\overline{X} + \left(\overline{X}\right)^2} = \overline{X^2} - 2\overline{X}\,\overline{X} + \left(\overline{X}\right)^2$$

which yields the extremely useful identity

$$\mathrm{Var}(X) = \overline{X^2} - \left(\overline{X}\right)^2.$$

(7) The above identities also show that if X is a random variable, and $a, b \in \mathbb{R}$, then

$$\mathrm{Var}\, aX + b = a^2 \,\mathrm{Var}\, X. \tag{3}$$

So (3) and (2) tell us how the variance and expected values change when we "scale a random variable by a" and "shift (or translate) it by b," i.e., take X and form a new random variable $aX + b$.

Hence, if you gain \$100 for every centimetre of rainfall over 18 cm in the example of Section 3, then your gain in dollars is $G = 100X - 1800$, whose expected value is $2217 - 1800 = 417$ (dollars), and whose variance is $18,211$ (dollars squared).

5. A Simple Example of Financial Instruments

To motivate and illustrate the Markowitz model, we will start with some simple, contrived data. We imagine that we can spend \$1000 to buy any one of five financial instruments—four stocks and one bond—and at the end of one month there are only four events, i.e., four conceivable future scenarios, each occuring with probability 0.25, where the increase in price of the instruments are as follows:

	Event 1	Ev 2	Ev 3	Ev 4	avg	var
Prob, p	1/4	1/4	1/4	1/4		
C (Coke)	8	8	12	12	10	4
P (Pepsi)	8	8	12	12	10	4
A (Amazon)	8	12	8	12	10	4
N (Nozama)	12	8	12	8	10	4
B (Bond)	9	9	9	9	9	0

Notice that:

(1) C, P, A, N all have expected values 10 and variance 4.
(2) B has expected value 9 and variance 0; in other words, B, the increase in price of the bond, has a certain value of 9.
(3) $C = P$, meaning that these random variables have the same values in all events. We have $2C = C + P = 5C - 3P$ (the latter represents the change in value upon buying 5 lots of Coke and short selling 3 lots of Pepsi); they all have expected value 20 (dollars, which represents dollars per \$1000 of investment) and variance 16 (dollars squared, which represents dollars squared per \$1000 of investment).
(4) Stocks in the same industries tend to move together. So while Coke and Pepsi are competitors, they tend to move alike.
(5) $A + N = 20$, meaning $A + N$ is 20 in all possible events.
(6) Nozama is Amazon spelled backwards, to indicate that it is a fictional stock that is perfectly negatively correlated with Amazon and would not exist in the real world. Nozama is the perfect "hedge" for Amazon, meaning a sure (and presumably legal) way to get an increase of \$20 for a \$2000 investment, i.e., 1% or 100 basis points[1] over one month.
(7) B, the increase in bond price, is lower than those of the stocks, but is a sure bet. This is typical of financial instruments.
(8) C and A are independent: each takes on the values $8, 12$ with 0.5 probability, and the joint distribution of (A, C) any of $(8, 8), (8, 12), (12, 8), (12, 12)$ is exactly $0.25 = (0.50)^2$.
(9) $A + C = 16$ with probability 0.25, $A + C = 20$ with probability 0.50, and $A + C = 24$ with probability 0.50.
(10) Hence $A + C$ has expected value 20 and variance

$$(0.25)(16-20)^2 + (0.50)(20-20)^2 + (0.25)(24-20)^2 = 8.$$

[1]A *basis point* is 1/100-th of one percent.

6. Risk and the Markowitz Model

The assumption in the *Markowitz* model is that from all the portfolios available to you (in a market composed of financial instruments like stocks and bonds), the most desirable one is the one that maximizes the utility function

$$\mathrm{Utility}(R,\mu) \stackrel{\mathrm{def}}{=} \overline{R} - \mu\,\mathrm{Var}(R), \tag{4}$$

where R is the return of your portfolio, and μ is parameter that depends on your *tolerance of risk* when risk is taken to mean the variance; furthermore, the choice of portfolio of n instruments, represented by random variables $R_1,\ldots,R_n$, is a function

$$R = R(\vec{w}) = w_1R_1 + \cdots + w_nR_n$$

[Vanderbei's textbook uses x_i's instead of w_i's; Wikipedia currently uses w_i's], where $\vec{w} \in \mathbb{R}^n$ which may be subject to certain restrictions, including budget restrictions, legal restrictions (e.g., under what circumstances can one short sell, i.e., when can a w_i be negative?), hedging restrictions, etc.

We remark that $\mu = 0$ means that you are indifferent to risk (meaning variance in your portfolio), and $\mu < 0$ means that you are willing to "pay for risk," which you might apply to people at the River Rock Casino. In portfolio theory, $\mu > 0$ where μ large indicates a great aversion to variance or "risk."

The utility, U, has the same units of R in (4), and constant μ has units of $1/R$. For example, if R is in units of

$$\frac{(\text{dollars of return})}{\big((\text{dollars of investment})\,(\text{months})\big)},$$

then the units of μ becomes

$$\frac{\big((\text{dollars of investment})\,(\text{months})\big)}{(\text{dollars of return})}.$$

Similarly, if you ignore the period of time, the units of μ becomes

$$\frac{(\text{dollars of investment})}{(\text{dollars of return})};$$

in this case, since μ, R, U all represent one type of dollars per another type of dollars, it may be tempting to view μ, R, U as dimensionless, although this is point of view is unlikely to be realistic (see the exercises).

6.1. Problems With Variance Aversion.

There are a standard set of remarks regarding people who are extremely risk averse when risk is measured as variance. For example, if L is a lottery ticket that with probability $p = 1/1000$ will be worth \$1000, and probability $q = 1 - p = 999/1000$ will be worth 0, then

$$\overline{L} = 1, \quad \mathrm{Var}(L) = \overline{L^2} - \big(\overline{L}\big)^2 = 1000 - 1 = 999.$$

It follows that if your aversion to variance, μ, is at least $1/999$ (in units of 1/dollars), then L alone has a negative utility, and you'd be happier to discard the lottery ticket (giving you a utility equal to zero). For similar reasons, for any $\mu > 0$, even a μ extremely small, it is possible to construct a lottery ticket where you cannot lose money in any scenario, and yet the ticket has a negative utility in terms of (4).

For similar reasons, someone who is very averse to variance might buy an umbrella from the UBC Bookstore in the middle of a hot day in August, with no rain in the forecast for two weeks, as a "hedge" against the highly unlikely event of rain.

The above remarks are standard remarks in mathematical finance, and apply to aversion to "risk" where "risk" is defined as variance or many other (seemingly reasonable) definitions of risk measured by how the random variable differs from its expected value. For this reason people people in finance introduce difference measures of risk, such as VaR (value at risk) and many others. Financial markets are driven by many forces, resulting in a large number of possible scenarios that are difficult to predict; attempting to extract from such markets some simple measurements that reliably describe "risk" and that work well in most situations is not easy.

Still, we will see that the Markowitz model (4) is extremely useful, once we understand when it works well and what are its limitations.

6.2. **The One Instrument Markowitz model.** To understand the Markowitz model, consider portfolios based on the instruments C, P, A, N, B of Section 5.

To begin, consider portfolios based on the single instrument A. Imagine that someone hands you 5 lots of Amazon (A) above, so that $R = 5A$, your utility is

$$\overline{R} - \mu \operatorname{Var}(R) = 5\overline{A} - \mu 25 \operatorname{Var}(A) = 5 \cdot 10 - \mu 25 \cdot 4 = 50 - 100\mu$$

(so here μ is in units of $(\text{dollar})^{-1}$). So if you are very risk averse, with $\mu > 2$, it may seem that you'd have a higher utility (of zero) by discarding your shares. This is absurd.

For this reason the Markowitz model is like the n-body problem for $n = 1$: the case of a single instrument is too degenerate to give interesting results. And similar to the $n = 1$ case of the n-body problem, there are two remedies:

(1) study the case of two instruments, or
(2) study a single instrument, but make an additional assumptions about the system of a single instrument.

[The 1-body problem becomes interesting when we assume there is a force acting on the single body, such as a *central force*. The 2-body problem, under Newtonian gravitation, can be reduced to a 1-body central force problem. We'll see something similar with the Markowitz model.]

So now imagine that we have a supply of cash on hand, say ten thousand dollars, which next month will have the same value (of ten thousand dollars) in all events (i.e., future scenarios). And imagine that the shares of Amazon, A, that we have must be bought from this supply of cash; by symmetry, assume that if we discard a share of Amazon then we recover the cost of this share. Now "discarding" is really "selling," and it makes sense that we might sell shares of Amazon (under some conditions).

Another way to describe the above idea is by a simple two-instrument Markov model.

6.3. **Some Two Instrument Markowitz Models.** Another way to way to desribe the above one-instrument Markowitz model is with a two-instrument Markowitz model where one instrument is a random varible M representing "money" (or "cash"). Hence $\overline{M} = 0$ and $\operatorname{Var}(M) = 0$, and for simplicity we set M to be units of

\$1000. Then our portfolio options are $R = w_1 A + w_2 M$, subject to $w_1 + w_2 = 10$ and $w_1, w_2 \geq 0$. We get the optimization problem

(1) maximize $U(\vec{w}, \mu) = 10w_1 - 4w_1^2\mu$,
(2) subject to $w_1 + w_2 = 10$, and
(3) $w_1, w_2 \geq 0$, i.e., $0 \leq w_1 \leq 10$.

The solution (see the computation in the next paragraph) is

$$w_1^* = \min\bigl(10, (5/4)\mu^{-1}\bigr), \quad w_2^* = 10 - w_1^* \ . \tag{5}$$

This becomes a lot more reasonable.

In more detail, to find the optimum feasible solution, take the partial derivative of

$$U(w_1) = 10w_1 - 4\mu w_1^2$$

with respect to w_1 (holding any other variables fixed—U above depends on μ, which we view as fixed)

$$\frac{\partial U}{\partial w_1} = 10 - 8\mu w_1.$$

Hence $\partial U/\partial w_1$ is positive for $0 \leq w_1 < 10/(8\mu)$ and is negative for $w_1 > 10/(8\mu)$. Since the region of feasibility is $0 \leq w_1 \leq 10$, we see that

(1) if $10 \leq 10/(8\mu)$, then $U(w_1)$ for $0 \leq w_1 \leq 10$ is maximized at $w_1^* = 10$; and
(2) if $10 > 1/(8\mu)$, then $U(w_1)$ is maximized at $w_1^* = 10/(8\mu)$.

This gives us (5).

Often there are financial instruments whose risk is close to zero, such as federally issued US or Canadian Treasury bonds or bills. [These governments can always avoid default or underpayment simply by printing more money under extreme conditions; such extreme circumstances might make your entire model chaotic, but any model of investment makes some assumptions and becomes chaotic when the assumptions are violated.] So more typically you would have a bond, such as B in Section 5, with a smaller expected rate of return but little or no variance. In this case you get an optimization problem for $R = w_1 A + w_2 B$, which is only slightly more complicated than the above model, since $B = 9$ is a constant random variable:

(1) maximize $U(\vec{w}, \mu) = 10w_1 + 9w_2 - 4w_1^2\mu$,
(2) subject to $w_1 + w_2 = 10$, and
(3) $w_1, w_2 \geq 0$, i.e., $0 \leq w_1 \leq 10$.

[The formula for $U(\vec{w}, \mu)$ above is based on the fact that $B = 9$ is constant, so (3) we have $\operatorname{Var} w_1 A + 9 = w_1^2 \operatorname{Var} A$.] Eliminating $w_2 = 10 - w_1$ we have

$$U(\vec{w}, \mu) = 10w_1 + 9w_2 - 4w_1^2\mu = 10w_1 + 9(10 - w_1) - 4w_1^2\mu = 90 + w_1 - 4\mu w_1^2.$$

To find the optimum of this we take the partial derivative of

$$U(w_1) = 90 + w_1 - 4\mu w_1^2$$

with respect to w_1 (holding any other variables fixed—U above depends on μ, which we view as fixed)

$$\frac{\partial U}{\partial w_1} = 1 - 8\mu w_1.$$

Hence $\partial U/\partial w_1$ is positive for $0 \leq w_1 < 1/(8\mu)$ and is negative for $w_1 > 1/(8\mu)$. Since the region of feasibility is $0 \leq w_1 \leq 10$, we see that

(1) if $10 \leq 1/(8\mu)$, then $U(w_1)$ for $0 \leq w_1 \leq 10$ is maximized at $w_1^* = 10$; and
(2) if $10 > 1/(8\mu)$, then $U(w_1)$ is maximized at $w_1^* = 1/(8\mu)$.

Hence we get the optimum feasible solution:

$$w_1^* = \min\bigl(10, (1/8)\mu^{-1}\bigr), \quad w_2^* = 10 - w_1^* \ . \tag{6}$$

7. Fundamental Identities, Part II: Covariance and Correlation

In the previous section, the variables M and B were constants; this meant that

$$\mathrm{Var}(w_1A + w_2M) = w_1^2 \,\mathrm{Var}(A) = \mathrm{Var}(w_1A),$$

and similarly for M replaced with B. However, if we consider a two-instrument portfolio $R = w_1A + w_2X$ where X is not a constant, then $\mathrm{Var}(R)$ depends on how A and X are related; we always have $\overline{R} = w_1\overline{A} + w_2\overline{X}$ by (1), but the situation with variance is different. The point of this section is to give some precise statements.

One important principle from Section 5 is that although $A+C$ and $2A$ (or $2C$ or $C+P$) all have the same expected values, the variance of $A+C$ is 8 while the others have variance 16. This hints at the general principle that if a collection of financial instruments have the same expected return, a portfolio with a lot of independent instruments will have a smaller variance than an undiversified portfolio. (Whether or not there exists true independence between instruments is another matter...)

Let us make a more precise study the variance of a sum of random variables. The convenient identity

$$\mathrm{Var}(R) = \overline{R^2} - \bigl(\overline{R}\bigr)^2$$

implies that

$$\mathrm{Var}(X+Y) = \overline{(X+Y)^2} - \bigl(\overline{X} + \overline{Y}\bigr)^2$$

which, after expanding terms (see the exercises), becomes

$$= \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\bigl(\overline{XY} - \overline{X}\ \overline{Y}\bigr).$$

In other words, the variance of $X+Y$ is the sum of their variances plus a correction term; we define the *covariance of X and Y* to be

$$\mathrm{Cov}(X,Y) \stackrel{\mathrm{def}}{=} \overline{XY} - \overline{X}\ \overline{Y}\ ,$$

a quantity which describes the correction term, i.e., we have

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\,\mathrm{Cov}(X,Y)\ . \tag{7}$$

Linear algebra (specifically the *Cauchy-Schwarz inequality*) implies that

$$-\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)} \le \mathrm{Cov}(X,Y) \le \sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)},$$

and that the *correlation of X and Y*

$$\mathrm{Corr}(X,Y) \stackrel{\mathrm{def}}{=} \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\,\mathrm{Var}(Y)}}$$

(assuming that $\mathrm{Var}(X)$ and $\mathrm{Var}(Y)$ are both nonzero), which is therefore between -1 and 1, represents the cosine between the vectors $X - \overline{X}$ and $Y - \overline{Y}$ measured by a "dot product" that is weighted according to event probabilites.

In the case at hand we have

$$\mathrm{Corr}(A,N) = -1, \quad \mathrm{Corr}(A,C) = 0, \quad \mathrm{Corr}(A,A) = 1,$$

and the fact that $\mathrm{Var}\,A = \mathrm{Var}\,N = \mathrm{Var}\,C = 4$ implies that

$$\mathrm{Var}(A+N) = 0, \quad \mathrm{Var}(A+C) = 8, \quad \mathrm{Var}(A+A) = \mathrm{Var}(2A) = 4\,\mathrm{Var}(A) = 16.$$

If $\mathrm{Cov}(X,Y)=0$ then we say that X *and* Y *are uncorrelated*, and we have

$$\mathrm{Var}(X+Y)=\mathrm{Var}(X)+\mathrm{Var}(Y)$$

which is called *Pythagoras's theorem* (in the context of weighted dot products). If X,Y are *independent*[2] then $\mathrm{Cov}(X,Y)=0$, but the condition $\mathrm{Cov}(X,Y)$ is much weaker and can be tested (to some extent) experimentally.

8. The Usual Formulation of the Markowitz Model

A generalization of (7) is

$$\mathrm{Var}(R_1+\cdots+R_n)=\sum_{i,j=1}^{n}\mathrm{Cov}(R_i,R_j),$$

or, since $\mathrm{Cov}(X,X)=\mathrm{Var}(X)$ and $\mathrm{Cov}(R_i,R_j)=\mathrm{Cov}(R_j,R_i)$, we often write

$$\mathrm{Var}(R_1+\cdots+R_n)=\sum_{i=1}^{n}\mathrm{Var}(R_i)+2\sum_{i<j}\mathrm{Cov}(R_i,R_j).$$

Upon scaling each R_i, we get

$$\mathrm{Var}(w_1R_1+\cdots+w_nR_n)=\sum_{i,j=1}^{n}w_iw_j\,\mathrm{Cov}(R_i,R_j),$$

or the equivalent formula

$$\mathrm{Var}(w_1R_1+\cdots+w_nR_n)=\sum_{i=1}^{n}w_i^2\,\mathrm{Var}(R_i)+2\sum_{i<j}w_iw_j\,\mathrm{Cov}(R_i,R_j).$$

The utility in the Markowitz model becomes

$$U(\vec{w},\mu)=\sum_{i=1}^{n}w_i\overline{R_i}-\mu\left(\sum_{i=1}^{n}w_i^2\,\mathrm{Var}(R_i)+2\sum_{i<j}w_iw_j\,\mathrm{Cov}(R_i,R_j)\right). \tag{8}$$

Typically we make some assumptions about the feasible region of $\vec{w}\in\mathbb{R}^n$ (or perhaps $\vec{w}\in\mathbb{Z}^n$ if we insist on purchasing a whole number of lot sizes).

9. Practical Aspects of the Markowitz Model

The Morkowitz model is important because of its simplicity and ability to give a reasonable model for how to combine various ("risky") assets into a portfolio with an acceptable amount of "risk."

We have seen that measuring risk as variance has its shortcomings (see Subsection 6.1), such as being able to invent lottery tickets which we would rather discard instead of keep. This is a reflection of the fact that variance aversion is not a fundamental law of markets, such as the law of gravitation in physics. Instead, the Markowitz model is a way to take a markets that are driven by an extremely large set of forces or factors, and to take some simple data about these markets that gives some realistic idea of how to optimize portfolios.

Here are some of the important properties of the Markowitz model:

---

[2]meaning that for all a,b, the probability that $X=a$ and $Y=b$ simultaneously is precisely the probability that $X=a$ times that of $Y=b$

(1) if $\vec{w}^*$ has the optimum utility of a feasible portfolio (for any set of constraints) and $\vec{w}$ is another feasible portfolio, then $\overline{R(\vec{w}^*)} = \overline{R(\vec{w})}$ implies that
$$\mathrm{Var}(R(\vec{w}^*)) \le \mathrm{Var}(R(\vec{w}))$$
(assuming $\mu > 0$);
(2) similarly if $\vec{w}^*$ has the optimum utility of a feasible portfolio (for any set of constraints) and $\vec{w}$ is another feasible portfolio, then $\mathrm{Var}(R(\vec{w}^*)) = \mathrm{Var}(R(\vec{w}))$ implies that
$$\overline{R(\vec{w}^*)} \ge \overline{R(\vec{w})};$$
(3) the expected values $\overline{R_i}$ and variances/covariances $\mathrm{Cov}(R_i, R_j)$ are often easy to estimate;
(4) with $\mu > 0$, the Markowitz model gives a "convex" quadratic optimization problem, which is readily solvable; with $\mu < 0$ the optimization isn't generally easy to solve, since one generally has to check all extreme parts of the feasible region, e.g., all the vertices of the region (see the exercises and examples given in class);
(5) the formulas involving expected values and variances/covariances make no assumptions about the random variables except that these expected values and variances/covariances are not infinite;
(6) the Markowitz model exhibits many known principles of investing, and can be used to test other princples; even if the model is not exact, it may indicate what to expect [you certainly wouldn't want to make investments based on principles that fail for the Markowitz model unless you know why these principles fail on the Markowitz model and wouldn't fail in the real world...].

For this reason, even though the Markowitz model can be improved upon, it is an excellent place to start.

10. Closing Remarks, Fat Tails, and the Black-Scholes Volatility Smile

After Newton's theory of gravitation was known, inaccuracies were discovered when Mercury passed behind the sun. These inaccuracies were later corrected by Einstein's theory of general relativity. So even if a first theory has some inaccuracies, it can be a good place to start.

As mentioned before, financial models are not based a few fundamental principles of how markets work, since markets are affected by a large number of forces that are difficult to predict. Rather such models try to focus on some quantities that one can measure, in a model that is simple enough to use for some predictions. You can't expect any model of a financial market to be perfectly accurate all the time.

A terrific example of an imperfect but useful model is the *Black-Scholes* formula. It is a model that predicts the value of a simple type of stock option (a European call or put, i.e., the option to purchase or sell a stock at a given price at some future given point in time). The formula is based on *geometric Brownian motion*, which—like a Gaussian distribution (bell curve)—is known to be unrealisitic: real world stock price distributions are known to have *fat tails*, meaning that the probability of being far from the expected value is much greater a Gaussian distribution would have (with the same variance). The result is that the *volatility* of a stock, which is one input to the Black-Scholes formula, appears to increase as the option's *strike* (i.e., price of purchase/sell) moves away from the expected value. Plotting this volatility as a function of the strike for known market values of a European option shows a well-known *volatility smile* of stock (whose volatility is supposed to be an intrinsic property of the stock that doesn't change as a function of the option's strike).

Yet the Black-Scholes formula is useful for many reasons. For one, it is a simple framework for pricing derivatives, and one can try to modify the framework to get more accurate pricing. For another, if you know the prices of a few European options and plot the volatility smile, you may be able to estimate a fair price of other such options. In this case, of course, trying to interpolate the data is likely a reasonable thing to do, while trying to extrapolate many be tricky; its not even clear that stock prices should be modeled by any distribution with a finite variance.

11. Exercises

(1) Consider the solution (5) to the Markowitz utility of the model $R = w_1 A + w_2 M$ subject to $w_1, w_2 \geq 0$ and $w_1 + w_2 = 10$.
   (a) What is w_1^* for the values $\mu = .05$, $\mu = 1$, and $\mu = 1000$?
   (b) Explain intuitively—in terms of μ representing risk aversion—why w_1 is very close to 0 for $\mu = 1000$.
   (c) For the analogous solution (6) for the model $R = w_1 A + w_2 B$ (subject to the same conditions), what is the value of w_1^* for the value $\mu = .05$?
   (d) Explain intuitively—in terms of the difference between the models $w_1 A + w_2 M$ and $w_1 A + w_2 B$—why for $\mu = .05$ the value of w_1 is 10 for one of the models and less than 10 for the other.

(2) Compute the Markowitz Utility $U(w_1, w_2, \mu)$ for the portfolio $R = w_1 A + w_2 N$. [Note the formula $\mathrm{Corr}(A, N) = -1$ in the Section 7, and note that the formulas in Section 8 imply that

$$\mathrm{Var}(w_1 X + w_2 Y) = w_1^2 \,\mathrm{Var}(X) + 2 w_1 w_2 \,\mathrm{Cov}(X, Y) + w_2^2 \,\mathrm{Var}(Y)\ .]$$

Then find the optimum feasible solution for this model under the conditions $w_1, w_2 \geq 0$ and $w_1 + w_2 = 10$.

(3) Compute the Markowitz Utility $U(w_1, w_2, \mu)$ for the portfolio $R = w_1 C + w_2 P$. Then find **all** optimum feasible solutions for this model under the conditions $w_1, w_2 \geq 0$ and $w_1 + w_2 = 10$.

Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4, CANADA, and Department of Mathematics, University of British Columbia, Vancouver, BC V6T 1Z2, CANADA.

E-mail address: jf@cs.ubc.ca or jf@math.ubc.ca

URL: http://www.math.ubc.ca/~jf