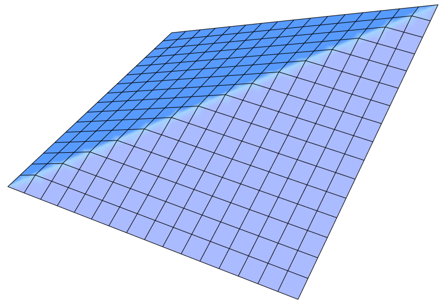


ACTSC 372: Corporate Finance II

Prof. Hany Fahmy • Spring 2014 • University of Waterloo
TR 10:00AM–11:20AM in DC 1351

Transcribed, LaTeXed, and edited by Rob Zimmerman

Note: These are personal lecture notes, not official course materials. They may contain errors (which by default you should attribute to me, Rob Zimmerman, rather than the course instructor). The permanent home of these — and all of my other lecture notes — is `https://rob-zimmerman.github.io/coursenotes`.

Contents

1. Foundations of Finance and Choice Under Uncertainty

Lecture 1 — Tuesday, May 06, 2014

The three main sectors of the economy and their roles in market activity are summarized in Figure 1.

Consumers generate income y, where y = savings + consumption. Suppose s = \$100,000. This amount goes either into **financial intermediation** or into the **financial market**. Our course deals with the financial market. Any market has two sides: demand and supply. Their interaction determines the **market outcome**, which has both price and quantity.

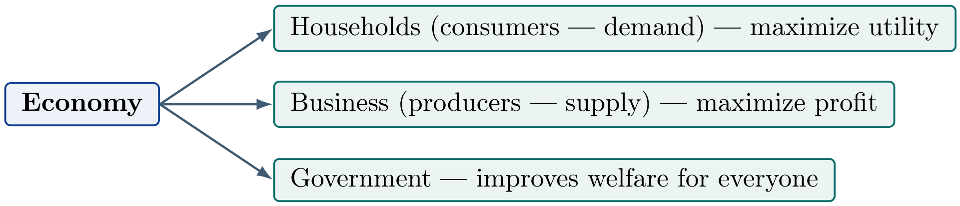


FIGURE 1

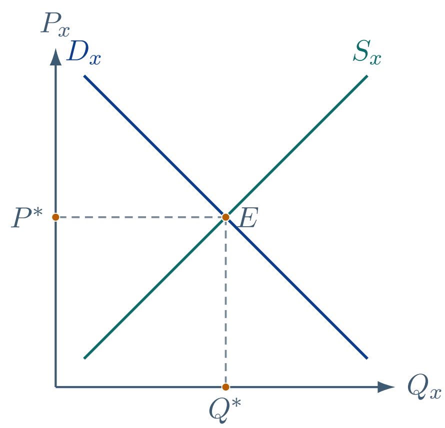


FIGURE 2

In Figure 2, P^* is the **market clearing price** for a good x and Q^* is the **market clearing quantity**. It is at these precise levels that demand meets supply. In a financial market, the number of shares outstanding may be approximately fixed over a short interval, but the supply curve is not thereby irrelevant: a shift in demand generally changes both the equilibrium price and the quantity traded. The figure records this joint determination rather than attributing every price movement to demand alone.

There are other kinds of markets besides the financial market. The **labour market** has labour demand N^d and labour supply N^s (number of hours devoted to work), and its market outcome determines the wage and equilibrium labour hours. The **housing market** has demand D^h and supply S^h, which determine P^h and Q^h. There is also a **market for goods and services**. The **money market** has demand M^d and supply M^s, which determine P^m (the market interest rate) and the quantity of money. The exchange of funds for securities in the financial market is summarized in Figure 3.

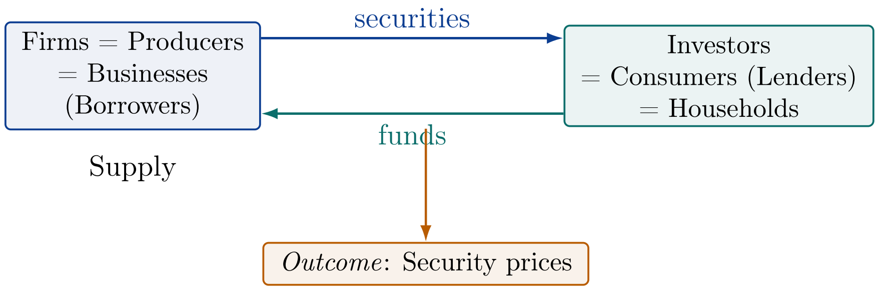


FIGURE 3

Lecture 2 — Thursday, May 08, 2014

2. Choice Theory Under Uncertainty

Expected Utility and Lotteries

In standard choice theory, the consumer chooses between certain alternatives, so no probability is involved. Let x_1 and x_2 be the amounts of two goods, and let $U(x_1, x_2)$ denote the consumer's **Cobb–Douglas utility function**. Utility is measured in units of utility, or "utils." For example, if $U(x_1, x_2) = x_1 \cdot x_2$, and $x_1 = 10$ units and $x_2 = 20$ units, then $U = 200$ utility units. At this stage there is no uncertainty. The two utility curves in Figure 4 are both convex to the origin.

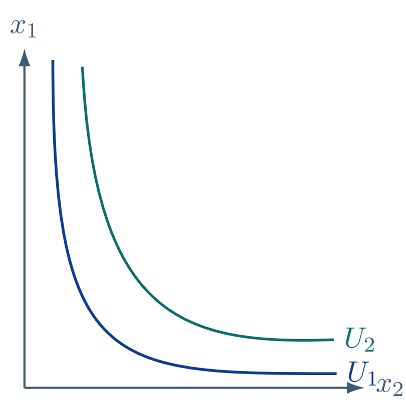


FIGURE 4

Marginal utility, denoted by MU_x, is the change in utility generated by an additional unit of good x. Total utility, denoted by TU_x, is the total satisfaction from all units consumed of good x;

a typical increasing, concave total utility curve is shown in Figure 5.

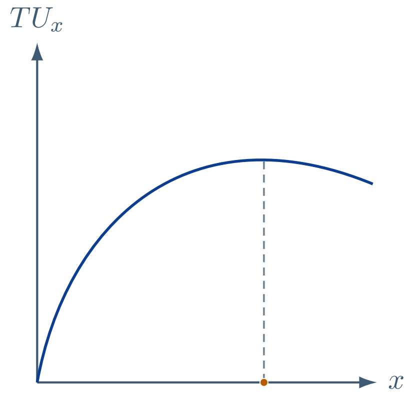


FIGURE 5

Here

$$MU_1 = \frac{\Delta TU_1}{\Delta x_1} = \frac{18-10}{2-1} = 8$$

utility units, while

$$MU_2 = \frac{25-18}{3-2} = 7$$

utility units, and so on.

More rigorously,

$$MU_{x_1} = \frac{\partial TU}{\partial x_1},$$

and it is decreasing as x_1 increases. This is the **law of diminishing marginal utility** from consumption, sketched in Figure 6.

Let us recall mathematical expectation. Suppose r_A denotes the return on Stock A, with

r_A units	Probability
2%	0.1
10%	0.8
20%	0.1

Then the expected return is

$$\mathbb{E}[r_A] = 2\%(0.1) + 10\%(0.8) + 20\%(0.1) = 10\%.$$

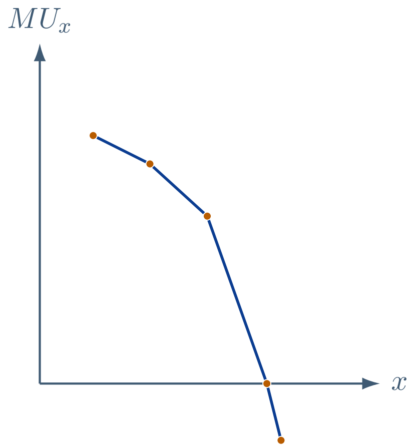


FIGURE 6

We model **risk** as a lottery. Denote the risk associated with a given lottery by $\tilde{z}$. For example,

$$\tilde{z} = \begin{cases} +\$1, & \text{with probability } 50\%, \\ -\$1, & \text{with probability } 50\%. \end{cases}$$

Clearly $\tilde{z}$ is a random variable. An equivalent notation is $\tilde{z}(z_1, z_2; \alpha)$, where α is the probability of the first outcome. In this example we may write $\tilde{z}(+1, -1; 0.5)$. We also denote **wealth** by $\tilde{w}$, and let z_1, z_2 represent the monetary outcomes of the lottery.

This notation is the starting point for the course: once wealth is random, investor preferences must be expressed over lotteries rather than over certain outcomes.

Example 2.1 Suppose you have \$4000 worth of goods at home and wish to bring \$8000 worth of goods from abroad by sea freight. The ship sinks with probability 1/10 and does not sink with probability 9/10. Then

$$\tilde{z} = \begin{cases} \$0, & \text{with probability } 10\%, \\ +\$8000, & \text{with probability } 90\% \end{cases} = \tilde{z}(0, 8000; \tfrac{1}{10}),$$

which represents the risk of the gamble.

Suppose your initial wealth is $w_0 = \$4000$. If the gamble is undertaken, then

$$\tilde{w} = \begin{cases} 4000, & \text{with probability } 10\%, \\ 12000, & \text{with probability } 90\%, \end{cases}$$

so

$$\mathbb{E}\left[\tilde{w}\right] = 4000 \cdot 10\% + 12000 \cdot 90\% = 11200.$$

To reduce risk, we may use two independent ships and split the goods between them. There are then four outcomes. Let x_1 denote "Ship 1 sinks and Ship 2 does not," x_2 denote "both ships sink," x_3 denote "both ships do not sink," and x_4 denote "Ship 1 does not sink and Ship 2 sinks." Then

$$\tilde{z} = \begin{cases} x_1 = 4000, & \text{with probability } 9\%, \\ x_2 = 0, & \text{with probability } 1\%, \\ x_3 = 8000, & \text{with probability } 81\%, \\ x_4 = 4000, & \text{with probability } 9\%. \end{cases}$$

This simplifies to

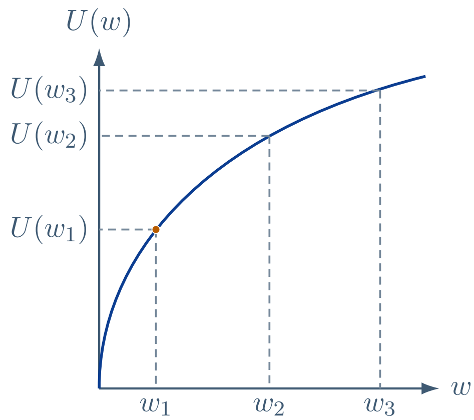
$$\tilde{z} = \begin{cases} 4000, & \text{with probability } 18\%, \\ 0, & \text{with probability } 1\%, \\ 8000, & \text{with probability } 81\%. \end{cases}$$

The question is which situation should be preferred. We still have $\mathbb{E}\left[\tilde{w}\right] = 11200$, so expected wealth alone cannot distinguish the two shipping arrangements even though splitting the cargo reduces risk.

Example 2.2 (St. Petersburg paradox) Suppose a coin is tossed repeatedly. The outcome lies in $\{H, T\}$, and you receive 2^n dollars if the first tail occurs on toss n. Then

$$\mathbb{E}\left[\tilde{z}\right] = \sum_{n=1}^{\infty} 2^n \frac{1}{2^n} = \infty,$$

so the expected value criterion would suggest paying an infinite amount to enter the game. The paradox is not that very large payoffs are impossible, but that a criterion based only on expected dollars treats every additional dollar as equally valuable. Bernoulli's explanation was to consider the expected utility of wealth, namely $\mathbb{E}\left[U(\tilde{w})\right]$, rather than $\mathbb{E}\left[\tilde{w}\right]$.



Marginal utility is high for the poor and lower for the rich. This leads to the **law of diminishing marginal utility of wealth**: as w increases, the marginal utility of wealth, $MU(w) = \partial U/\partial w$, decreases. Returning to the St. Petersburg paradox, Bernoulli's logarithmic utility gives the finite expected utility

$$\mathbb{E}[U(\tilde{z})] = \sum_{n=1}^{\infty} \frac{\log(2^n)}{2^n} = \log(2) \sum_{n=1}^{\infty} \frac{n}{2^n} = 2\log(2) < \infty.$$

This conclusion uses the growth rate of logarithmic utility; diminishing marginal utility by itself does not guarantee finite expected utility for every possible concave utility function.

Von Neumann and Morgenstern developed an expected utility function for risky outcomes, now called the **von Neumann–Morgenstern expected utility**. Let $x, y, z, \ldots$ denote monetary outcomes, let $\tilde{w}$ denote wealth, and let $\alpha, \beta, \ldots$ denote probabilities. If $\tilde{w}(x, y; \alpha)$ is a lottery, then

$$U(\tilde{w}) = U(x) \cdot \alpha + U(y) \cdot (1 - \alpha)$$

is the von Neumann–Morgenstern expected utility. A proof appears in the subsequent development.

Lecture 3 — Tuesday, May 13, 2014

Recall that we model risk as $\tilde{z} = \begin{cases} z_1 \\ z_2 \end{cases}$, so wealth becomes $\tilde{w} = w_0 + \tilde{z}$, where w_0 is initial wealth. Recall the law of diminishing marginal utility of wealth: as w increases, marginal utility

MU decreases, where

$$MU = \frac{\Delta U}{\Delta w} = \frac{\mathrm{d}U}{\mathrm{d}w}$$

is the slope of the utility function U. A utility function consistent with this law must be concave, so as w increases, the derivative $\mathrm{d}U/\mathrm{d}w$ decreases.

Example 3.1 A common utility function is $U(w) = \log(w)$. If $w_{\text{poor}} = \$10$ and $w_{\text{rich}} = \$1000$, then

$$U_{\text{poor}} = U(w_{\text{poor}}) = \log(10) \approx 2.3 \qquad \text{and} \qquad U_{\text{rich}} = U(w_{\text{rich}}) = \log(1000) \approx 6.9.$$

Moreover,

$$MU_{\text{poor}} = U'(w_{\text{poor}}) = \frac{1}{10} \qquad \text{and} \qquad MU_{\text{rich}} = U'(w_{\text{rich}}) = \frac{1}{1000}.$$

Also,

$$\frac{\mathrm{d}^2U}{\mathrm{d}w^2} = -\frac{1}{w^2} < 0,$$

so the function is concave for $w > 0$.

This example makes the economics precise: the same dollar amount changes utility more at low wealth than at high wealth, which is exactly the intuition behind concavity.

Preferences are defined over lotteries (risky wealth outcomes), not only over certain wealth levels. Let $\mathcal{L}$ be the set of finite lotteries, and write

$$\tilde{w}(w_1, \ldots, w_n; p_1, \ldots, p_n), \qquad p_i \geq 0, \qquad \text{and} \qquad \sum_{i=1}^{n} p_i = 1.$$

A preference relation $\succeq$ on $\mathcal{L}$ means "weakly preferred to." As usual, $\tilde{w} \succ \tilde{v}$ means strict preference, and $\tilde{w} \sim \tilde{v}$ means indifference. For $\alpha \in [0, 1]$, the mixture $\alpha\tilde{w} + (1 - \alpha)\tilde{v}$ means "play $\tilde{w}$ with probability α, otherwise play $\tilde{v}$." The von Neumann–Morgenstern theorem starts from the following behavioral axioms.

(1) **Completeness.** For any $\tilde{w}, \tilde{v} \in \mathcal{L}$, either $\tilde{w} \succeq \tilde{v}$ or $\tilde{v} \succeq \tilde{w}$ (or both). This axiom says that the decision maker can compare any two available lotteries. It is a natural requirement for choice, because without comparability an optimal decision may not be well-defined.

(2) **Transitivity.** If $\tilde{w} \succeq \tilde{v}$ and $\tilde{v} \succeq \tilde{u}$, then $\tilde{w} \succeq \tilde{u}$. This axiom ensures that the ranking is internally consistent and does not produce cycles. It is behaviorally reasonable, because

cyclical preferences lead to contradictory decisions.

(3) **Continuity.** If $\tilde{w} \succ \tilde{v} \succ \tilde{u}$, then there exists $\alpha \in (0,1)$ such that

$$\tilde{v} \sim \alpha\tilde{w} + (1-\alpha)\tilde{u}.$$

This axiom states that an intermediate lottery can be represented as a mixture of a better and a worse lottery. It is reasonable to impose continuity because small probability changes should not create abrupt jumps in preference.

(4) **Independence.** For any $\tilde{w}, \tilde{v}, \tilde{u} \in \mathcal{L}$ and $\alpha \in (0,1)$,

$$\tilde{w} \succeq \tilde{v} \qquad \text{if and only if} \qquad \alpha\tilde{w} + (1-\alpha)\tilde{u} \succeq \alpha\tilde{v} + (1-\alpha)\tilde{u}.$$

This axiom says that adding the same common consequence to both lotteries does not reverse their ranking. It is reasonable because only the differing components of two prospects should matter for comparison.

(5) **Ranking (Monotonicity on Canonical Lotteries).** Let b be a best outcome and a a worst outcome. If $\alpha \geq \beta$, then

$$\tilde{w}(b, a; \alpha) \succeq \tilde{w}(b, a; \beta).$$

This axiom says that a lottery with a higher probability of the best outcome should be weakly preferred. It is reasonable because it fixes the direction of preference on benchmark two-point lotteries.

(6) **Measurability.** For each sure outcome x between a and b, there exists $\alpha_x \in [0,1]$ such that

$$x \sim \tilde{w}(b, a; \alpha_x).$$

This axiom says that each sure outcome can be matched to an equivalent canonical lottery between the best and worst outcomes. It is reasonable because it gives a numerical scale for utility through equivalent probabilities.

(7) **Bounded Set.** There exist outcomes a, b such that for every outcome x, $b \succeq x \succeq a$. This axiom says that the domain contains a worst and a best outcome. It is reasonable because these endpoints are needed to construct canonical lotteries and normalize utility.

Theorem 3.2 (von Neumann–Morgenstern Expected Utility Theorem) Suppose a preference relation $\succeq$ over finite lotteries satisfies the seven axioms above. Then there exists a continuous utility index U on sure outcomes such that for any lotteries

$$\tilde{w} = (w_1, \ldots, w_n; p_1, \ldots, p_n) \qquad \text{and} \qquad \tilde{v} = (v_1, \ldots, v_m; q_1, \ldots, q_m),$$

$$\tilde{w} \succeq \tilde{v} \qquad \text{if and only if} \qquad \sum_{i=1}^{n} p_i U(w_i) \geq \sum_{j=1}^{m} q_j U(v_j).$$

In particular, for a two-outcome lottery $\tilde{w}(x, y; \alpha)$,

$$U(\tilde{w}(x, y; \alpha)) = \alpha U(x) + (1 - \alpha) U(y).$$

Moreover, U is unique up to positive affine transformations: if V represents the same preference ordering, then $V = a + bU$ for some $a \in \mathbb{R}$, $b > 0$.

Proof sketch. Let a and b be least and most preferred outcomes (boundedness). By measurability and continuity, for each sure outcome x there exists $\alpha_x \in [0, 1]$ such that

$$x \sim \tilde{w}(b, a; \alpha_x).$$

Define $U(x) := \alpha_x$, so $U(a) = 0$, $U(b) = 1$, and ranking implies that $x \succeq y$ if and only if $U(x) \geq U(y)$. Independence extends this representation from sure outcomes to mixtures and, after reducing compound lotteries, forces linearity in probabilities. Hence each lottery is evaluated by expected utility $\sum_i p_i U(w_i)$. Affine uniqueness follows because order and mixture comparisons are preserved exactly by maps of the form $a + bU$ with $b > 0$. □

Risk Aversion and Risk Premia

Any financial contract involves both time and risk. On one side is the borrower, namely the business corporation issuing securities, and on the other side are the lenders, namely households or savers who purchase those securities. The lenders bear both time and risk.

The main reason savers participate in financial contracts is to smooth consumption over time — that is, to maintain a more consistent standard of living. Consider a consumer who lives for two periods, the current period (Period 0) and the future period (Period 1). Recall that

$$y_0 = c_0 + s_0 \qquad \text{and} \qquad y_1 = c_1 + s_1.$$

Such a consumer may save in Period 0 in order to have more resources in Period 1. Suppose utility is time-additive, with the same undiscounted one-period utility function U in both periods. If U

is strictly concave, the consumer prefers $c_0 = c_1 = 9$ units to $c_0 = 8$ and $c_1 = 10$. Thus

$$U(c_0 = c_1 = 9) > U(c_0 = 8, c_1 = 10),$$

so

$$2U(9) > U(8) + U(10),$$

which implies

$$U(9) > \frac{1}{2}(U(8) + U(10)).$$

This is consistent with a concave utility function.

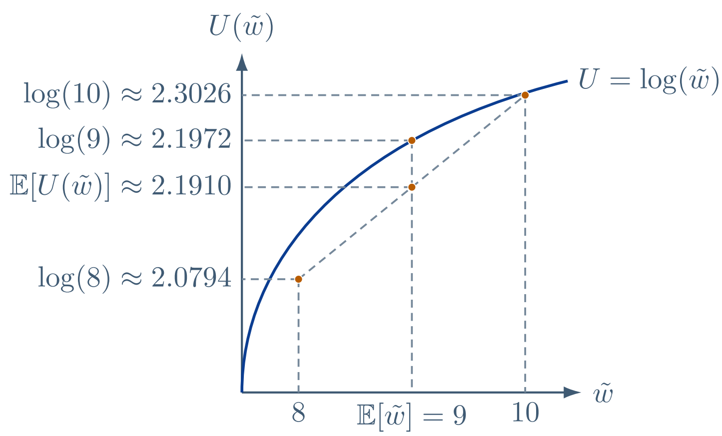


FIGURE 7

The concavity calculation is illustrated in Figure 7. Consumption smoothing implies a concave utility function. To see the connection with risk, consider an investor with wealth \$9 who faces the gamble

$$\tilde{z} = \begin{cases} +\$1 \text{ with probability } 1/2, \\ -\$1 \text{ with probability } 1/2, \end{cases}$$

then

$$\tilde{w} = \begin{cases} \$10 \text{ with probability } 1/2, \\ \$8 \text{ with probability } 1/2. \end{cases}$$

Of course, $\mathbb{E}[\tilde{w}] = 10(1/2) + 8(1/2) = 9$. The outcomes \$8 and \$10 are uncertain, whereas \$9 is certain. Using a strictly concave von Neumann–Morgenstern utility function, a risk-averse investor strictly prefers certainty, so

$$U(9) > \frac{1}{2}[U(8) + U(10)].$$

Equivalently,

$$U(\mathbb{E}[\tilde{w}]) > \mathbb{E}[U(\tilde{w})].$$

For a merely concave utility function Jensen's inequality is weak; strictness requires a nondegenerate gamble and strict concavity over the relevant interval.

As another example, let $U(w) = \log(w)$ and $w_0 = \$10$. If the contract yields

$$\tilde{z} = \begin{cases} 1 \text{ with probability } 1/2, \\ -1 \text{ with probability } 1/2, \end{cases}$$

then

$$\tilde{w} = w_0 + \tilde{z} = \begin{cases} \$11 \text{ with probability } 1/2 \\ \$9 \text{ with probability } 1/2 \end{cases}.$$

We verify that this investor is risk averse because

$$U(\mathbb{E}[\tilde{w}]) = U(10) = \log(10) > \frac{1}{2}[\log(9) + \log(11)] = \mathbb{E}[U(\tilde{w})].$$

Lecture 4 — Thursday, May 15, 2014

For an investor whose objective is to maximize $\mathbb{E}[U(\tilde{w})]$, **strict risk aversion** means that utility from certainty exceeds expected utility from every nondegenerate uncertain outcome. That is,

$$U(\mathbb{E}[\tilde{w}]) > \mathbb{E}[U(\tilde{w})].$$

This is equivalent to strict concavity over the relevant wealth range. Weak risk aversion corresponds to concavity and gives a weak inequality. Similarly, a **risk-loving** investor satisfies $U(\mathbb{E}[\tilde{w}]) < \mathbb{E}[U(\tilde{w})]$ for nondegenerate risks when utility is strictly convex. If an investor is **risk-neutral**, then utility is affine and $U(\mathbb{E}[\tilde{w}]) = \mathbb{E}[U(\tilde{w})]$. The three utility shapes are compared in Figure 8.

The **risk premium** $\pi(w_0, \tilde{z})$ is the maximum amount an investor is willing to pay to avoid risk. Recall the previous example: if $w_0 = \$10$ and

$$\tilde{z} = \begin{cases} \$1 \text{ with probability } 1/2, \\ -\$1 \text{ with probability } 1/2, \end{cases}$$

then

$$\tilde{w} = \begin{cases} \$11 \text{ with probability } 1/2, \\ \$9 \text{ with probability } 1/2, \end{cases}$$

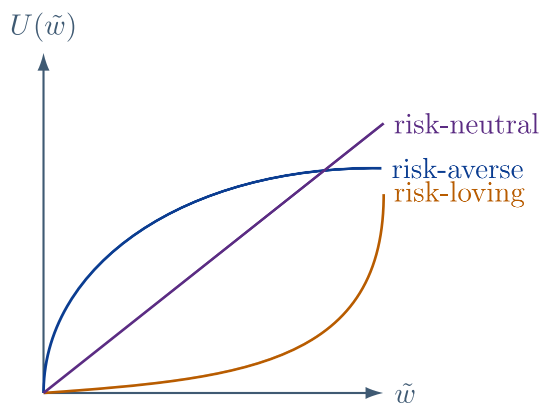


FIGURE 8

with utility function $U(\tilde{w}) = \log(w)$. Then $\mathbb{E}[\tilde{w}] = \$10$, which is the certainty outcome. Also, $U(\mathbb{E}[\tilde{w}]) = \log(10) = 2.3026$ utility units and $\mathbb{E}[U(\tilde{w})] = 2.2976$ utility units. We seek $\pi(w_0, \tilde{z})$.

First define w_{CE} to be the certainty-equivalent level of wealth. Then $2.2976 = \log(w_{CE})$, which implies $w_{CE} = \$9.95$. Therefore,

$$\pi(w_0, \tilde{z}) = \mathbb{E}[\tilde{w}] - w_{CE} = 10 - 9.95 = 0.05.$$

In general,

$$\pi = \mathbb{E}[\tilde{z} + w_0] - w_{CE}.$$

Actuarially fair gambles have zero expected value. For example, if

$$\tilde{z} = \begin{cases} 1 \text{ with probability } 1/2 \\ -1 \text{ with probability } 1/2 \end{cases},$$

then $\mathbb{E}[\tilde{z}] = 0$. If $\mathbb{E}[\tilde{z}] \neq 0$, then the gamble is not actuarially fair.

The risk premium π depends on both w_0 and $\tilde{z}$. There are two ways to compute π: from the graph by using $\pi = \mathbb{E}[\tilde{z} + w_0] - w_{CE}$, or from the risk-aversion condition

$$U(\mathbb{E}[\tilde{w}] - \pi) = \mathbb{E}[U(\tilde{w})].$$

In our running example,

$$U(10 - \pi) = 2.2976 \implies \log(10 - \pi) = 2.2976 \implies \pi = 10 - e^{2.2976} = 0.05.$$

The **certainty equivalent** (or **cash equivalent**) of the gamble's payoff, $CE(w_0, \tilde{z})$, is the sure payoff that has the same effect on welfare as bearing the risk. It equals the gamble's expected payoff less the risk premium. Thus w_{CE} is certainty-equivalent wealth, while $CE = w_{CE} - w_0$ is the certainty-equivalent payoff.

Recall that $\mathbb{E}[U(\tilde{w})] = U(w_{CE})$. Then

$$\mathbb{E}[U(\tilde{w})] = U(w_0 + \mathbb{E}[\tilde{z}] - \pi).$$

The term $\mathbb{E}[\tilde{z}] - \pi$ is therefore the certainty-equivalent payoff. If the gamble is actuarially fair, then $CE = -\pi$. This provides another way to compute π. Thus the certainty equivalent is measured on the same dollar scale as wealth, while the risk premium is the discount applied to expected wealth because the payoff is risky. For a risk-averse investor, a larger risk premium means that a larger sure payment is required to compensate for bearing the same gamble.

Example 4.1 Suppose Mr. U and Mr. V are both risk-averse. Mr. U's utility function is $\sqrt{w}$ and Mr. V's is $\log(w)$. Both have $w_0 = 4000$ and both face the actuarially fair gamble

$$\tilde{z} = \begin{cases} 2000 \text{ with probability } 1/2 \\ -2000 \text{ with probability } 1/2 \end{cases},$$

so

$$\tilde{w} = w_0 + \tilde{z} = \begin{cases} 6000 \text{ with probability } 1/2 \\ 2000 \text{ with probability } 1/2 \end{cases}.$$

We seek π_u and π_v.

For Mr. U:

$$\mathbb{E}[U(\tilde{w})] = U(w_0 + \mathbb{E}[\tilde{z}] - \pi_u) \implies \frac{1}{2}\sqrt{2000} + \frac{1}{2}\sqrt{6000} = \sqrt{4000 - \pi_u} \implies \pi_u = 267.9.$$

For Mr. V:

$$\frac{1}{2}\log(2000) + \frac{1}{2}\log(6000) = \log(4000 - \pi_v) \implies \pi_v = 4000 - \sqrt{12{,}000{,}000} \approx 535.9.$$

Thus there are genuine degrees of risk aversion: both investors dislike risk, but one dislikes it more strongly.

The Arrow–Pratt Measure of Risk Aversion

To compare degrees of risk aversion, we now turn to the **Arrow–Pratt measure of absolute risk aversion**. The next lecture derives its local approximation to the risk premium.

Lecture 5 — Tuesday, May 20, 2014

Assume an actuarially fair gamble, so $\mathbb{E}[\tilde{z}] = 0$, $\mathrm{Var}(\tilde{z}) = \mathbb{E}\left[\tilde{z}^2\right] = \sigma^2$, and initial wealth is w. Then $\pi(w, \tilde{z})$ must satisfy $\mathbb{E}[U(w + \tilde{z})] = U(w + \mathbb{E}[\tilde{z}] - \pi)$. Taking a second-order Taylor approximation about w gives

$$\mathbb{E}\left[U(w) + \tilde{z} \cdot U'(w) + \frac{1}{2}\tilde{z}^2 \cdot U''(w) + \text{Remainder}\right] \approx U(w) - \pi \cdot U'(w) + \text{Remainder}.$$

This gives the approximation below.

$$U(w) + \frac{1}{2}\sigma^2 \cdot U''(w) \approx U(w) - \pi \cdot U'(w) \implies \pi \approx \frac{1}{2}\sigma^2 \left(-\frac{U''(w)}{U'(w)}\right).$$

The term

$$-\frac{U''(w)}{U'(w)}$$

is the **Arrow–Pratt measure of absolute risk aversion**, which we denote by $ARA(w)$. The approximation

$$\pi \approx \frac{1}{2}\sigma^2 \left(-\frac{U''(w)}{U'(w)}\right)$$

is the **Arrow–Pratt approximation** of the risk premium.

The sign of π depends on the sign of ARA. Risk aversion occurs when $U' > 0$ and $U'' < 0$, which gives positive ARA. Risk neutrality has $U'' = 0$, which gives $\pi = 0$. Risk seeking occurs when $U' > 0$ and $U'' > 0$, which gives negative ARA and therefore $\pi < 0$.

The risk premium π is proportional to σ^2, which means that the variance of risk is central to its determination. A small $\tilde{z}$ implies small π. The Arrow–Pratt measure is based on $\mathbb{E}[\tilde{z}] = 0$ and on $\tilde{z}$ being small. If either condition fails, we can instead use the **Markowitz measure of risk**

aversion,

$$\mathbb{E}\left[U(\tilde{w})\right] = U(\mathbb{E}\left[\tilde{w}\right] - \pi),$$

where π is computed directly from the indifference condition, giving the **Markowitz definition** of the risk premium. The Arrow–Pratt and Markowitz measures agree for small risks because both approximate the same calculation of the certainty equivalent.

Example 5.1 Let $U = \log(w)$, $w_0 = \$10$, and $\tilde{z} = (\pm 1; 1/2)$, so $\tilde{w} = w_0 + \tilde{z} = (11, 9; 1/2)$. The Arrow–Pratt approximation gives

$$\pi \approx \frac{1}{2}\sigma^2 \cdot \mathrm{ARA}(w_0) = \frac{1}{2}\mathbb{E}\left[\tilde{z}^2\right] \cdot \mathrm{ARA}(w_0) = \frac{1}{2}\left(\frac{1}{2}(1)^2 + \frac{1}{2}(-1)^2\right)\left(\frac{1}{w_0}\right) = \frac{1}{20}.$$

The Markowitz definition gives

$$\mathbb{E}\left[U(w_0 + \tilde{z})\right] = U(\mathbb{E}\left[\tilde{w}\right] - \pi) \implies \frac{1}{2}\log(9) + \frac{1}{2}\log(11) = \log(10 - \pi),$$

which yields $\pi = 10 - \sqrt{99} \approx 0.05013$, very close to the Arrow–Pratt approximation $1/20 = 0.05$.

Any positive affine transformation of U has no effect on π. Also,

$$\mathrm{ARA}(w) = -\frac{U''(w)}{U'(w)} = -\frac{\mathrm{d}}{\mathrm{d}w}\log U'(w),$$

so ARA measures the proportional rate at which marginal utility declines as wealth increases. Its numerical value depends on the unit in which wealth is measured.

If two investors have the same wealth w but different utility functions, then investor A is locally more risk-averse at w than investor B when $\mathrm{ARA}_A(w) \geq \mathrm{ARA}_B(w)$. Equivalently, investor A demands at least as large a risk premium to second order for every sufficiently small risk at that wealth. A global comparison over all wealth levels requires this inequality for every w in the common domain. A fixed utility function exhibits **decreasing absolute risk aversion** (**DARA**) when $\mathrm{ARA}(w)$ decreases with wealth.

We next consider **relative risk aversion** (**RRA**), which is a unitless measure. It is the negative wealth elasticity of marginal utility:

$$\mathrm{RRA} = -\frac{\%\Delta MU}{\%\Delta w},$$

and therefore

$$\mathrm{RRA}(w) = w \cdot \mathrm{ARA}(w).$$

To express the premium and the risk relative to wealth, define $\pi_R \equiv \pi/w$ and $\tilde{z}_R \equiv \tilde{z}/w$. These are the **relative premium** and **relative risk**, respectively. From the Markowitz definition,

$$\mathbb{E}\left[U(w + w \cdot \tilde{z}_R)\right] = U(w + \mathbb{E}\left[\tilde{z}\right] - w \cdot \pi_R),$$

so, in general,

$$\mathbb{E}\left[U(w(1 + \tilde{z}_R))\right] = U\left(w\left(1 + \mathbb{E}[\tilde{z}_R] - \pi_R\right)\right).$$

For an actuarially fair risk, $\mathbb{E}[\tilde{z}_R] = 0$, and the right-hand side reduces to $U(w(1 - \pi_R))$. This exact indifference relation can be used to compute π_R even when $\tilde{z}$ is large. For a small actuarially fair risk, we may instead use the Arrow–Pratt formulation. Since $\pi \approx (1/2)\sigma^2 \cdot \mathrm{ARA}(w)$ and $\sigma_R^2 = \mathrm{Var}(\tilde{z}_R)$, we have

$$\sigma^2 = \mathrm{Var}(\tilde{z}) = \mathrm{Var}(w \cdot \tilde{z}_R) = w^2\sigma_R^2.$$

Therefore

$$\pi_R = \frac{\pi}{w} \approx \frac{1}{2}\sigma_R^2 \cdot \mathrm{RRA}.$$

Under DARA, absolute risk aversion decreases as wealth increases. Relative risk aversion, however, may increase, decrease, or remain constant because $\mathrm{RRA}(w) = w \cdot \mathrm{ARA}(w)$ combines the level of wealth with absolute risk aversion.

Stochastic Dominance

The idea of **stochastic dominance** is to compare risky alternatives. Consider two projects, A and B. Let X_A and X_B be random variables representing payoffs, and let x_A and x_B denote particular payoffs.

x_A	$\mathbb{P}(X_A = x_A)$	x_B	$\mathbb{P}(X_B = x_B)$
10	0.2	10	0.2
20	0.6	50	0.4
100	0.2	60	0.4

Then

$$\mu_A = 34, \qquad \sigma_A \approx 33.23, \qquad \mu_B = 46, \qquad \text{and} \qquad \sigma_B \approx 18.55.$$

Both the mean and the standard deviation favour B, but two moments alone do not determine the preference of every risk-averse investor. Stochastic dominance provides a comparison of the entire distributions.

Lecture 6 — Thursday, May 22, 2014

Stochastic Dominance Criteria

Consider an investor who evaluates risky wealth $\tilde{w}$ through $\mathbb{E}[U(\tilde{w})]$. Assume that more is preferred to less; that is, $U' > 0$. Therefore, the investor places more weight on higher payoffs.

Definition 6.1 Let F_A and F_B be two cdfs for random payoffs X_A and X_B, respectively. We say that B **first-order stochastically dominates** A when $F_A(x) \geq F_B(x)$ for every payoff level x. The dominance is strict when the inequality is strict for at least one x. Thus F_B lies everywhere weakly to the right of and below F_A.

The cdf comparison has a direct economic interpretation. At every payoff threshold x, project B assigns weakly less probability to receiving a payoff no larger than x. Thus B gives at least as much chance of exceeding every threshold and, under strict dominance, strictly more chance at some threshold. Every investor who simply prefers more wealth to less therefore weakly prefers B.

Theorem 6.2 B first-order stochastically dominates A if and only if

$$\mathbb{E}[U(X_B)] \geq \mathbb{E}[U(X_A)]$$

for all nondecreasing utility functions U for which the expectations exist.

Example 6.3 The distributions of X_A and X_B are as follows. The pmf of A is

$$\mathbb{P}(X_A = x_A) = \begin{cases} 0.2 & \text{if } x_A = 10 \\ 0.6 & \text{if } x_A = 20 \\ 0.2 & \text{if } x_A = 100 \end{cases},$$

so the cdf of A is

$$F_A(x_A) = \begin{cases} 0 & x_A < 10 \\ 0.2 & 10 \le x_A < 20 \\ 0.8 & 20 \le x_A < 100 \\ 1 & x_A \ge 100 \end{cases}.$$

The pmf of B is

$$\mathbb{P}(X_B = x_B) = \begin{cases} 0.2 & \text{if } x_B = 10 \\ 0.4 & \text{if } x_B = 50 \\ 0.4 & \text{if } x_B = 60 \end{cases},$$

so the cdf of B is

$$F_B(x_B) = \begin{cases} 0 & x_B < 10 \\ 0.2 & 10 \le x_B < 50 \\ 0.6 & 50 \le x_B < 60 \\ 1 & x_B \ge 60 \end{cases}.$$

Neither cdf is everywhere below the other, so there is no first-order dominance. We return to these distributions after defining second-order dominance.

Another notion takes into account the attitude towards risk.

Definition 6.4 We say that B **second-order stochastically dominates** A when

$$\int_{-\infty}^{x} F_A(t)\,\mathrm{d}t \ge \int_{-\infty}^{x} F_B(t)\,\mathrm{d}t$$

for all x.

Theorem 6.5 B second-order stochastically dominates A if and only if

$$\mathbb{E}[U(X_B)] \ge \mathbb{E}[U(X_A)]$$

for all nondecreasing and concave utility functions U for which the expectations exist.

Second-order stochastic dominance is therefore the appropriate comparison once risk aversion is imposed. The integrated cdf condition measures cumulative downside probability rather than just pointwise dominance. It allows B to fail first-order dominance, provided that the remaining distributional difference is still preferred by every investor with increasing and concave utility.

For the preceding distributions, direct integration gives

$$\int_{-\infty}^{x} (F_A(t) - F_B(t))\, \mathrm{d}t = \begin{cases} 0, & x < 20, \\ 0.6(x-20), & 20 \le x < 50, \\ 18 + 0.2(x-50), & 50 \le x < 60, \\ 20 - 0.2(x-60), & 60 \le x < 100, \\ 12, & x \ge 100. \end{cases}$$

This expression is nonnegative for every x, so B second-order stochastically dominates A. Thus every investor with increasing concave utility weakly prefers B, despite the absence of first-order dominance.

Example 6.6 Let

$$X_A = \begin{cases} 2, & \text{with probability } 1/3, \\ 4, & \text{with probability } 2/3, \end{cases} \quad \text{and} \quad X_B = \begin{cases} 3, & \text{with probability } 1/4, \\ 5, & \text{with probability } 3/4. \end{cases}$$

Their cdfs are

$$F_A(x) = \begin{cases} 0, & x < 2, \\ \frac{1}{3}, & 2 \le x < 4, \\ 1, & x \ge 4, \end{cases} \quad \text{and} \quad F_B(x) = \begin{cases} 0, & x < 3, \\ \frac{1}{4}, & 3 \le x < 5, \\ 1, & x \ge 5. \end{cases}$$

Define

$$D(x) \equiv \int_{-\infty}^{x} (F_A(t) - F_B(t))\, \mathrm{d}t.$$

A direct piecewise calculation gives

$$D(x) = \begin{cases} 0, & x < 2, \\ (x-2)/3, & 2 \le x < 3, \\ 1/3 + (x-3)/12, & 3 \le x < 4, \\ 5/12 + (3/4)(x-4), & 4 \le x < 5, \\ 7/6, & x \ge 5. \end{cases}$$

Therefore $D(x) \ge 0$ for all x, with strict inequality for some x. Hence

$$\int_{-\infty}^{x} F_A(t)\, \mathrm{d}t \ge \int_{-\infty}^{x} F_B(t)\, \mathrm{d}t \quad \text{for all } x,$$

so B second-order stochastically dominates A. In fact, because $F_A(x) \geq F_B(x)$ for all x with strict inequality on some interval, B also first-order stochastically dominates A.

3. Portfolio Theory and Equilibrium Asset Pricing

Mean–Variance Portfolio Choice

Let $\tilde{r}_i = (P_{i1} - P_{i0})/P_{i0}$ denote the rate of return on asset i, for $i = 1, \ldots, n$. Let r_f denote the risk-free rate of return, that is, the constant return on an asset with no risk. The **mean**, or **expected rate of return**, is $\mathbb{E}[\tilde{r}_i] = \mu_i$. The **variance** is

$$\mathrm{Var}(\tilde{r}_i) = \mathbb{E}\left[(\tilde{r}_i - \mathbb{E}[\tilde{r}_i])^2\right] = \sigma_i^2 = \sigma_{ii}.$$

For two risky assets, the **covariance** is

$$\mathrm{Cov}(\tilde{r}_i, \tilde{r}_j) = \mathbb{E}\left[(\tilde{r}_i - \mathbb{E}[\tilde{r}_i])(\tilde{r}_j - \mathbb{E}[\tilde{r}_j])\right] = \sigma_{ij}.$$

The vector of asset returns is $\tilde{\boldsymbol{r}} = \begin{bmatrix} \tilde{r}_1 \\ \vdots \\ \tilde{r}_n \end{bmatrix}$, and the $n \times 1$ vector of asset weights is $\boldsymbol{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}$, where $\boldsymbol{w}$ no longer denotes wealth.

For a portfolio of n assets, we write $\tilde{r}_p = w_1\tilde{r}_1 + \cdots + w_n\tilde{r}_n$, or more compactly $\tilde{r}_p = \boldsymbol{w}^\top \tilde{\boldsymbol{r}}$. Then

$$\mu_p = \mathbb{E}[\tilde{r}_p] = \sum_{i=1}^{n} w_i \mathbb{E}[\tilde{r}_i] = \boldsymbol{w}^\top \mathbb{E}[\tilde{\boldsymbol{r}}].$$

Example 6.7 For a portfolio of two assets, let $\boldsymbol{w} = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$, $\tilde{\boldsymbol{r}} = \begin{bmatrix} \tilde{r}_1 \\ \tilde{r}_2 \end{bmatrix}$, and $\boldsymbol{\mu} = \begin{bmatrix} \mathbb{E}[\tilde{r}_1] \\ \mathbb{E}[\tilde{r}_2] \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$. Then $\tilde{r}_p = \boldsymbol{w}^\top \tilde{\boldsymbol{r}}$ and $\mathbb{E}[\tilde{r}_p] = \boldsymbol{w}^\top \boldsymbol{\mu}$.

This compact matrix notation is used throughout portfolio theory because it lets us write return and variance formulas in a way that extends immediately from two assets to n assets.

The variance–covariance matrix $\boldsymbol{\Omega}$ of portfolio returns is symmetric and positive semidefinite, with variances on the diagonal and covariances off the diagonal:

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \sigma_{nn} \end{bmatrix}$$

In our portfolio example with two assets,

$$\begin{aligned} \operatorname{Var}(\tilde{r}_p) &= \operatorname{Var}(w_1\tilde{r}_1) + \operatorname{Var}(w_2\tilde{r}_2) + 2\operatorname{Cov}(w_1\tilde{r}_1, w_2\tilde{r}_2) \\ &= w_1^2 \operatorname{Var}(\tilde{r}_1) + w_2^2 \operatorname{Var}(\tilde{r}_2) + 2w_1w_2 \operatorname{Cov}(\tilde{r}_1, \tilde{r}_2) \\ &= w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_{12} \\ &= \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \\ &= \boldsymbol{w}^\top \boldsymbol{\Omega} \boldsymbol{w}. \end{aligned}$$

There is an immediate connection to the previous consumer theory. Consider a representative consumer who lives for two periods, Period 0 (today) and Period 1 (the future). Let y_0 denote today's wealth and let $\tilde{y}_1$ denote future wealth. Also let P_{i0} be the price of stock i today and P_{i1} its future price, so that

$$\tilde{r}_i = \frac{P_{i1} - P_{i0}}{P_{i0}}.$$

At time $t = 0$, wealth is allocated according to

$$y_0 = a_1P_{10} + a_2P_{20} + \dots + a_nP_{n0}.$$

The weight of asset i in the portfolio is therefore

$$w_i = \frac{a_iP_{i0}}{y_0},$$

and hence

$$\sum_{i=1}^{n} w_i = 1.$$

If $\tilde{r}_p = \boldsymbol{w}^\top \tilde{\boldsymbol{r}}$, then future wealth is

$$\tilde{y}_1 = a_1P_{11} + \dots + a_nP_{n1}.$$

We may rewrite this as

$$\tilde{y}_1 = \sum_{i=1}^{n} a_i(P_{i1} - P_{i0}) + \sum_{i=1}^{n} a_i P_{i0} = \sum_{i=1}^{n} a_i(P_{i1} - P_{i0}) + y_0$$

and therefore

$$\begin{aligned}\tilde{y}_1 &= \sum_{i=1}^{n} a_i P_{i0} \cdot \frac{P_{i1} - P_{i0}}{P_{i0}} + y_0 \\ &= y_0 \left(1 + \sum_{i=1}^{n} w_i \tilde{r}_i\right) \\ &= y_0(1 + \boldsymbol{w}^\top \tilde{\boldsymbol{r}}).\end{aligned}$$

Thus future wealth depends on both portfolio weights and asset returns.

Lecture 7 — Tuesday, May 27, 2014

For initial wealth y_0, portfolio weights $\boldsymbol{w}$, and random asset returns $\tilde{\boldsymbol{r}}$, future wealth is $\tilde{y}_1 = y_0(1 + \boldsymbol{w}^\top \tilde{\boldsymbol{r}})$. Taking expectations gives

$$\mathbb{E}\left[\tilde{y}_1\right] = y_0(1 + \boldsymbol{w}^\top \mathbb{E}\left[\tilde{\boldsymbol{r}}\right]) = y_0(1 + \mu_p).$$

Since y_0 is nonrandom at the portfolio-choice date, taking variances gives directly

$$\begin{aligned}\operatorname{Var}(\tilde{y}_1) &= y_0^2 \operatorname{Var}(\boldsymbol{w}^\top \tilde{\boldsymbol{r}}) \\ &= y_0^2 \boldsymbol{w}^\top \cdot \text{Var–Cov}(\tilde{\boldsymbol{r}}) \cdot \boldsymbol{w} \\ &= y_0^2 \boldsymbol{w}^\top \boldsymbol{\Omega} \boldsymbol{w} \\ &= y_0^2 \sigma_p^2.\end{aligned}$$

Mean–variance analysis connects to choice under uncertainty under the conditions described below. Given a target return μ_p, investors choose $\boldsymbol{w}_{n\times 1}$ to minimize σ_p^2. This criterion is exact with quadratic utility, and it is also exact for expected-utility comparisons of jointly normally distributed returns when investors have increasing concave utility and the relevant expectations exist.

To justify this equivalence, assume a risk-averse investor, so that $U' > 0$ and $U'' < 0$. A second-order Taylor approximation about $\mathbb{E}[\tilde{y}_1]$ gives

$$U(\tilde{y}_1) = U(\mathbb{E}[\tilde{y}_1]) + U'(\mathbb{E}[\tilde{y}_1])(\tilde{y}_1 - \mathbb{E}[\tilde{y}_1]) + \frac{1}{2}U''(\mathbb{E}[\tilde{y}_1])(\tilde{y}_1 - \mathbb{E}[\tilde{y}_1])^2 + \text{Remainder}.$$

Taking expectations and retaining terms through second order yields

$$\mathbb{E}[U(\tilde{y}_1)] \approx U(\mathbb{E}[\tilde{y}_1]) + U'(\mathbb{E}[\tilde{y}_1]) \cdot \cancel{\mathbb{E}[(\tilde{y}_1 - \mathbb{E}[\tilde{y}_1])]} + \frac{1}{2}U''(\mathbb{E}[\tilde{y}_1]) \cdot \mathbb{E}\left[(\tilde{y}_1 - \mathbb{E}[\tilde{y}_1])^2\right].$$

Therefore,

$$\mathbb{E}[U(\tilde{y}_1)] \approx U(\mathbb{E}[\tilde{y}_1]) + \frac{1}{2}U''(\mathbb{E}[\tilde{y}_1]) \cdot \mathrm{Var}(\tilde{y}_1).$$

The approximation becomes exact when U is quadratic. Alternatively, suppose the vector of asset returns is jointly multivariate normal, so every portfolio return is normal. It is not enough that the individual returns merely have normal marginal distributions. Under joint normality,

$$Z := \frac{\tilde{r}_p - \mu_p}{\sigma_p} \sim \mathcal{N}(0,1),$$

so $\tilde{r}_p = \mu_p + Z \cdot \sigma_p$. At a fixed mean, a smaller standard deviation is preferred by every increasing concave expected-utility maximizer for whom expected utility exists.

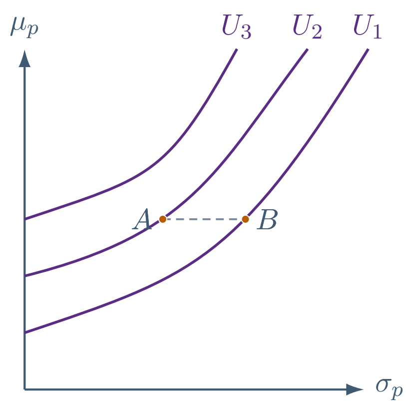


Figure 9

In Figure 9, A is strictly preferred to B because it lies northwest, toward higher μ_p and lower σ_p.

Constructing the Minimum Variance Frontier

Consider n risky assets and a risk-averse investor who cares only about one-period μ_p and σ_p^2. Assume $\boldsymbol{\Omega}$ is positive definite and that $\boldsymbol{\mu}$ is not proportional to $\mathbf{1}$; then $\boldsymbol{\Omega}^{-1}$ exists and $d = ac - b^2 > 0$. We have

$$\mu_p = \mathbb{E}\left[\tilde{r}_p\right] = \mathbb{E}\left[w_1 r_1 + \cdots + w_n r_n\right] = w_1\mu_1 + \cdots + w_n\mu_n = \boldsymbol{w}^\top \boldsymbol{\mu}$$

and $\sigma_p^2 = \boldsymbol{w}^\top \boldsymbol{\Omega} \boldsymbol{w}$ for $\boldsymbol{\Omega} = \begin{bmatrix} \sigma_{11} & \dots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \dots & \sigma_{nn} \end{bmatrix}$, the variance–covariance matrix.

The investor's objective is to choose optimal weights $\boldsymbol{w}$ so that σ_p is minimized subject to a given μ_p. That is,

$$\min_{\boldsymbol{w}} \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{\Omega} \boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}^\top \boldsymbol{\mu} = \mu_p \qquad \text{and} \qquad \boldsymbol{w}^\top \mathbf{1} = 1.$$

To solve this optimization problem, we use Lagrange multipliers. Let

$$\mathcal{L}(\boldsymbol{w}, \lambda_1, \lambda_2) = \frac{1}{2}\boldsymbol{w}^\top \boldsymbol{\Omega} \boldsymbol{w} + \lambda_1[\mu_p - \boldsymbol{w}^\top \boldsymbol{\mu}] + \lambda_2[1 - \boldsymbol{w}^\top \mathbf{1}].$$

The multiplier λ_1 enforces the constraint on the target return, and λ_2 enforces the full investment constraint. To solve the problem, we set

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}}, \qquad \frac{\partial \mathcal{L}}{\partial \lambda_1}, \qquad \text{and} \qquad \frac{\partial \mathcal{L}}{\partial \lambda_2}$$

equal to zero:

$$\begin{aligned}
&\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \mathbf{0} \implies \frac{1}{2} \cdot 2\boldsymbol{\Omega}\boldsymbol{w} - \lambda_1 \boldsymbol{\mu} - \lambda_2 \mathbf{1} = \mathbf{0} \\
&\frac{\partial \mathcal{L}}{\partial \lambda_1} = 0 \implies \mu_p - \boldsymbol{w}^\top \boldsymbol{\mu} = 0 \implies \boldsymbol{w}^\top \boldsymbol{\mu} = \mu_p \implies \boldsymbol{\mu}^\top \boldsymbol{w} = \mu_p \\
&\frac{\partial \mathcal{L}}{\partial \lambda_2} = 0 \implies 1 - \boldsymbol{w}^\top \mathbf{1} = 0 \implies \mathbf{1}^\top \boldsymbol{w} = 1
\end{aligned}$$

Solving for $\boldsymbol{w}$ gives

$$\boldsymbol{w} = \lambda_1 \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} + \lambda_2 \boldsymbol{\Omega}^{-1} \mathbf{1}.$$

λ_1 and λ_2 are unknown, so we now solve for them explicitly.

Lecture 8 — Thursday, May 29, 2014

To continue, multiply the expression for $\boldsymbol{w}$ by $\boldsymbol{\mu}^\top$. This gives

$$\boldsymbol{\mu}^\top \boldsymbol{w} = \lambda_1 \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\mu} + \lambda_2 \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}.$$

Next multiply the same expression by $\mathbf{1}^\top$. This gives

$$\mathbf{1}^\top \boldsymbol{w} = \lambda_1 \mathbf{1}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\mu} + \lambda_2 \mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}.$$

Let $a = \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\mu}$, $b = \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1}\mathbf{1} = \mathbf{1}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\mu}$, and $c = \mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}$. Then $\mu_p = \lambda_1 a + \lambda_2 b$ and $1 = \lambda_1 b + \lambda_2 c$. Solving this system gives

$$\begin{bmatrix} \mu_p \\ 1 \end{bmatrix} = \begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix},$$

so

$$\lambda_1 = \frac{c\mu_p - b}{d} \qquad \text{and} \qquad \lambda_2 = \frac{a - b\mu_p}{d},$$

where $d = ac - b^2$.

Substituting these values into the optimal weights gives

$$\begin{aligned} \hat{\boldsymbol{w}} = \underbrace{\frac{a\boldsymbol{\Omega}^{-1}\mathbf{1} - b\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}}{d}}_{=:\Phi} + \underbrace{\frac{c\boldsymbol{\Omega}^{-1}\boldsymbol{\mu} - b\boldsymbol{\Omega}^{-1}\mathbf{1}}{d}}_{=:\Theta} \mu_p \implies \hat{\boldsymbol{w}} \\ = \Phi + \Theta \mu_p. \end{aligned}$$

The minimum variance equation of this portfolio is obtained from $\hat{\sigma}_p^2 = \hat{\boldsymbol{w}}^\top \boldsymbol{\Omega} \hat{\boldsymbol{w}}$. Once μ_p is fixed and the Markowitz solution $\hat{\boldsymbol{w}}$ is known, we obtain $\hat{\sigma}_p^2$. The shape of the **minimum variance frontier**, which is the graphical representation of the Markowitz problem, is either a hyperbola in (μ, σ)-space or a parabola in (μ, σ^2)-space. Algebraic manipulation gives

$$\hat{\sigma}_p^2 = \frac{c}{d}\left(\mu_p - \frac{b}{c}\right)^2 + \frac{1}{c},$$

which we call the **relation between risk and reward** along the minimum variance frontier.

Recall that, in general, $y = (x - B)^2 + E$ is a parabola with vertex at (B, E). In the present case, the vertex is at $(b/c, 1/c)$. Therefore the **minimum variance portfolio** has expected return b/c and variance $1/c$.

The lower branch of this parabola is inefficient whenever the upper branch offers the same variance with a higher expected return. Consequently, the efficient frontier is the upper branch of the minimum variance frontier beginning at the minimum variance portfolio. This is the part of the opportunity set relevant for a risk-averse investor who prefers more expected return for the same level of risk. Figure 10 shows both branches and the minimum variance portfolio.

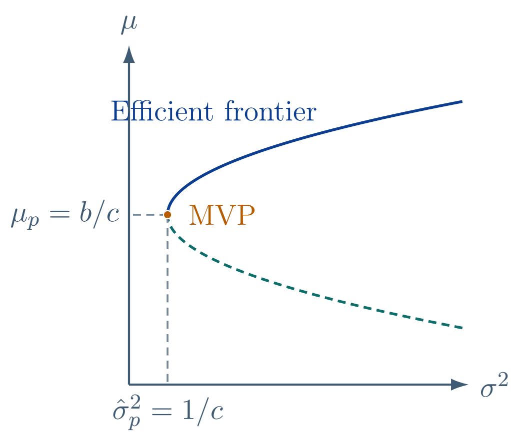


FIGURE 10

Example 8.1 Consider a portfolio of two stocks, with

$$\boldsymbol{\mu}_{2\times 1} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} \mathbb{E}[\tilde{r}_1] \\ \mathbb{E}[\tilde{r}_2] \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix},$$

and

$$\boldsymbol{\Omega} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 4 \end{bmatrix}.$$

Then

$$\boldsymbol{\Omega}^{-1} = \frac{1}{3}\begin{bmatrix} 4 & 1 \\ 1 & 1 \end{bmatrix},$$

so

$$a = \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{\mu} = 4,$$

$$b = \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1} \mathbf{1} = 3,$$

$$c = \mathbf{1}^\top \boldsymbol{\Omega}^{-1} \mathbf{1} = \frac{7}{3},$$

and

$$d = ac - b^2 = \frac{1}{3}.$$

Therefore,

$$\Phi = \frac{1}{d}(a\boldsymbol{\Omega}^{-1}\mathbf{1} - b\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}) = 3(4\boldsymbol{\Omega}^{-1}\mathbf{1} - 3\boldsymbol{\Omega}^{-1}\boldsymbol{\mu}) = \begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

and

$$\Theta = \frac{1}{d}(c\boldsymbol{\Omega}^{-1}\boldsymbol{\mu} - b\boldsymbol{\Omega}^{-1}\mathbf{1}) = \begin{bmatrix} -1 \\ 1 \end{bmatrix}.$$

So

$$\hat{\boldsymbol{w}} = \Phi + \Theta\mu_p = \begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix}\mu_p$$

is the vector of optimal weights for any μ_p.

Also,

$$\hat{\sigma}_p^2 = \frac{c}{d}\left(\mu_p - \frac{b}{c}\right)^2 + \frac{1}{c} = \frac{7/3}{1/3}\left(\mu_p - \frac{3}{7/3}\right)^2 + \frac{1}{7/3} = 7\left(\mu_p - \frac{9}{7}\right)^2 + \frac{3}{7},$$

which is the minimum variance frontier.

To identify the minimum variance portfolio, we have $(\mu_p)_{mvp} = b/c = 9/7$, $(\sigma_p^2)_{mvp} = 1/c = 3/7$, and so

$$\hat{\boldsymbol{w}}_{mvp} = \Phi + \Theta\mu_p = \begin{bmatrix} 2 \\ -1 \end{bmatrix} + \begin{bmatrix} -1 \\ 1 \end{bmatrix}\frac{9}{7} = \begin{bmatrix} 5/7 \\ 2/7 \end{bmatrix}.$$

As a sanity check, we observe that

$$\sum_{i=1}^{2} w_i = 1$$

as always.

Lecture 9 — Tuesday, June 03, 2014

Geometry of the Portfolio Frontier

We now describe the geometry of the portfolio frontier in several special cases. Begin with two risky assets, so that $w_1 + w_2 = 1$ and hence $w_2 = 1 - w_1$. The portfolio return is

$$\tilde{r}_p = w_1\tilde{r}_1 + (1 - w_1)\tilde{r}_2.$$

The covariance may be written as

$$\sigma_{12} = \rho_{12}\sigma_1\sigma_2,$$

where ρ_{12} is the correlation coefficient between the two assets. Therefore

$$\mu_p = \mathbb{E}[\tilde{r}_p] = w_1\mu_1 + (1 - w_1)\mu_2$$

and

$$\sigma_p^2 = \mathrm{Var}(\tilde{r}_p) = w_1^2\sigma_1^2 + (1 - w_1)^2\sigma_2^2 + 2w_1(1 - w_1)\rho_{12}\sigma_1\sigma_2.$$

If the two risky assets have $\rho_{12} = 1$, then the variance is a perfect square:

$$\sigma_p^2 = [w_1\sigma_1 + (1 - w_1)\sigma_2]^2.$$

Along the line segment between the two assets, this gives

$$\sigma_p = w_1\sigma_1 + (1 - w_1)\sigma_2.$$

Solving for the portfolio weights,

$$w_1 = \frac{\sigma_p - \sigma_2}{\sigma_1 - \sigma_2} \qquad \text{and} \qquad w_2 = \frac{\sigma_1 - \sigma_p}{\sigma_1 - \sigma_2}.$$

Substituting these weights into the equation for expected return gives

$$\mu_p = \mu_1 + \frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1}(\sigma_p - \sigma_1),$$

so the mean-standard-deviation frontier is a straight line. Its slope is

$$\frac{\mathrm{d}\mu_p}{\mathrm{d}\sigma_p} = \frac{\mu_2 - \mu_1}{\sigma_2 - \sigma_1}.$$

When $w_1 = 1$, the portfolio is Asset 1; when $w_1 = 0$, it is Asset 2. Thus, with perfect positive correlation, diversification does not bend the frontier, as Figure 11 makes clear.

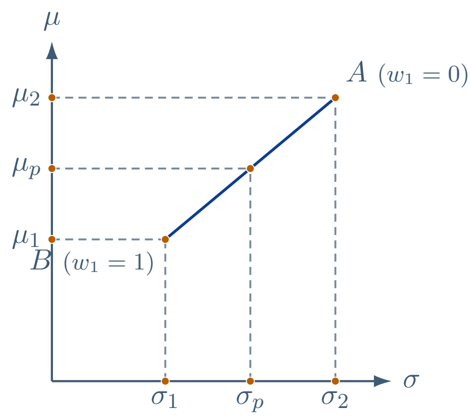


FIGURE 11

When $-1 < \rho_{12} < 1$, imperfect correlation makes diversification valuable. Relative to the straight line in the perfectly positively correlated case, the attainable set bends to the left in mean–standard-deviation space, as shown in Figure 12. In mean–variance space, the minimum variance frontier is a parabola; in mean–standard-deviation space, the upper branch is the efficient frontier.

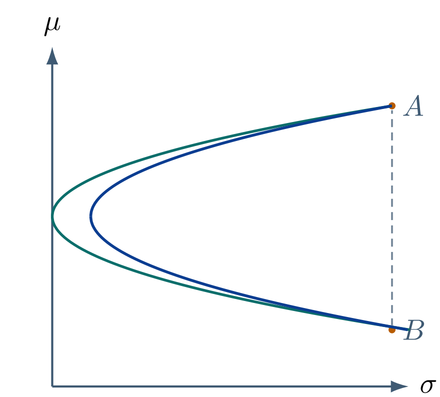


FIGURE 12

When $\rho_{12} = -1$, perfect negative correlation gives

$$\sigma_p^2 = [w_1\sigma_1 - (1 - w_1)\sigma_2]^2.$$

Solving for w_1 gives

$$w_1 = \frac{\pm\sigma_p + \sigma_2}{\sigma_1 + \sigma_2}.$$

Substituting this expression into the portfolio mean equation gives

$$\mu_p = \left(\frac{\sigma_2}{\sigma_1+\sigma_2}\right)\mu_1 + \left(\frac{\sigma_1}{\sigma_1+\sigma_2}\right)\mu_2 \pm \left(\frac{\mu_1-\mu_2}{\sigma_1+\sigma_2}\right)\sigma_p.$$

Thus $\mu_p = \alpha \pm \beta\sigma_p$ in this case. The minimum variance portfolio can be risk-free, since the two assets can offset each other's risk exactly. The resulting frontier with two segments is shown in Figure 13.

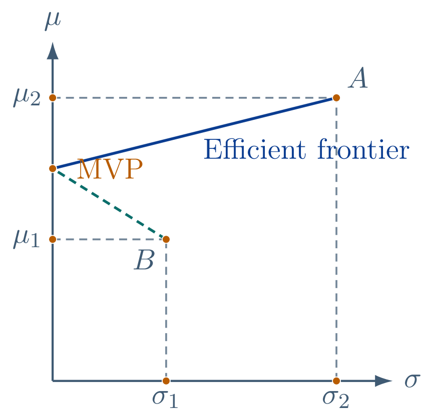


FIGURE 13

For n risky assets, suppose the covariance matrix is positive definite and expected returns are not all equal. From the algebra of the portfolio frontier, the minimum variance frontier is

$$\hat{\sigma}_p^2 = \frac{c}{d}\left(\mu_p - \frac{b}{c}\right)^2 + \frac{1}{c},$$

where

$$a = \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\mu}, \qquad b = \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}, \qquad c = \mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}, \qquad \text{and} \qquad d = ac - b^2.$$

The minimum variance portfolio has expected return b/c and variance $1/c$.

Now suppose Asset 1 is risk-free, so $\mu_1 = r_f$, and Asset 2 is risky, so $\mu_2 > r_f$ and $\sigma_1 = 0$.

For a lending portfolio with $0 \le w_1 \le 1$, the variance becomes

$$\sigma_p^2 = \mathrm{Var}(\tilde{r}_p) = (1-w_1)^2\sigma_2^2.$$

Thus $\sigma_p = (1-w_1)\sigma_2$ and $w_1 = 1 - \sigma_p/\sigma_2$. Substituting into the equation for expected return gives

$$\mu_p = w_1 r_f + (1-w_1)\mu_2 = r_f + \left(\frac{\mu_2 - r_f}{\sigma_2}\right)\sigma_p.$$

The relation between risk and return is again linear, now starting from the risk-free rate on the vertical axis, as in Figure 14.

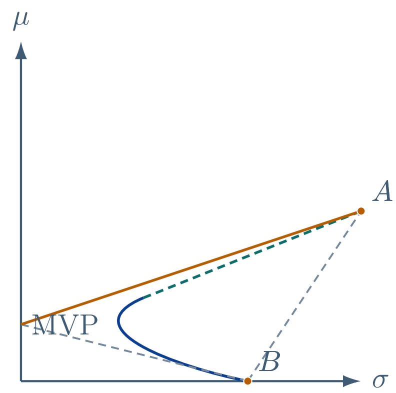


Figure 14

With many risky assets and one risk-free asset, the frontier for risky assets remains curved, but the efficient frontier becomes the straight line from r_f tangent to the risky frontier. The tangency point is denoted by T. Once the risk-free asset is available, any point on this line can be obtained by lending at r_f and investing in T, or by borrowing at r_f and investing more than 100% in T. Therefore

$$\mu_p = r_f + \left(\frac{\mu_T - r_f}{\sigma_T}\right)\sigma_p$$

on the efficient frontier.

The tangency condition is important because every other line drawn from r_f to the risky frontier has a smaller slope. Since the slope is excess return per unit of risk, the tangent portfolio is the risky portfolio with the highest attainable Sharpe ratio. Different investors may choose different positions along this line, but they all use the same risky fund T and only change the amount borrowed or lent at r_f. This geometry is shown in Figure 15.

The efficient frontier is the straight line from r_f through T. A conservative investor's relatively steep indifference map is shown in Figure 16.

Figure 17 then superimposes the risky frontier and its tangency line from the risk-free rate.

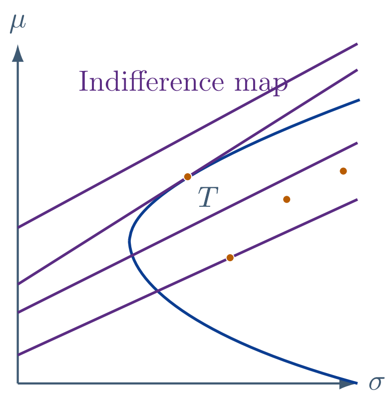


Figure 15

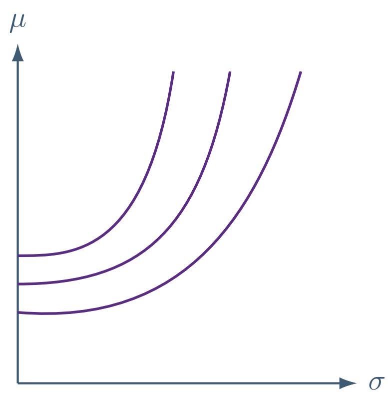


Figure 16

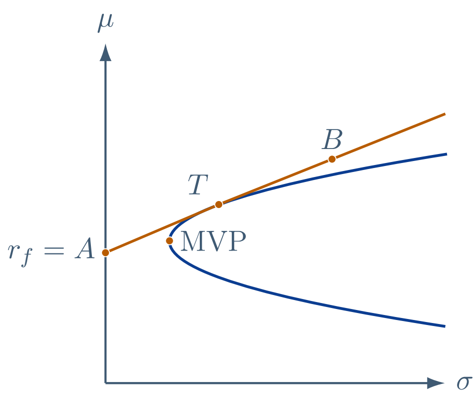


Figure 17

Lecture 10 — Thursday, June 05, 2014

Two-Fund Separation and the Capital Market Line

Under the Markowitz assumptions, investors are risk-averse, care only about $\mathbb{E}[\tilde{r}_p] = \mu_p$ and $\mathrm{Var}(\tilde{r}_p) = \sigma_p^2$, and invest over one period. The minimum variance portfolio weights for a target return μ_p then take the affine form

$$\hat{\boldsymbol{w}} = \Phi + \Theta\mu_p.$$

Theorem 10.1 (Two-fund spanning property) Any affine combination of two portfolios on the minimum variance frontier is again on the minimum variance frontier.

Proof. Let portfolios 1 and 2 lie on the minimum variance frontier. Then

$$\hat{\boldsymbol{w}}_1 = \Phi + \Theta\mu_1 \qquad \text{and} \qquad \hat{\boldsymbol{w}}_2 = \Phi + \Theta\mu_2.$$

Consider the portfolio p with weights

$$\hat{\boldsymbol{w}}_p = \alpha\hat{\boldsymbol{w}}_1 + (1-\alpha)\hat{\boldsymbol{w}}_2, \quad \alpha \in \mathbb{R}.$$

Its expected return is $\mu_p = \alpha\mu_1 + (1-\alpha)\mu_2$, and

$$\begin{aligned}
\hat{\boldsymbol{w}}_p &= \alpha(\Phi + \Theta\mu_1) + (1-\alpha)(\Phi + \Theta\mu_2) \\
&= \Phi + \Theta\big(\alpha\mu_1 + (1-\alpha)\mu_2\big) \\
&= \Phi + \Theta\mu_p.
\end{aligned}$$

Thus p has the frontier weight representation for its expected return, so it lies on the minimum variance frontier. $\square$

Theorem 10.2 (Two-fund theorem) Given two frontier portfolios with distinct expected returns, every minimum variance portfolio can be obtained as an affine combination of those two portfolios. In this sense, two mutual funds are enough to span the frontier.

Proof. Let the two spanning portfolios have expected returns $\mu_1 \neq \mu_2$. For any target return μ_p, choose

$$\alpha = \frac{\mu_p - \mu_2}{\mu_1 - \mu_2}.$$

Then $\mu_p = \alpha\mu_1 + (1-\alpha)\mu_2$. By the preceding spanning property,

$$\alpha\hat{\boldsymbol{w}}_1 + (1-\alpha)\hat{\boldsymbol{w}}_2 = \Phi + \Theta\mu_p,$$

which is exactly the minimum variance portfolio for the target return μ_p. $\square$

First, in (μ–σ)-space, minimizing σ^2 is equivalent to maximizing $\mathbb{E}[u(\tilde{y})]$ under the mean–variance assumptions. The equilibrium choices of conservative and aggressive investors are compared in Figure 18.

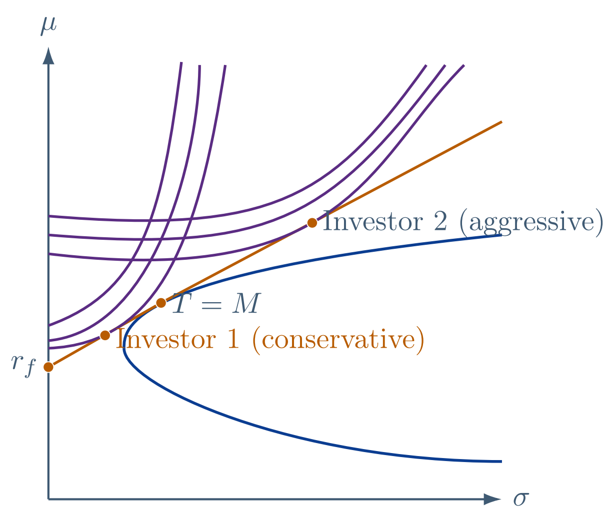


FIGURE 18

Second, introducing a risk-free asset allows borrowing and lending at the risk-free rate. If all investors mix the risk-free asset with the tangency portfolio T, then $T = M$, the market portfolio. To include both lending and borrowing, we extend the frontier line. There are three natural cases: $[r_f + L]$, corresponding to a conservative investor; $[r_f + B]$, corresponding to an aggressive investor; and $[r_f + M]$, corresponding to a moderate investor.

For a lending portfolio, let w_L denote the weight on portfolio L, and let $1 - w_L$ be the weight on the risk-free asset. Here $0 \le w_L \le 1$, so the remaining wealth is invested in the risk-free asset.

$$\begin{aligned}
\tilde{r}_L &= w_L\tilde{r}_M + (1-w_L)r_f, \\
\mu_L &= \mathbb{E}[\tilde{r}_L] = w_L\mathbb{E}[\tilde{r}_M] + (1-w_L)r_f = w_L\mu_M + (1-w_L)r_f, \\
\mathrm{Var}(\tilde{r}_L) &= w_L^2\sigma_M^2 + (1-w_L)^2\cancel{\mathrm{Var}(r_f)} + 2w_L(1-w_L)\cancel{\mathrm{Cov}(\tilde{r}_M, r_f)}.
\end{aligned}$$

Therefore $\sigma_L^2 = w_L^2 \sigma_M^2$, so $\sigma_L = w_L \sigma_M$. The risk of L is a fraction of market risk.

For a borrowing portfolio, borrow $w_B - 1$ at the risk-free rate, where $w_B \geq 1$, and invest w_B in the market portfolio. Then

$$\tilde{r}_B = w_B \tilde{r}_M - (w_B - 1) r_f,$$

so

$$\mu_B = \mathbb{E}[\tilde{r}_B] = w_B \mu_M - (w_B - 1) r_f,$$
$$\mathrm{Var}(\tilde{r}_B) = w_B^2 \sigma_M^2 \implies \sigma_B = w_B \sigma_M.$$

Thus, regardless of risk aversion, all investors construct efficient portfolios by combining the risk-free asset with the same market portfolio M. The relation $\sigma_k = w_k \sigma_M$ holds for any such portfolio k, including portfolios formed through borrowing or lending.

For every risky portfolio with positive volatility, the **Sharpe ratio** is its excess expected return per unit of volatility:

$$\frac{\mathbb{E}[\tilde{r}_p] - r_f}{\sigma_p}.$$

On the capital market line, this ratio is the slope of the efficient frontier. If $p = M$, the maximum Sharpe ratio is

$$\frac{\mathbb{E}[\tilde{r}_M] - r_f}{\sigma_M}.$$

Finally, the equilibrium assumptions are:

1. Borrowing and lending are possible at the risk-free rate.
2. Aggregate borrowing equals aggregate lending when the credit market clears.
3. Investors are (μ–σ^2)-utility maximizers.
4. Investors have homogeneous beliefs about asset characteristics, so all investors face the same efficient frontier.

These assumptions imply that M is the market portfolio and that the line through r_f and M is the capital market line. Its equation is

$$\mathbb{E}[\tilde{r}_p] = r_f + \left(\frac{\mathbb{E}[\tilde{r}_M] - r_f}{\sigma_M} \right) \sigma_p.$$

Thus, in equilibrium, optimal portfolios lie on the capital market line, where expected utility is maximized.

Lecture 12 — Thursday, June 12, 2014

Capital Asset Pricing Model

The capital market line gives the relationship between expected return and total risk only for efficient portfolios. Individual assets need not be efficient portfolios, so their required returns cannot be read directly from the CML. CAPM answers this by asking how much an individual asset contributes to the risk of the market portfolio.

Theorem 12.1 (CAPM interpretation) If CAPM holds, only asset j's covariance with the market, rather than all of its stand-alone volatility, is priced.

Proof. The CAPM relation, derived in the next lecture, gives

$$\begin{aligned}\mathbb{E}[\tilde{r}_j] - r_f &= \frac{\sigma_{j,m}}{\sigma_m} \cdot \frac{\mathbb{E}[\tilde{r}_m] - r_f}{\sigma_m} \\ &= \text{signed market-linked exposure} \cdot \text{market price of risk.}\end{aligned}$$

By definition,

$$\rho_{j,m} = \frac{\sigma_{j,m}}{\sigma_j \sigma_m}.$$

Hence

$$\sigma_{j,m} = \rho_{j,m}\sigma_j\sigma_m.$$

Substituting this identity above gives

$$\mathbb{E}[\tilde{r}_j] - r_f = \rho_{j,m}\sigma_j \frac{\mathbb{E}[\tilde{r}_m] - r_f}{\sigma_m}.$$

Thus the priced exposure is the signed quantity $\rho_{j,m}\sigma_j = \mathrm{Cov}(\tilde{r}_j, \tilde{r}_m)/\sigma_m$, which may be negative. The full CAPM relation is derived in the next lecture. □

The relevant contribution is covariance with the market, not the asset's stand-alone standard deviation.

Unsystematic risk is firm-specific risk, that is, idiosyncratic risk. It is not rewarded in equilibrium because it can be diversified away in a broad portfolio. Systematic risk cannot be diversified away, so the market compensates investors for bearing it through the beta term in CAPM.

Recall that Markowitz analysis is based on two assumptions: first, investors are risk-averse; second, they care only about μ and σ. In this simplified presentation, the **capital asset pricing model** (**CAPM**) adds five assumptions:

1. Investors are risk-averse: $u' > 0$ and $u'' < 0$, and they maximize $\mathbb{E}[u(\tilde{w}_1)]$.
2. Investors have homogeneous beliefs about μ and Σ.
3. Investors can borrow and lend at r_f.
4. Aggregate borrowing equals aggregate lending when the credit market clears.
5. There are no transaction costs.

Together with the implicit one-period competitive setting, common investment horizon, divisibility of securities, and inclusion of all traded risky assets in the market portfolio, these assumptions imply that investors face the same tangency portfolio. Market clearing then makes that tangency portfolio the market portfolio, so all assets must be priced on the same security market line. Therefore the expected return of any asset j must satisfy

$$\mathbb{E}[\tilde{r}_j] = r_f + \beta_{j,m}(\mathbb{E}[\tilde{r}_m] - r_f).$$

This is the **CAPM formula**.

Example 12.2 (Inferring beta from the security market line) Consider two stocks, A and B, with

$$\mathbb{E}[\tilde{r}_A] = 10.5\%, \qquad \beta_A = 0.7, \qquad \mathbb{E}[\tilde{r}_B] = 13\%, \qquad \text{and} \qquad \beta_B = 1.2.$$

For a third stock D, suppose $\mathbb{E}[\tilde{r}_D] = 12\%$.

First determine β_D. If CAPM holds, A and B both lie on the security market line, so

$$\mathbb{E}[\tilde{r}_A] = r_f + \beta_A(\mathbb{E}[\tilde{r}_m] - r_f) \qquad \text{and} \qquad \mathbb{E}[\tilde{r}_B] = r_f + \beta_B(\mathbb{E}[\tilde{r}_m] - r_f).$$

Subtracting the two equations gives

$$13\% - 10.5\% = (1.2 - 0.7)(\mathbb{E}[\tilde{r}_m] - r_f),$$

so $\mathbb{E}[\tilde{r}_m] - r_f = 5\%$. Substituting into the equation for A gives $r_f = 7\%$, hence $\mathbb{E}[\tilde{r}_m] = 12\%$. Therefore

$$12\% = 7\% + \beta_D(5\%),$$

so $\beta_D = 1$.

The location of D relative to A and B on the security market line is shown in Figure 19.

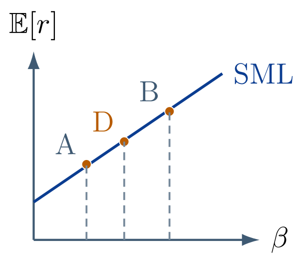


FIGURE 19

Next, given $P_1 = \$124$ and $D_1 = \$2$, compute P_0 for Stock A using

$$\mathbb{E}[\tilde{r}_A] = \frac{P_1 - P_0 + D_1}{P_0}.$$

Thus

$$0.105 = \frac{124 - P_0 + 2}{P_0},$$

so $P_0 = \$114.03$.

In general, risk decomposes into systematic and idiosyncratic components. Volatility is measured by variance or standard deviation.

The CAPM equation

$$\mathbb{E}[\tilde{r}_j] = r_f + \beta_{j,m}(\mathbb{E}[\tilde{r}_m] - r_f),$$

is a relation between expected returns. An empirical time-series regression allows both an intercept and an error term:

$$\tilde{r}_{jt} - r_{ft} = \alpha_j + \beta_{j,m}(\tilde{r}_{mt} - r_{ft}) + \tilde{\varepsilon}_{jt}.$$

Under CAPM, the population intercept satisfies $\alpha_j = 0$. The term $\tilde{\varepsilon}_{jt}$ captures idiosyncratic shocks. Diversification is used to reduce this component.

Theorem 12.3 (Diversification of idiosyncratic risk) Suppose the residuals have mean zero, are uncorrelated with the market, are pairwise uncorrelated across stocks, and have uniformly bounded variances. For an equally weighted portfolio, the residual variance converges to zero as the number of stocks tends to infinity.

Proof. Let

$$\begin{aligned}\tilde{r}_p &= \sum_{i=1}^{n} w_i \tilde{r}_{it} \\ &= \sum_{i=1}^{n} w_i \left[r_f + \beta_{i,m}(\tilde{r}_{mt} - r_f) + \tilde{\varepsilon}_{it} \right] \\ &= \sum_{i=1}^{n} w_i r_f + \sum_{i=1}^{n} w_i \beta_{i,m}(\tilde{r}_{mt} - r_f) + \sum_{i=1}^{n} w_i \tilde{\varepsilon}_{it}.\end{aligned}$$

Let r_f be constant. Since

$$\sum_{i=1}^{n} w_i = 1 \qquad \text{and} \qquad \sum_{i=1}^{n} w_i \beta_{i,m} = \beta_{p,m},$$

we get

$$\tilde{r}_p = r_f + \beta_{p,m}(\tilde{r}_{mt} - r_f) + \sum_{i=1}^{n} w_i \tilde{\varepsilon}_{it}.$$

This is the return decomposition for portfolio p. For an equally weighted portfolio, $w_i = 1/n$, and the residual term converges to zero in mean square under the stated assumptions.

Writing $\bar{\sigma}_\varepsilon^2 = n^{-1} \sum_{i=1}^{n} \mathrm{Var}(\tilde{\varepsilon}_{it})$, we obtain

$$\begin{aligned}\sigma_p^2 &= \mathrm{Var}(\tilde{r}_p) \\ &= \mathrm{Var}\big(\beta_{p,m}\tilde{r}_{mt}\big) + \mathrm{Var}\left(\sum_{i=1}^{n} w_i \tilde{\varepsilon}_{it}\right) \\ &= \beta_{p,m}^2 \sigma_m^2 + \mathrm{Var}\left(\sum_{i=1}^{n} \frac{1}{n} \tilde{\varepsilon}_{it}\right)\end{aligned}$$

$$= \beta_{p,m}^2 \sigma_m^2 + \frac{1}{n} \bar{\sigma}_\varepsilon^2.$$

As $n \to \infty$,

$$\sigma_p^2 - \beta_{p,m}^2 \sigma_m^2 \longrightarrow 0.$$

If the sequence of portfolio betas also converges to some β, then $\sigma_p^2 \to \beta^2 \sigma_m^2$. $\square$

Lecture 13 — Tuesday, June 17, 2014

Theorem 13.1 (Sharpe–Lintner CAPM) In equilibrium, expected returns are pinned down by the risk-free rate and market beta. Equivalently, only systematic risk is priced:

$$\mathbb{E}[\tilde{r}_j] = r_f + \beta_{j,m}(\mathbb{E}[\tilde{r}_m] - r_f).$$

Proof. Assume $\mathbf{\Omega}$ is positive definite. Let r_f denote the risk-free rate, let

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

be the vector of expected returns on the risky assets, and let $\boldsymbol{w}$ be the vector of weights assigned to them. Once the risk-free asset is available, the expected return on a portfolio is not simply $\boldsymbol{w}^\top \boldsymbol{\mu}$, because some wealth may be placed in the risk-free asset. The correct expression is

$$\mu_p = \boldsymbol{w}^\top \boldsymbol{\mu} + (1 - \boldsymbol{w}^\top \mathbf{1}) r_f = r_f + \boldsymbol{w}^\top (\boldsymbol{\mu} - r_f \mathbf{1}).$$

To derive the efficient portfolios, fix a target excess return $\mu_p - r_f$ and minimize variance:

$$\min_{\boldsymbol{w}} \frac{1}{2} \boldsymbol{w}^\top \mathbf{\Omega} \boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}^\top (\boldsymbol{\mu} - r_f \mathbf{1}) = \mu_p - r_f.$$

The Lagrangian is

$$\mathcal{L}(\boldsymbol{w}, \lambda) = \frac{1}{2} \boldsymbol{w}^\top \mathbf{\Omega} \boldsymbol{w} + \lambda \left[\mu_p - r_f - \boldsymbol{w}^\top (\boldsymbol{\mu} - r_f \mathbf{1}) \right].$$

The first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{w}} = \mathbf{0} \Longleftrightarrow \boldsymbol{\Omega w} - \lambda(\boldsymbol{\mu} - r_f \mathbf{1}) = \mathbf{0},$$

and

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0 \Longleftrightarrow \boldsymbol{w}^\top(\boldsymbol{\mu} - r_f \mathbf{1}) = \mu_p - r_f.$$

The first condition shows that every efficient risky portfolio must be proportional to

$$\boldsymbol{\Omega}^{-1}(\boldsymbol{\mu} - r_f \mathbf{1}),$$

so the covariance matrix and the vector of excess expected returns jointly determine the optimal direction of investment.

Solving the first condition for $\boldsymbol{w}$ gives

$$\boldsymbol{w} = \lambda \boldsymbol{\Omega}^{-1}(\boldsymbol{\mu} - r_f \mathbf{1}).$$

Substituting this into the return constraint yields

$$\lambda \left[\boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\mu} - r_f \mathbf{1}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\mu} - r_f \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1}\mathbf{1} + r_f^2 \mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1}\right] = \mu_p - r_f.$$

If

$$a = \boldsymbol{\mu}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\mu}, \qquad b = \mathbf{1}^\top \boldsymbol{\Omega}^{-1}\boldsymbol{\mu}, \qquad \text{and} \qquad c = \mathbf{1}^\top \boldsymbol{\Omega}^{-1}\mathbf{1},$$

and $h = a - 2br_f + cr_f^2$, then $h > 0$ provided that $\boldsymbol{\mu} - r_f \mathbf{1} \neq \mathbf{0}$, and

$$\lambda = \frac{\mu_p - r_f}{h},$$

so the efficient portfolio weights are

$$\boldsymbol{w} = \boldsymbol{\Omega}^{-1}(\boldsymbol{\mu} - r_f \mathbf{1}) \left(\frac{\mu_p - r_f}{h}\right).$$

The key covariance step is that for any efficient portfolio p and any portfolio q, where $\boldsymbol{w}_q$ records the risky-asset weights and the remaining wealth is invested at r_f,

$$\mathrm{Cov}(\tilde{r}_p, \tilde{r}_q) = \boldsymbol{w}_p^\top \boldsymbol{\Omega} \boldsymbol{w}_q = \lambda \boldsymbol{w}_q^\top (\boldsymbol{\mu} - r_f \mathbf{1}).$$

By the portfolio return identity, $\boldsymbol{w}_q^\top(\boldsymbol{\mu} - r_f\mathbf{1}) = \mu_q - r_f$, so

$$\mathrm{Cov}(\tilde{r}_p, \tilde{r}_q) = \left(\frac{\mu_p - r_f}{h}\right)(\mu_q - r_f).$$

Setting $q = p$ gives

$$\mathrm{Var}(\tilde{r}_p) = \frac{(\mu_p - r_f)^2}{h},$$

and therefore, on the upper efficient branch where $\mu_p \geq r_f$,

$$\sigma_p = \frac{\mu_p - r_f}{\sqrt{h}}.$$

This is the capital market line written in terms of the efficient frontier with a risk-free asset:

$$\mu_p = r_f + \sqrt{h}\,\sigma_p.$$

Finally, define the **beta** relative to portfolio p by

$$\beta_{q,p} = \frac{\mathrm{Cov}(\tilde{r}_q, \tilde{r}_p)}{\mathrm{Var}(\tilde{r}_p)}.$$

Then

$$\mu_q - r_f = \beta_{q,p}(\mu_p - r_f),$$

so

$$\mathbb{E}[\tilde{r}_q] = r_f + \beta_{q,p}\big(\mathbb{E}[\tilde{r}_p] - r_f\big).$$

Taking p to be the market portfolio yields the familiar CAPM formula

$$\mathbb{E}[\tilde{r}_q] = r_f + \beta_{q,m}(\mathbb{E}[\tilde{r}_m] - r_f).$$

□

Recall that the Sharpe ratio is

$$\frac{\mu_p - r_f}{\sigma_p}.$$

Among portfolios of risky assets, the tangency portfolio T maximizes this ratio. Every positive exposure to T, combined with borrowing or lending at the risk-free rate, lies on the same capital market line and has the same Sharpe ratio. The pure risk-free portfolio has zero volatility, so its Sharpe ratio is undefined rather than equal to this common value.

The **security market line** (**SML**) is the graphical representation of CAPM in μ-β space. The **capital market line** (**CML**) is the graphical representation of Markowitz theory in μ-σ space. Here β denotes systematic risk, while σ denotes volatility. The CML applies to efficient portfolios, whereas the SML prices individual assets and portfolios by beta. Their different horizontal axes and slopes are compared in Figure 20.

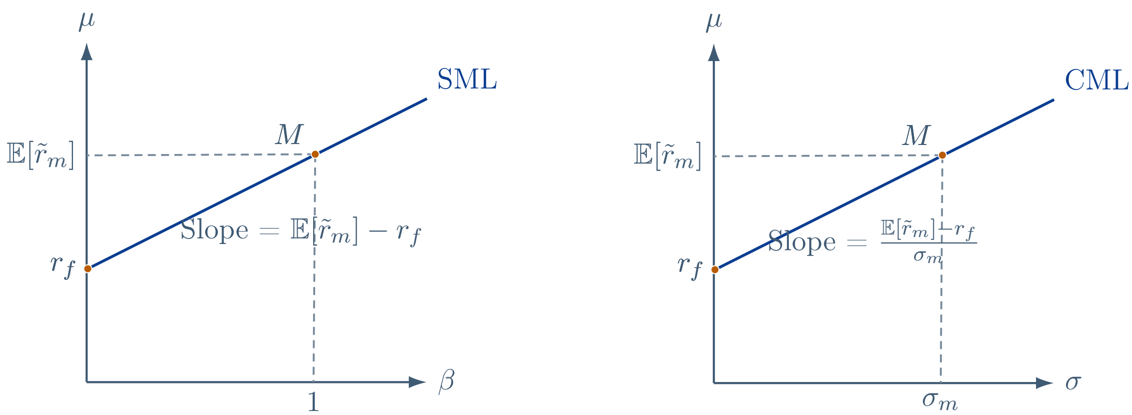


FIGURE 20

We have

$$\mathbb{E}[\tilde{r}_i] = r_f + \beta_{i,m}(\mathbb{E}[\tilde{r}_m] - r_f)$$

so

$$\mu_p = \sum_{i=1}^{n} w_i \mathbb{E}[\tilde{r}_i] \qquad \text{and} \qquad \beta_{p,m} = \sum_{i=1}^{n} w_i \beta_{i,m},$$

and therefore

$$\mathbb{E}[\tilde{r}_p] = r_f + \beta_{p,m}(\mathbb{E}[\tilde{r}_m] - r_f).$$

The CML is limited to efficient portfolios. The SML is the CAPM pricing relation and applies to individual assets as well as portfolios.

Lecture 14 — Thursday, June 19, 2014

4. Factor Models and Arbitrage Pricing Theory

Arbitrage pricing theory (**APT**) explains variation in the rate of return $\tilde{r}_i$, namely $\tilde{r}_i - \mathbb{E}[\tilde{r}_i]$, through sensitivity to common factors. The factor in deviation form is $\tilde{F} = \tilde{f} - \mathbb{E}[\tilde{f}]$, where $\tilde{f}$ is the raw factor. Factor models and the absence of arbitrage are the foundation of APT.

Factor Models and Common Risk Exposures

Common factors are economy-wide shocks that affect many securities at once. Examples include changes in GDP growth, inflation, interest rates, or labour costs. The point of a factor model is not that every firm reacts in the same way, but that the same underlying shock can matter with different magnitudes across sectors. Transportation firms, banks, and utilities may all respond to changes in interest rates, but their sensitivities need not be equal.

In a **single factor model**, we write

$$\tilde{r}_i - \mathbb{E}[\tilde{r}_i] = b_i \tilde{F} + \tilde{\varepsilon}_i.$$

Here $b_i \tilde{F}$ is the **systematic component** and $\tilde{\varepsilon}_i$ is **idiosyncratic**. We normalize $\mathbb{E}[\tilde{F}] = \mathbb{E}[\tilde{\varepsilon}_i] = 0$ and impose $\operatorname{Cov}(\tilde{F}, \tilde{\varepsilon}_i) = 0$, with $\operatorname{Var}(\tilde{F}) > 0$. The empirical version is $\tilde{r}_i = \alpha_i + b_i \tilde{f} + \tilde{\varepsilon}_i$. If we take $\tilde{f}$ to be a broad index of securities, such as $\tilde{r}_m$, then

$$\tilde{r}_i - \mathbb{E}[\tilde{r}_i] = b_i(\tilde{r}_m - \mathbb{E}[\tilde{r}_m]) + \tilde{\varepsilon}_i,$$

which is the **single-index market model**. Therefore

$$b_i = \frac{\operatorname{Cov}(\tilde{r}_i, \tilde{r}_m)}{\operatorname{Var}(\tilde{r}_m)} \equiv \beta_{i,m}.$$

If $y = a + bx$, then

$$b = \frac{\operatorname{Cov}(x, y)}{\operatorname{Var}(x)}.$$

Hence $b_i = \beta_{i,m}$ if and only if we have a single factor model and the factor is $\tilde{r}_m$.

In a **multifactor model**, we write

$$\tilde{r}_i - \mathbb{E}[\tilde{r}_i] = b_{i1}\tilde{F}_1 + b_{i2}\tilde{F}_2 + \cdots + b_{ik}\tilde{F}_k + \tilde{\varepsilon}_i.$$

Here b_{i1} is not generally the same as a CAPM beta. The model still has systematic and idiosyncratic components, but now there are k factors affecting stock i, and b_{ik} is the partial sensitivity of the return on stock i to factor k. The loading vector is the population multiple-regression coefficient vector, with the residual orthogonal to the span of all factors. When factors are correlated, an individual loading is not obtained from a single covariance ratio that ignores the other factors.

Example 14.1 Let $\tilde{IN}$ denote the market interest rate. Let M^D be money demand and M^S be money supply. Their interaction determines the equilibrium interest rate IN^*.

More explicitly, money supply is

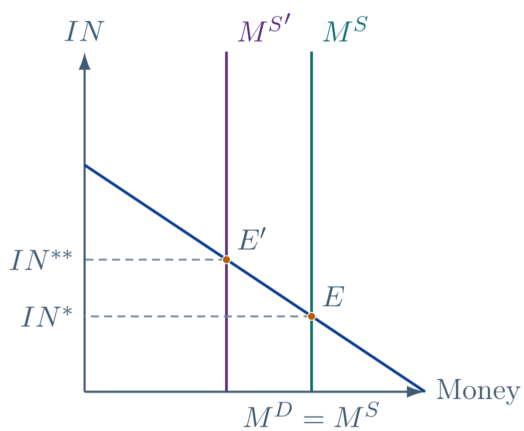
$$M^S = C + D,$$

that is, currency in circulation plus bank deposits. In this simple discussion it is treated as fixed and controlled by the Bank of Canada. Money demand is written as

$$M^D = \gamma Y - \delta \tilde{IN},$$

where γ and δ measure the sensitivity of money demand to income and to the interest rate.

The economic intuition is straightforward: the interest rate is the opportunity cost of holding money rather than an interest-bearing asset. When IN rises, holding idle cash becomes more expensive and money demand falls. When IN falls, that opportunity cost declines and people are willing to hold more money. This is why money demand decreases with the interest rate in the equation above.



Lecture 15 — Tuesday, June 24, 2014

Arbitrage Pricing Theory and Factor Risk Premia

APT has its roots in the multifactor model and is grounded in the absence of arbitrage. In an **arbitrage-free market**, there is no persistent opportunity to buy low and sell high without bearing corresponding risk.

The no-arbitrage restrictions behind APT can be seen from two simple arbitrage arguments.

Example 15.1 (Arbitrage with equal factor loadings) Consider two well-diversified portfolios A and B, so idiosyncratic risk has been diversified away. Suppose $b_A = b_B$, while $\mathbb{E}[\tilde{r}_A] = 12\%$ and $\mathbb{E}[\tilde{r}_B] = 8\%$. Since the two portfolios have the same factor loading, they carry the same systematic risk. Going long in A gives $\mathbb{E}[\tilde{r}_A] + b_A\tilde{F}$, while shorting B gives $-\mathbb{E}[\tilde{r}_B] - b_B\tilde{F}$. The factor terms cancel because $b_A = b_B$, so the net return is the riskless amount

$$\mathbb{E}[\tilde{r}_A] - \mathbb{E}[\tilde{r}_B] = 4\%.$$

This is an arbitrage opportunity.

The first no-arbitrage condition is therefore that well-diversified portfolios with equal factor loadings must have the same expected return. Otherwise we can form a long–short position with zero risk and a positive payoff.

Example 15.2 (Arbitrage from a mispriced factor exposure) Suppose $\mathbb{E}[\tilde{r}_m] = 11\%$ in a single factor model and $b_m = 1$. Also suppose a well-diversified portfolio D has $\mathbb{E}[\tilde{r}_D] = 7\%$ and $b_D = 0.5$, while $r_f = 5\%$. Since $b_m = 1$, we take M to be the market portfolio. The security market line is then

$$\mathbb{E}[\tilde{r}_i] = 5\% + 6\% \cdot b_i.$$

Since D lies off the security market line, it is overpriced or underpriced relative to its factor exposure. Construct a portfolio N by mixing the risk-free asset with M so that $\mathbb{E}[\tilde{r}_N] = 8\%$. Then

$$8\% = wr_f + (1-w)\mathbb{E}[\tilde{r}_m] = w(5\%) + (1-w)(11\%),$$

which gives $w = 0.5$. Thus we place 50% in the risk-free asset and 50% in the market portfolio. This portfolio has the same beta as D but a higher expected return, so going long in N and short in D yields a profit of 1% with no factor mismatch.

The second no-arbitrage condition is that expected excess returns on well-diversified portfolios must be linear in their factor loadings. In a one-factor model, this reduces to proportionality between expected excess return and the single loading.

Ross's APT pricing relation is derived from the following multifactor model:

$$\tilde{r}_i = \mathbb{E}[\tilde{r}_i] + b_{i1}\tilde{F}_1 + b_{i2}\tilde{F}_2 + \cdots + b_{ik}\tilde{F}_k + \tilde{\varepsilon}_i. \tag{1}$$

For a portfolio p of n stocks:

$$\tilde{r}_p = \sum_{i=1}^{n} w_i \tilde{r}_i. \tag{2}$$

Substituting (1) into (2) gives

$$\tilde{r}_p = \sum_{i=1}^{n} w_i \mathbb{E}[\tilde{r}_i] + \sum_{j=1}^{k} \tilde{F}_j \left(\sum_{i=1}^{n} w_i b_{ij} \right) + \sum_{i=1}^{n} w_i \tilde{\varepsilon}_i.$$

According to APT, a no-arbitrage portfolio should satisfy three properties:

1. *Self-financed*: the portfolio is constructed with zero change in wealth, so $\boldsymbol{w}^\top \mathbf{1} = 0$.
2. *Well-diversified*: no position is large enough for idiosyncratic risk to survive aggregation, so $\max_i |w_i|$ is small.
3. *Zero sensitivity to common factors*: $\boldsymbol{w}^\top \boldsymbol{b}_j = 0$ for each factor j.

In the idealized large-portfolio limit, these three conditions eliminate common factor exposure and diversify away idiosyncratic risk, while the self-financing condition ensures that no net investment is required. The limiting portfolio would therefore create an arbitrage unless its expected return were zero. Hence the APT restriction is

$$\mathbb{E}[\tilde{r}_p] = \sum_{i=1}^{n} w_i \mathbb{E}[\tilde{r}_i] = 0,$$

so $\boldsymbol{w}^\top \boldsymbol{\mu} = 0$.

Because every admissible well-diversified portfolio with zero investment and zero beta must also satisfy $\boldsymbol{w}^\top \boldsymbol{\mu} = 0$, the expected return vector $\boldsymbol{\mu}$ must lie in the span of the vectors that already define those restrictions. The intuition is a version of no arbitrage based on linear algebra: if expected returns had a component outside the span of the constant payoff and the factor loading vectors, then we could choose weights orthogonal to all priced risks but not orthogonal to expected returns, producing an arbitrage portfolio with zero cost. In the case with one factor, that means $\boldsymbol{\mu}$ must be a linear combination of $\mathbf{1}$ and $\boldsymbol{b}$:

$$\boldsymbol{\mu} = \lambda_0 \mathbf{1} + \lambda_1 \boldsymbol{b}.$$

For asset i and one factor k, $\mathbb{E}[\tilde{r}_i] = \lambda_0 + \lambda_k b_{ik}$. For one asset i and many factors,

$$\mathbb{E}[\tilde{r}_i] = \lambda_0 + \lambda_1 b_{i1} + \cdots + \lambda_k b_{ik}.$$

To interpret the λ_k, assume for simplicity that there is one common factor, so $\mathbb{E}[\tilde{r}_i] = \lambda_0 + \lambda_k b_{ik}$. We identify $\lambda_0 = r_f$, so λ_k measures the premium per unit of exposure to factor k. Thus

$$\mathbb{E}[\tilde{r}_i] = r_f + \lambda_k b_{ik}.$$

This is the APT analogue of the CAPM pricing equation. If portfolio p is a pure factor portfolio with unit sensitivity to factor k, so $b_{pk} = 1$, then the slope of the APT pricing line is

$$\lambda_k = \mathbb{E}[\tilde{r}_p] - r_f.$$

Hence

$$\mathbb{E}[\tilde{r}_i] = r_f + b_{ik}(\mathbb{E}[\tilde{r}_p] - r_f).$$

If the single priced factor is the market return and $p = M$ is a unit-loading factor portfolio, this equation coincides with CAPM. APT is otherwise a different, more flexible framework: it does not require a special role for the market portfolio, does not impose a particular distribution for asset returns, and allows several common factors to price expected returns.

Lecture 16 — Thursday, June 26, 2014

The same representation in terms of factor prices can be written in vector form. To do so, summarize the main APT conditions once more.

1. $\boldsymbol{w}^\top \mathbf{1} = 0$, so the portfolio is self-financed.

2. The portfolio is well diversified, so idiosyncratic risk is diversified away.
3. $\boldsymbol{w}^\top \boldsymbol{b}_k = 0$ for every factor k, that is,

$$[w_1 \dots w_n] \begin{bmatrix} b_{1k} \\ \vdots \\ b_{nk} \end{bmatrix} = 0.$$

   The constructed portfolio then has zero net exposure to every common factor; in the diversified limit, no arbitrage requires $\mathbb{E}[\tilde{r}_p] = 0$.
4. $\boldsymbol{w}^\top \boldsymbol{\mu} = 0$.

Conditions (1), (3), and (4) imply that $\boldsymbol{\mu} = \lambda_0 \mathbf{1} + \lambda_1 \boldsymbol{b}$. For one factor k,

$$\begin{bmatrix} \mathbb{E}[\tilde{r}_1] \\ \vdots \\ \mathbb{E}[\tilde{r}_n] \end{bmatrix} = \lambda_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \lambda_1 \begin{bmatrix} b_{1k} \\ \vdots \\ b_{nk} \end{bmatrix}.$$

For $K > 1$ factors,

$$\begin{bmatrix} \mathbb{E}[\tilde{r}_1] \\ \vdots \\ \mathbb{E}[\tilde{r}_n] \end{bmatrix} = \lambda_0 \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} + \lambda_1 \begin{bmatrix} b_{1,1} \\ \vdots \\ b_{n,1} \end{bmatrix} + \lambda_2 \begin{bmatrix} b_{1,2} \\ \vdots \\ b_{n,2} \end{bmatrix} + \cdots + \lambda_K \begin{bmatrix} b_{1,K} \\ \vdots \\ b_{n,K} \end{bmatrix}.$$

5. Capital Structure and Dividend Policy

This discussion follows the standard managerial finance treatment of the cost of capital and capital structure, including Gitman's *Principles of Managerial Finance.*

Cost of Capital and Financing Sources

The **cost of capital** is the rate of return that a firm must earn on projects in which it invests to maintain the market value of its stock. It can also be thought of as the rate of return required by the market suppliers of capital to attract their funds to the firm.

In general, if risk is held constant, projects with a rate of return above the cost of capital increase the value of the firm, while projects with a rate of return below the cost of capital decrease it.

Thus the cost of capital is a hurdle rate, but only for projects with risk comparable to the firm's existing operations. If a project has materially different business risk, the required return must be adjusted rather than mechanically using the firm's current WACC.

To isolate the basic structure of the cost of capital, we set forth the following assumptions regarding risk and taxes:

1. *Business risk* — the risk to the firm of being unable to cover operating costs — is assumed to be unchanged.
2. *Financial risk* — the risk to the firm of being unable to cover its financial obligations — is assumed to be unchanged.
3. *After-tax costs* are considered relevant, that is, the cost of capital is measured on an after-tax basis.

The four basic sources of long-term funds for a firm are long-term debt, preferred stock, common stock, and retained earnings.

Balance sheet

Assets	Current liabilities
	Long-term debt
	Equities
	· Preferred stock
	· Common stock
	· Retained earnings

These sources appear on the right-hand side of the firm's balance sheet. The specific cost of each source of financing is the after-tax cost of obtaining that financing today.

For long-term debt, let k_b denote the before-tax cost of debt, let I denote the annual interest payment, let n denote the years to maturity, let N_b denote the net proceeds from selling debt after flotation costs, let F denote face value, let t denote the tax rate, and let k_d denote the after-tax cost of debt. The exact yield k_b solves

$$N_b = \sum_{j=1}^{n} \frac{I}{(1+k_b)^j} + \frac{F}{(1+k_b)^n}.$$

A commonly used approximation is

$$k_b \approx \frac{I + \frac{F - N_b}{n}}{\frac{N_b + F}{2}},$$

and therefore

$$k_d = k_b(1 - t).$$

The cost of common stock is the required rate of return demanded by equity investors. Under the **Gordon growth model**,

$$P_0 = \frac{D_1}{k_s - g},$$
$$k_s = \frac{D_1}{P_0} + g.$$

where $k_s > g$. Here P_0 is the current stock price, D_1 is the expected dividend per share with $D_1 = D_0(1 + g)$, D_0 is the last dividend paid, and g is the dividend growth rate. Alternatively, the capital asset pricing model gives

$$k_s = r_f + b(\mathbb{E}[\tilde{r}_M] - r_f),$$

where k_s is the required return on common stock.

For retained earnings, let k_{RE} denote the cost of retained earnings. The cost of retained earnings to the firm is the same as the cost of an equivalent fully subscribed issue of additional common stock, so

$$k_{RE} = k_s.$$

If new common stock is issued and N_n denotes net proceeds per share after flotation costs, its cost is instead

$$k_n = \frac{D_1}{N_n} + g,$$

which generally exceeds $k_{RE} = k_s$ because $N_n < P_0$.

For preferred stock, let k_p denote the cost of preferred stock, let D_p denote the fixed preferred stock dividend, and let N_p denote net proceeds from stock issuance. Then

$$k_p = \frac{D_p}{N_p}.$$

Because preferred stock dividends are paid out of after-tax cash flow, no tax adjustment is required.

Given these component costs, let w_d, w_p, and w_s denote the proportions of long-term debt, preferred stock, and common stock in the capital structure, respectively. Then

$$w_d + w_p + w_s = 1.$$

The **weighted average cost of capital** (**WACC**) is then

$$\text{WACC} = w_d k_d + w_p k_p + w_s \begin{cases} k_n, & \text{if new common stock is issued,} \\ k_{RE}, & \text{if retained earnings are used.} \end{cases}$$

Capital Structure Irrelevance and Tax Shields

The central references for this part are Copeland, Weston, and Shastri's *Financial Theory & Corporate Policy*, Gitman and Zutter's *Principles of Managerial Finance*, and Modigliani and Miller's 1958 paper on the cost of capital, corporation finance, and the theory of investment.

The question of whether an *optimal* capital structure exists is one of the most important questions in corporate finance. At present, there is no unique methodology used by financial managers to determine the optimal capital structure of a particular firm. Scholarly research suggests that there is a range of optimal capital structures. Although it does not specify the optimal mix, it helps explain how a firm's chosen financing mix affects its value.

There are two approaches to explaining optimal capital structure: a theoretical approach and a practical approach. We begin with the former.

The theoretical approach starts by identifying three key frictions: the increased probability of bankruptcy due to debt obligations, the agency costs associated with lender monitoring of the firm's actions, and costs associated with managers having more information about the firm's prospects than investors.

We begin with the seminal paper of Modigliani and Miller (1958), which develops the capital-structure irrelevance result in a frictionless setting.

We first study this effect in a world with only corporate taxes, and then we make the model more realistic by adding personal taxes.

Modigliani and Miller (1958) and their 1963 correction assumed the following:

A1. Capital markets are frictionless.

A2. Individuals can borrow and lend at the risk-free rate.

A3. There are no costs to bankruptcy or to business disruption.

A4. Firms issue only two types of claims: risk-free debt and risky equity.

A5. All firms are assumed to be in the same risk class (same operating risk).

A6. Corporate taxes are the only form of government tax, that is, no wealth taxes on corporations and no personal taxes.

A7. All cash flow streams are perpetuities, that is, no growth.

A8. Corporate insiders and outsiders have the same information, that is, no signaling opportunities.

A9. Managers always maximize shareholders' wealth, that is, no agency costs.

A10. Operating cash flows are completely unaffected by changes in capital structure.

These assumptions define the benchmark in which the Modigliani–Miller conclusion holds. Relaxing bankruptcy, tax, information, or agency assumptions can change the capital-structure conclusion, sometimes substantially.

Assumption A5 deserves some clarification. When we say that all firms have the same risk class, we mean that they have the same *operating risk*. In other words, differences across firms may change the scale of operating cash flows, but not the pattern of risk across states. Formally, this means that the risky future net operating cash flows can differ by at most a scale factor:

$$\widetilde{CF}_i = \lambda \widetilde{CF}_j,$$

where $\widetilde{CF}$ is the risky net operating cash flow before interest and taxes, and $\lambda > 0$ is a constant scale factor.

This implies that the expected future cash flows from the two firms (or projects) are perfectly correlated. This perfect correlation is easiest to see by focusing on returns:

$$\tilde{r}_{i,t} = \frac{\widetilde{CF}_{i,t} - \widetilde{CF}_{i,t-1}}{\widetilde{CF}_{i,t-1}},$$

which is the rate of return on firm i at time t. Since $\widetilde{CF}_{i,t} = \lambda\widetilde{CF}_{j,t}$, we have

$$\begin{aligned}\tilde{r}_{i,t} &= \frac{\lambda\widetilde{CF}_{j,t} - \lambda\widetilde{CF}_{j,t-1}}{\lambda\widetilde{CF}_{j,t-1}} \\ &= \tilde{r}_{j,t}.\end{aligned}$$

This shows that if two cash flow streams differ only by a positive scale factor, then they have the same percentage returns in every state. Consequently they have the same return distribution, the same risk, and the same required expected return.

Theorem 16.1 (Modigliani–Miller Proposition 1 without taxes) In a frictionless market without taxes, the market value of any firm is independent of its capital structure and is obtained by capitalizing its expected operating income at a rate ρ appropriate to its risk class. In other words, the method of financing is irrelevant.

Proof. It suffices to show that the market value of a levered firm is the same as that of an unlevered firm when the corporate tax rate is zero:

$$V_L = V_U \quad \text{if } \tau_c = 0,$$

where V_L is the value of a levered firm, V_U is the value of an unlevered firm, and τ_c is the corporate tax rate. The goal is to show that, absent corporate taxes, financing with debt rather than equity does not change the total value created by the firm's assets.

Before we proceed, it is useful to note three facts. First, a pro forma income statement for a typical firm looks as follows:

	REV	Revenues
$-$	VC	Variable costs of operations
$-$	FCC	Fixed cash costs (for example, administrative costs and real estate taxes)
$-$	DEP	Noncash charges (for example, depreciation and deferred taxes)
	$EBIT$	Earnings before interest and taxes
$-$	J	Annual interest on debt.
	EBT	Earnings before taxes
$-$	T	Taxes $= \tau_c(EBT)$, where τ_c is the corporate tax rate.
	NI	Net income

Second, total fixed costs are partitioned into two parts: FCC, the cash component, and DEP, the noncash component. Third, it is important to distinguish cash flows from accounting profit. For an **unlevered firm**, meaning $D = 0$, after-tax cash flow from operations equals after-tax operating income plus depreciation and any other noncash expenses. Depreciation reduces accounting profit, but it does not represent an actual cash outflow in the current period:

$$\begin{aligned}
\text{After-tax cash flow} &= \text{after-tax operating income} + DEP \\
&= \underbrace{(REV - VC - FCC - DEP)}_{=:EBIT} - T + DEP \\
&= EBIT - \tau_c EBIT + DEP \\
&= EBIT(1-\tau_c) + DEP.
\end{aligned}$$

Equivalently,

$$\begin{aligned}
\text{after-tax cash flow for an unlevered firm} &= \widetilde{EBIT}(1-\tau_c) + DEP \\
&= (\widetilde{REV} - \widetilde{VC} - FCC - DEP)(1-\tau_c) + DEP
\end{aligned}$$

Under the assumption of no growth, all cash flows are perpetuities. Depreciation each year must be replaced by investment I in order to keep the same amount of capital in place, so $DEP = I$. We also assume the firm has enough taxable income to use the interest deduction in every period. Therefore the after-tax free cash flow, FCF, available for payment to creditors and shareholders is

$$\begin{aligned}
\widetilde{FCF}_U &= (\widetilde{REV} - \widetilde{VC} - FCC - DEP)(1-\tau_c) + DEP - I \\
&= \underbrace{(\widetilde{REV} - \widetilde{VC} - FCC - DEP)}_{=\widetilde{EBIT}}(1-\tau_c), \quad \text{since } DEP = I.
\end{aligned}$$

For a levered firm, meaning $D \neq 0$, after-tax cash flows must be allocated between debt holders and shareholders. Let J denote the annual interest payment. Shareholders receive $\widetilde{NI} + DEP - I$, while debt holders receive J. Therefore

$$\begin{aligned}
\widetilde{FCF}_L &= \widetilde{NI} + DEP - I + J \\
&= \widetilde{NI} + J, \quad \text{since } DEP = I, \\
&= (\widetilde{REV} - \widetilde{VC} - FCC - DEP - J)(1-\tau_c) + J \\
&= \underbrace{(\widetilde{REV} - \widetilde{VC} - FCC - DEP)}_{=:\widetilde{EBIT}}(1-\tau_c) + J\tau_c.
\end{aligned}$$

The last line makes the tax shield explicit: relative to the unlevered firm, leverage adds the term

$J\tau_c$.

Let ρ denote the discount rate for a firm financed entirely by equity, and let k_b denote the before-tax market required rate of return on a risk-free perpetual bond. Its market value is therefore $B = J/k_b$. $\mathbb{E}[\widetilde{FCF_L}]$ denotes the expected perpetual after-tax free cash flow of a levered firm, and $\mathbb{E}[\widetilde{FCF_U}]$ denotes the corresponding quantity for an unlevered firm.

The value of an unlevered firm is

$$\begin{aligned} V_U &= \frac{\mathbb{E}[\widetilde{FCF_U}]}{1+\rho} + \frac{\mathbb{E}[\widetilde{FCF_U}]}{(1+\rho)^2} + \frac{\mathbb{E}[\widetilde{FCF_U}]}{(1+\rho)^3} + \cdots \\ &= \frac{\mathbb{E}[\widetilde{FCF_U}]}{1+\rho}\left(1 + \frac{1}{1+\rho} + \left(\frac{1}{1+\rho}\right)^2 + \cdots\right) \\ &= \frac{\mathbb{E}[\widetilde{FCF_U}]}{\rho} = \frac{\mathbb{E}[\widetilde{EBIT}(1-\tau_c)]}{\rho}. \end{aligned} \tag{3}$$

The value of a levered firm is

$$\begin{aligned} V_L &= PV(\text{unlevered operating cash flow}) + PV(\text{interest tax shield}) \\ &= \left[\frac{\mathbb{E}[\widetilde{FCF_U}]}{1+\rho} + \frac{\mathbb{E}[\widetilde{FCF_U}]}{(1+\rho)^2} + \cdots\right] + \left[\frac{J\tau_c}{1+k_b} + \frac{J\tau_c}{(1+k_b)^2} + \cdots\right] \\ &= \frac{\mathbb{E}[\widetilde{EBIT}(1-\tau_c)]}{\rho} + \tau_c\left\{\frac{J}{1+k_b}\left(1 + \frac{1}{1+k_b} + \left(\frac{1}{1+k_b}\right)^2 + \cdots\right)\right\} \\ &= \underbrace{\frac{\mathbb{E}[\widetilde{EBIT}(1-\tau_c)]}{\rho}}_{=V_U} + \tau_c \underbrace{\frac{J}{k_b}}_{=B}. \end{aligned}$$

Therefore

$$V_L = V_U + \tau_c B. \tag{4}$$

Equation (4) states that the value of a levered firm, V_L, is equal to the value of an unlevered firm, V_U, plus the present value of the tax shield provided by debt, $\tau_c B$. This extra value created by the interest tax shield on debt is known as the *gain from leverage*.

If $\tau_c = 0$ in (4), the equation becomes

$$V_L = V_U \quad \text{if } \tau_c = 0. \tag{5}$$

Equation (5) states that, in the absence of market imperfections, including corporate taxes, the value of the firm is independent of the type of financing used for its projects. Under the assumption of no growth and with $\tau_c = 0$,

$$V_L = V_U = \frac{\mathbb{E}[\widetilde{EBIT}]}{\rho},$$

where ρ is the WACC. Thus the market value of any firm is independent of its capital structure and is given by the ratio of expected $\widetilde{EBIT}$ to the rate ρ appropriate to its risk class. □

Example 16.2 (Arbitrage argument for Modigliani–Miller Proposition 1)

	Company U	*Company L*
$EBIT$	10 000	10 000
$-K_d \cdot D$	0	1500
NI	10 000	8500
K_s	10%	11%
Equity $= S$	100 000	77 272
Debt $= B$	0	30 000
$V = B + S$	100 000	107 272
$WACC$	10%	9.3%
$B/S \equiv$ Leverage	0%	38.8%

The income statements of two firms, U and L, are given in the table above. Both companies have the same perpetual cash flows from operations, $EBIT$, but Company U has no debt whereas Company L has \$30,000 of debt paying 5% interest. The cost of equity is

$$K_s = \frac{NI}{S},$$

where S is the value of shares. From Gordon's model, the value of common stock is

$$k_s = \frac{D_1}{P_0} + g.$$

If $g = 0$ and we value all equity, then

$$K_s = \frac{D_1 \times \text{number of shares}}{P_0 \times \text{number of shares}} + 0 = \frac{NI}{S}.$$

Since

$$(K_s)_L = 11\% > (K_s)_U = 10\%,$$

holding Company L is riskier. Also,

$$\underbrace{V_L > V_U}_{\text{higher value}} \qquad \text{and} \qquad WACC_L = \frac{EBIT}{V_L} < WACC_U = \frac{EBIT}{V_U},$$

which means that Company L has the lower weighted average cost of capital.

This raises a natural question: is the difference in values a violation of Modigliani–Miller Proposition 1?

Modigliani and Miller argued that although $V_L \neq V_U$ initially, the difference will *not* persist. Their demonstration is one of the earliest arbitrage arguments in finance theory.

The arbitrage strategy is to sell shares in L, borrow, and buy shares in U, under the assumption that we initially own shares in L.

We can create homemade leverage as follows:

1. Sell stock in L. If we own 1% of the equity, then we sell \$772.72.
2. Borrow an amount equal to 1% of the debt in L, namely $1\% \times 30{,}000 = \$300$, at 5% interest.
3. Buy 1% of the shares in U, that is, buy \$1000 worth of shares and earn 10%.

Up to this point, we have the same income as in the levered position in L before arbitrage. However, the cash inflows and outflows are:

$$\begin{aligned} \text{IN: } & \$772.72 \text{ (from selling shares in } L) + \$300.00 \text{ (from borrowing)} \\ \text{OUT: } & -\$1000.00 \text{ (used to buy shares in } U) \\ \text{NET: } & \$72.72 \text{ available to invest in } U \text{ and earn } 10\%. \end{aligned}$$

$$\text{Total income} = \$85.00 + \$7.27 = \$92.27.$$

$$ROE = \frac{\$92.27}{\$772.72} = 11.94\%.$$

Thus, in terms of return,

$$\text{ROE}_{\text{before arbitrage}} = \frac{\$85}{\$772.72} = 11\% \quad \text{and} \quad \text{ROE}_{\text{after arbitrage}} = \frac{\$92.27}{\$772.72} = 11.94\%.$$

In terms of risk,

$$\text{Leverage}_{\text{before}} = \frac{\text{Debt}}{\text{Equity}} = 38.8\% \quad \text{and} \quad \text{Leverage}_{\text{after}} = \frac{300}{772.72} = 38.8\%.$$

After the homemade arbitrage, we achieve a higher return on equity at the same risk. Therefore, anyone holding equity in L has an incentive to do the same thing until the market values of the two firms become identical.

Therefore, in a perfect world without taxes, the market values of levered and unlevered firms are identical.

Personal Taxes and the Gain from Leverage

Miller (1977) modified the result in (5) by introducing personal taxes as well as corporate taxes into the model. This revision made both the model and the implied effect of leverage more realistic; after all, we do not observe firms with 100% debt in their capital structures.

Let τ_{PS} denote the personal tax rate on income received from holding shares, and let τ_{PB} denote the personal tax rate on income from bonds. These are the two personal tax rates in the model. Let J be the perpetual interest payment and let r_B be the discount rate appropriate to the bondholder's after-personal-tax cash flow.

The value of a firm financed entirely by equity, that is, the value of the unlevered firm, is

$$V_U = \frac{\mathbb{E}[\widetilde{EBIT}(1-\tau_c)(1-\tau_{PS})]}{\rho}.$$

On the other hand, if the firm has both bonds and shares outstanding, the earnings stream is partitioned into two parts. The first part is cash flow to shareholders after corporate and personal taxes:

$$(EBIT - J)(1 - \tau_c)(1 - \tau_{PS}).$$

The second part is payment to bondholders after personal taxes:

$$J(1 - \tau_{PB}).$$

The total cash payments to suppliers of capital are therefore

$$\underbrace{\mathbb{E}[\widetilde{EBIT}(1 - \tau_c)(1 - \tau_{PS})]}_{\text{cash flows to owners of the unlevered firm}} \quad -J(1 - \tau_c)(1 - \tau_{PS}) + J(1 - \tau_{PB}).$$

The first term is the same as the cash flows to owners of the unlevered firm and should therefore be discounted at the cost of equity for a firm financed entirely by equity, ρ. The remaining after-tax debt-related cash flow is discounted at r_B.

$$V_L = \frac{\mathbb{E}[\widetilde{EBIT}(1 - \tau_c)(1 - \tau_{PS})]}{\rho} + \frac{J\left[(1 - \tau_{PB}) - (1 - \tau_c)(1 - \tau_{PS})\right]}{r_B}. \tag{6}$$

Equivalently,

$$V_L = V_U + \underbrace{\left[1 - \frac{(1 - \tau_c)(1 - \tau_{PS})}{1 - \tau_{PB}}\right] B}_{\text{gain from leverage}}. \tag{7}$$

where

$$B = \frac{J(1 - \tau_{PB})}{r_B}$$

is the market value of debt.

With the introduction of personal taxes, the gain from leverage is the second term on the right-hand side of (7). When personal tax rates are set equal to zero, $r_B = k_b$ under the same debt assumptions, and equation (6) reduces to (4). Thus the gain from leverage in (6) is the same as the one derived in (4).

The bracketed term in (7) is the net tax advantage of debt. It is positive when the corporate tax shield dominates the disadvantage that personal taxes impose on bond income, zero when the two effects exactly offset, and negative when personal taxes make debt financing unattractive to investors.

The tax advantage of debt is exactly eliminated when

$$(1-\tau_{PB}) = (1-\tau_c)(1-\tau_{PS})$$

When bond income is taxed more heavily than stock income, the personal-tax disadvantage of debt offsets some or all of the corporate interest tax shield; the displayed equality is the exact zero-advantage condition in this model.

Optimal Capital Structure and WACC Tradeoffs

It is generally believed that the value of the firm is maximized when the cost of capital is minimized.

In the simple no-growth valuation heuristic, suppose the firm's expected after-tax operating cash flow is fixed while its financing mix varies. Then

$$V = \frac{\mathbb{E}[FCF]}{\text{WACC}},$$

so minimizing WACC is equivalent to maximizing firm value within that model. This is a tradeoff-theory statement, not a claim that the unlevered discount rate in (3) changes with the levered firm's financing mix.

This is the basic tradeoff view of capital structure. At low levels of debt, replacing some equity with cheaper debt that offers tax advantages tends to reduce WACC. At higher levels of debt, the added financial risk raises both the cost of debt and the cost of equity, and eventually those increases outweigh the tax benefit.

Figure 21 illustrates the weighted average cost of capital together with the component cost functions for equity, k_s, and debt, k_d, as functions of financial leverage measured by the debt ratio.

k_d remains low because of the tax shield, but it increases slowly as leverage increases in order to compensate lenders for the additional risk.

k_s is above the cost of debt. It increases as financial leverage increases, and it generally rises more rapidly than the cost of debt because stockholders require a higher return to compensate for the higher degree of financial risk.

The WACC is a weighted average of the firm's debt and equity capital costs. At a debt ratio of

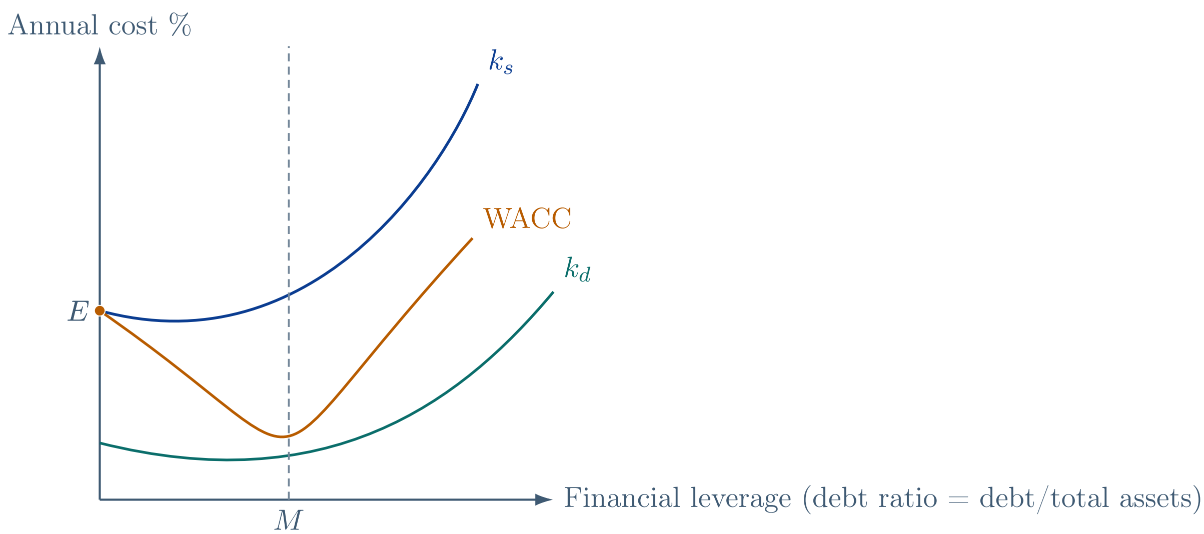


FIGURE 21

zero, the firm is financed entirely by equity, as shown at point E in Figure 21. As the debt ratio increases, the WACC initially declines because the cost of debt is less than the cost of equity, that is, $k_d < k_s$. As the debt ratio continues to increase, the increased debt and equity costs eventually cause the WACC to rise after point M. This behavior produces a U-shaped WACC function.

Because firm value is maximized when the overall cost of capital is minimized, the optimal capital structure in this tradeoff model occurs at point M in Figure 21.

Lecture 17 — Thursday, July 03, 2014

For cost-of-capital analysis, the sources of funds are long-term debt (issued through bonds), preferred stock, common stock, and retained earnings. Preferred stock, common stock, and retained earnings are all equity financing.

For long-term debt, let k_d denote the before-tax yield to maturity, let I denote annual interest payments with $I = \text{coupon rate} \cdot \text{face value}$, let n denote the number of years to maturity, let N_b denote the net proceeds from selling the bond after flotation and administrative costs, let t denote the corporate tax rate, and let $k_i = k_d(1-t)$ denote the after-tax cost of debt. The shortcut formula given earlier is an approximation to k_d; the exact yield equates N_b to the present value of the coupons and principal.

Using CAPM, the cost of common stock is

$$k_s = r_f + b_{i,m}(\mathbb{E}[\tilde{r}_m] - r_f).$$

If D_p is the preferred stock dividend and N_p is net proceeds, then the cost of preferred stock is

$$k_p = \frac{D_p}{N_p}.$$

Under the Gordon model, the cost of a new common stock issue is

$$k_n = \frac{D_1}{N_n} + g.$$

This formula requires $k_n > g$.

The **weighted marginal cost of capital** (**WMCC**) changes at break points, which are the levels of total new financing at which the cost of one of the components rises. The break point for source j is

$$BP_j = \frac{AF_j}{w_j},$$

where AF_j is the amount of funds available from source j and w_j is its weight in the capital structure.

Example 17.1 Suppose the optimal capital weights are $w_d = 40\%$ for long-term debt, $w_p = 10\%$ for preferred stock, and $w_s = 50\%$ for common stock. We want the firm's break points and weighted marginal cost of capital.

The break points are

$$BP_{\text{common stock}} = \frac{300{,}000}{0.5} = 600{,}000$$

and

$$BP_{\text{long-term debt}} = \frac{400{,}000}{0.40} = 1{,}000{,}000.$$

Range	*Source*	*Weight*	*Cost*	*weight · cost*
\$0–\$600,000	Debt, preferred stock, common stock	$0.4, 0.1, 0.5$	$5.6, 10.6, 13$	$\sum = 9.8$
\$600,000–\$1,000,000	Debt, preferred stock, common stock	$0.4, 0.1, 0.5$	$5.6, 10.6, 14$	$\sum = 10.3$
\$1,000,000–∞	Debt, preferred stock, common stock	$0.4, 0.1, 0.5$	$8.4, 10.6, 14$	$\sum = 11.42$

The corresponding stepwise WMCC schedule is shown in Figure 22.

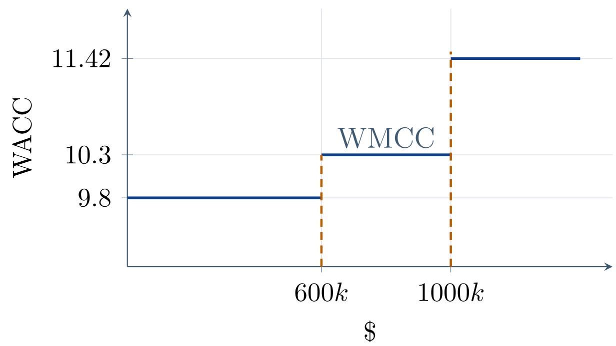


FIGURE 22

Lecture 18 — Tuesday, July 08, 2014

When only corporate taxes are considered, the value of the levered firm is

$$V_L = V_U + \tau_c B.$$

The accounting identities

$$\text{REV} - \text{VC} - \text{FCC} - \text{DEP} = \text{EBIT}, \qquad \text{EBIT} - J = \text{EBT}, \qquad \text{and} \qquad \text{EBT} - T = \text{NI}$$

show how debt service enters taxable income and therefore creates a tax advantage.

If personal taxes are also introduced, with τ_{ps} denoting the tax rate on stock income and τ_{pb} denoting the tax rate on personal bond income, then

$$V_U = \frac{\mathbb{E}[EBIT(1-\tau_c)(1-\tau_{ps})]}{\rho}$$

and

$$V_L = V_U + \left[1 - \frac{(1-\tau_c)(1-\tau_{ps})}{(1-\tau_{pb})}\right] B.$$

Thus the value of a levered firm equals the value of an unlevered firm plus the gain from leverage, where

$$B = \frac{J(1-\tau_{pb})}{r_B}$$

is the market value of debt.

Two comparative statics observations follow immediately. If

$$\tau_c = \tau_{pb} = \tau_{ps} = 0,$$

then $V_L = V_U$. If τ_{pb} becomes much larger than τ_{ps}, then the gain from leverage falls, which weakens the benefit of debt financing.

Optimal capital structure occurs when WACC is minimized.

Dividend Policy and Payout Irrelevance

Dividend policy theory asks whether payout policy affects firm value.

Under the **residual theory of dividends**, dividends are whatever remains after the firm has financed all projects with positive value using its target capital structure. The process is to determine optimal capital expenditures, estimate the equity financing needed, and then use retained earnings first whenever

$$K_{RE} < K_{\text{new issue}}.$$

Example 18.1 To isolate the residual calculation, suppose equity finances the entire CapEx $= 500k$ program. Retained earnings are either $RE_1 = 300k$ or $RE_2 = 800k$. If RE_2 is available, then the firm uses $500k$ of retained earnings to finance capital expenditures and pays the $300k$ surplus as dividends. If only RE_1 is available, then the firm uses $300k$ of retained earnings and issues $200k$ of new stock; no dividend is paid.

Under this interpretation, dividends are residual payments rather than an independent source of value. The standard irrelevance argument says investor return is unaffected by dividend policy because firm value depends on earning power and asset risk, as reflected in the appropriate cost of capital. Any dividend effect then works through the information content of dividends about expected earnings, while clientele effects lead shareholders to sort into firms with the payout patterns they prefer.

Lecture 20 — Tuesday, July 15, 2014

6. General Equilibrium and Exchange Economies

Let Φ_t denote the information available at time t. The **efficient market hypothesis** says that the innovation in the return for the next period can be written as

$$\tilde{z}_{i,t+1} = \tilde{r}_{i,t+1} - \mathbb{E}[\tilde{r}_{i,t+1} \mid \Phi_t].$$

If the efficient market hypothesis holds, then this innovation has conditional expectation zero.

The frictionless benchmark supporting this version of the efficient market hypothesis assumes no transaction costs, freely available information, and common agreement about the implications of information at time t for future asset prices. These are convenient sufficient idealizations, not logically necessary conditions for every form of market efficiency.

If the efficient market hypothesis does not hold, then observed investor behaviour may deviate from the Markowitz model. Investors may lie on, above, or below the CML and may fail to lie on the efficient frontier.

Intertemporal Competitive Equilibrium

This section develops a model of macroeconomics built from microeconomic foundations.

The model assumes, as usual, that there are two periods: Period 0 (today) and Period 1 (the future). Preferences are described by $U(\underbrace{c_0}_{\text{consumption}}, \underbrace{l_0}_{\text{leisure}})$ and $U(c_1, l_1)$.

Consumers are homogeneous, so all consumers share the same preferences. The total time h satisfies

$$h = N^s + l,$$

where N^s is the number of hours supplied to work and l is leisure. Income satisfies

$$I_t = w \cdot N^s = w(h - l),$$

so

$$I_0 = w(h - l_0) \qquad \text{and} \qquad I_1 = w(h - l_1).$$

The control variables are N_0^s, l_0, c_0, N_1^s, l_1, and c_1. We fix N^s and l to reduce the model to a problem in the two variables c_0 and c_1. Here t_0 and t_1 denote income taxes.

With labour and leisure held fixed, the central choice is how to allocate purchasing power across the two dates. The interest rate is therefore the relative price of future consumption in terms of current consumption: saving one unit today produces $1 + r$ units tomorrow, while borrowing one unit today requires repayment of $1 + r$ units tomorrow.

The consumer also faces a budget. Savings are $s_0 = y_0 - t_0 - c_0$ and $s_1 = 0$, since there is no period 2 in this setup. The Period 0 budget constraint is $c_0 + s_0 = y_0 - t_0$, while the Period 1 budget constraint is $c_1 + s_1 = (y_1 - t_1) + s_0(1 + r)$.

Lecture 21 — Thursday, July 17, 2014

Recall that in this model built from microeconomic foundations,

$$c_0 + s_0 = y_0 - t_0 \tag{8}$$
$$c_1 + s_1 = (y_1 - t_1) + s_0(1 + r) \tag{9}$$

We now derive the intertemporal budget constraint. From (8) we have $s_0 = (y_0 - t_0) - c_0$. Substituting this into (9) gives

$$c_1 = (y_1 - t_1) + [(y_0 - t_0) - c_0](1 + r).$$

Rearranging yields

$$c_1 + c_0(1 + r) = (y_1 - t_1) + (y_0 - t_0)(1 + r),$$

and dividing through by $1 + r$ gives the form in present values

$$c_0 + \frac{c_1}{1 + r} = (y_0 - t_0) + \frac{y_1 - t_1}{1 + r}.$$

The left-hand side is the present value of lifetime consumption, while the right-hand side is the present value of lifetime disposable income. We denote that wealth measure by W.

The consumer chooses c_0 and c_1 to maximize $U(c_0, c_1)$ subject to the intertemporal budget constraint. Using Lagrange multipliers, let

$$\mathcal{L}(c_0, c_1, \lambda) = U(c_0, c_1) + \lambda \left[W - c_0 - \frac{c_1}{1+r} \right].$$

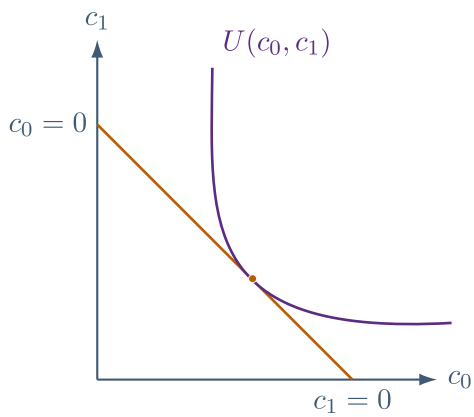


FIGURE 23

In Figure 23, the budget line has slope $-(1 + r)$, and at the optimum that slope must equal the slope of the indifference curve. Therefore the consumer's choice is characterized by two conditions: the tangency condition

$$1 + r = \frac{\mu_0}{\mu_1},$$

and the budget constraint

$$c_0 + \frac{c_1}{1+r} = W.$$

Equivalently, the first-order conditions are

$$\frac{\partial \mathcal{L}}{\partial \lambda} = 0, \qquad \frac{\partial \mathcal{L}}{\partial c_0} = 0, \qquad \text{and} \qquad \frac{\partial \mathcal{L}}{\partial c_1} = 0.$$

The first condition reproduces the budget constraint, while the other two imply that the marginal rate of substitution equals $1 + r$. In words, the consumer chooses consumption so that the utility cost of giving up one unit today exactly matches the utility gain from moving that unit into the future at the market interest rate.

The production and labour decisions are suppressed in this reduced exchange model. If a labour market is included, clearing requires $N^d = N^s$; if a goods market is included, aggregate demand must equal aggregate supply. We treat the equilibrium wage as exogenous in what follows.

Government spending in Period 0 equals G_0, and total taxes collected are $T_0 = t_0^1 + t_0^2 + \cdots + t_0^n$. If current government expenditure exceeds current tax revenue, the government runs a deficit and finances it domestically by issuing bonds B_0 to consumers. Then the government's Period-0 and Period-1 constraints are

$$G_0 = T_0 + B_0 \qquad \text{and} \qquad G_1 = T_1 - (1+r)B_0.$$

Eliminating B_0 gives

$$G_0 + \frac{G_1}{1+r} = T_0 + \frac{T_1}{1+r}.$$

Thus the present value of government expenditure must equal the present value of government revenue. This is simply the government's intertemporal budget constraint.

Equilibrium is imposed in the goods market and credit market, while the labour side is held fixed in this reduced model. The consumption-good price is normalized, and credit-market clearing determines the equilibrium interest rate r. For the goods market, we require equilibrium in both periods:

$$C_0 + G_0 = Y_0 \qquad \text{and} \qquad C_1 + G_1 = Y_1,$$

where C_0 = aggregate consumption $= c_0^1 + c_0^2 + \cdots + c_0^n$, with the analogous definition in Period 1, and Y_0 = aggregate income $= y_0^1 + \cdots + y_0^n$. For the credit market,

$$S^p + S^G = \text{private saving} + \text{government saving} = 0,$$

with $S^p = Y_0 - C_0 - T_0$ and $S^G = T_0 - G_0$.

At this stage the normalization is $P^* = 1$, the equilibrium wage w^* is treated as fixed, and the remaining price to determine is the equilibrium interest rate r^*.

The outcome of the model is a **competitive equilibrium** (**CE**). A competitive equilibrium is a vector of prices and decision rules subject to the following conditions:

1. The consumer chooses c_0 and c_1, given h, w, t_0, t_1, y_0, and y_1, so as to maximize $U(c_0, c_1)$ subject to the intertemporal budget constraint

$$c_0 + \frac{c_1}{1+r} = (y_0 - t_0) + \frac{y_1 - t_1}{1+r}.$$

2. The government satisfies its budget constraint

$$G_0 + \frac{G_1}{1+r} = T_0 + \frac{T_1}{1+r}.$$

3. All markets clear: in the goods market, $AD_0 = AS_0$ and $AD_1 = AS_1$; in the credit market,

$S_0^p + S_0^G = 0.$

Lecture 22 — Tuesday, July 22, 2014

We work with an economy that has two periods, homogeneous agents, and no risk. The competitive equilibrium is

$$CE = \left[\begin{bmatrix} P_0^* \\ P_1^* \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},\ r^*,\ w^* \text{ fixed} \right].$$

Thus $P_0 = P_1 = 1$ is the price normalization.

Example 22.1 (Exchange economy with two consumers) There are two periods and two identical consumers. Preferences are $U^i = \log(c_0^i) + 0.8 \log(c_1^i)$ for $i = 1, 2$. Endowments are

$$y_0^1 = 12.5, \qquad y_0^2 = 12.5, \qquad y_1^1 = 16, \qquad \text{and} \qquad y_1^2 = 16.$$

Government values are

$$G_0 = 5, \qquad G_1 = 12, \qquad T_0 = 5, \qquad \text{and} \qquad T_1 = 12.$$

The exogenous variables are y_0, y_1, t_0, t_1, and w, and the endogenous choices are c_0^1, c_1^1, c_0^2, and c_1^2.

Consumer 1 solves

$$\max_{c_0^1, c_1^1} \log(c_0^1) + 0.8 \log(c_1^1)$$

subject to

$$W = c_0^1 + \frac{c_1^1}{1+r}.$$

Using the tangency condition,

$$\frac{\partial U / \partial c_0}{\partial U / \partial c_1} = 1 + r \iff \frac{1/c_0^1}{0.8/c_1^1} = 1 + r \iff c_1^1 = 0.8(1+r)c_0^1.$$

Substituting this into the budget constraint gives

$$c_0^{1*} = \frac{W}{1.8} \qquad \text{and} \qquad c_1^{1*} = \frac{0.8}{1.8}W(1+r).$$

Because the two consumers are identical, Consumer 2 chooses the same optimal consumption plan:

$$c_0^{2*} = c_0^{1*} \qquad \text{and} \qquad c_1^{2*} = c_1^{1*}.$$

To determine equilibrium, use the credit market condition $S_0^p + S_0^G = 0$, which here becomes

$$(y_0^1 - t_0^1 - c_0^1) + (y_0^2 - t_0^2 - c_0^2) + (T_0 - G_0) = 0.$$

Since taxes are split evenly across the two consumers, $T_0 = 5$ implies $t_0 = 2.5$, and similarly $t_1 = 6$. Each consumer's lifetime wealth is therefore

$$W = 10 + \frac{10}{1+r}.$$

Substituting the optimal consumption rules into the market clearing condition gives

$$\left(12.5 - 2.5 - \frac{W}{1.8}\right) + \left(12.5 - 2.5 - \frac{W}{1.8}\right) + 0 = 0,$$

so $W = 18$. Solving $10 + 10/(1+r) = 18$ gives $r^* = 0.25 = 25\%$, the equilibrium interest rate.

State-Contingent Claims and Arrow–Debreu Prices

We now introduce uncertainty in Period 1 within an **Arrow–Debreu economy**.

A **state-contingent claim** is a security whose payoff depends on the realized state of nature. Assume there are n states in Period 1. Then any security has n possible future payoffs, one for each state. A **primitive Arrow–Debreu security** pays only \$1 in one particular future State Θ and pays \$0 in every other state. These securities are useful because they isolate pure state-contingent consumption.

To construct Arrow–Debreu securities from existing traded securities, the traded payoff vectors must span the state space. Thus, with n states, the payoff matrix must have rank n; there may be exactly n linearly independent securities or more than n securities with redundant payoffs. If this rank condition holds, the market is **complete**, meaning that any state-contingent payoff can be replicated.

It is useful to organize payoffs in a matrix whose rows are states and whose columns are traded securities. Completeness means that this matrix has full row rank. In the square case it is invertible: if A is the payoff matrix, the portfolio replicating primitive claim e_j solves $A\boldsymbol{x}_j = e_j$, so $\boldsymbol{x}_j = A^{-1}e_j$. The columns of A^{-1} therefore give the replicating portfolios.

Example 22.2 (Replicating primitive Arrow–Debreu securities) Suppose there are three states, Θ_1, Θ_2, and Θ_3, and three stocks with payoffs

$$a = (1, 0, 0), \qquad b = (0, 1, 1), \qquad \text{and} \qquad c = (0, 1, 4).$$

The market is complete because these payoff vectors are linearly independent. Put them into the payoff matrix

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 4 \end{bmatrix}.$$

To replicate the primitive securities, solve $AX = I$, so $X = A^{-1}$. This gives

$$X = A^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{4}{3} & -\frac{1}{3} \\ 0 & -\frac{1}{3} & \frac{1}{3} \end{bmatrix}.$$

Each column of X tells us how to combine the traded securities to obtain one unit of consumption in exactly one state. For example, the second column shows that the security paying \$1 only in State Θ_2 is obtained by going long 4/3 units of b and short 1/3 units of c.

The **Arrow–Debreu price** P_{Θ_i} is the price today of one unit of consumption delivered only if State Θ_i occurs tomorrow. Thus P_2 is the price of the claim with payoff vector $(0, 1, 0)$. In general, these prices measure how expensive consumption is in each future state.

Example 22.3 (Solving for Arrow–Debreu prices) Consider a complete market with two states and two securities. Security a has payoff (\$10, \$30), Security b has payoff (\$40, \$20), and their current prices are $P_a = \$10$ and $P_b = \$20$. Let P_1 and P_2 be the Arrow–Debreu prices of the two primitive securities with payoffs $(1, 0)$ and $(0, 1)$. Since every traded security can be valued as the sum of its state payoffs times the corresponding state prices,

$$P_a = P_1 Q_{1a} + P_2 Q_{2a}, \qquad\qquad 10 = 10P_1 + 30P_2,$$

$$P_b = P_1 Q_{1b} + P_2 Q_{2b}, \qquad 20 = 40P_1 + 20P_2.$$

Solving gives $P_1 = \$0.4$ and $P_2 = \$0.2$. Therefore a claim that pays one unit of consumption only in State Θ_1 costs \$0.4 today, while the analogous claim for State Θ_2 costs \$0.2.

In the example with two consumers, Consumer 1 has the following values in each period:

Period 0	Period 1
$y_0^1 = 12.5$	$y_1^1 = 16$
$t_0^1 = 2.5$	$t_1^1 = 6$
$c_0^1 = \dfrac{W}{1.8} = \dfrac{10 + (10/1.25)}{1.8} = 10$	$c_1^1 = 10.$

If $S_0 > 0$, the consumer is a lender; if $S_0 < 0$, the consumer is a borrower; and if $S_0 = 0$, the consumer is neither.

$$S_0^1 = y_0^1 - t_0^1 - c_0^1 = 12.5 - 2.5 - 10 = 0.$$

Therefore Consumer 1 is neither lending nor borrowing. In this economy with two people, Consumer 2 is also neither lending nor borrowing. The lending and borrowing regions around the endowment point are shown in Figure 24.

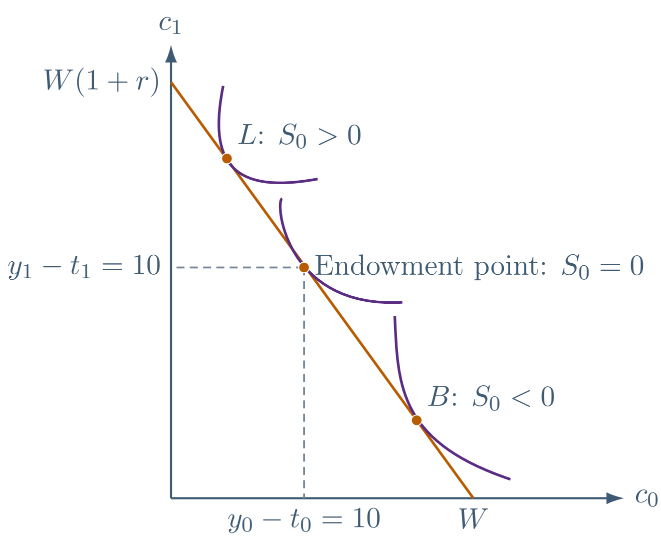


FIGURE 24

There are three states: $\Theta_1, \Theta_2, \Theta_3$. A security with payoff vector $(0, 0, 1)$ pays only in state Θ_3, so it is the primitive security associated with that state.

Lecture 23 — Thursday, July 24, 2014

Arrow–Debreu Exchange with Uncertainty

The economy is closed and consists only of a household sector with two consumers who live for two periods; there is no government or business sector. There are two states in Period 1: Θ_1 (good) and Θ_2 (bad). The probabilities are

$$\pi_1 = \mathbb{P}(\Theta_1) = \frac{1}{3} \qquad \text{and} \qquad \pi_2 = \mathbb{P}(\Theta_2) = \frac{2}{3}.$$

Endowments are

$$y_0^1 = 10, \qquad y_0^2 = 5, \qquad y_1^1(\Theta_1) = 1, \qquad y_1^1(\Theta_2) = 2, \qquad y_1^2(\Theta_1) = 4, \qquad \text{and} \qquad y_1^2(\Theta_2) = 6.$$

Preferences are homogeneous and take the form

$$U^i = \frac{1}{2} c_0^i + \delta^i \sum_{j=1}^{2} \pi_j u_1^i(c_1^i(\Theta_j)), \quad i = 1, 2,$$

where δ^i is consumer i's discount factor. The first term captures utility from current consumption, while the second term is expected discounted utility from state-contingent future consumption. In this example,

$$u_1^i(c_1^i(\Theta_j)) = \log\bigl(c_1^i(\Theta_j)\bigr) \qquad \text{and} \qquad \delta^1 = \delta^2 = 0.9.$$

A **competitive equilibrium** is a set of Arrow–Debreu prices and consumer decision rules such that both consumers maximize utility subject to their intertemporal budget constraints, given π_1, π_2, δ, the endowments, and the state prices P_1 and P_2, and such that all markets clear. Goods market clearing requires $AD_0 = AS_0$, so $c_0^1 + c_0^2 = Y_0$, and also $AD_1 = AS_1$ in each future state, so aggregate Period 1 consumption equals aggregate Period 1 endowment state by state.

The state prices P_1 and P_2 are the prices, in units of Period 0 consumption, of one unit of Period 1 consumption delivered in the corresponding state. A high state price means that consumption in that state is expensive today, either because the state is likely, because consumption in that state has high marginal utility, or both.

We now find the optimal consumption allocations for both consumers as functions of P_1 and P_2.

For Consumer 1, write $c_1^1(\Theta_1) \equiv c_1^1$ and $c_1^1(\Theta_2) \equiv c_2^1$. The problem is

$$\max_{c_0^1, c_1^1, c_2^1} \left\{ \frac{1}{2} c_0^1 + 0.9 \left(\frac{1}{3} \log\big(c_1^1\big) + \frac{2}{3} \log\big(c_2^1\big) \right) \right\}$$

subject to

$$y_0^1 + P_1 y_1^1 + P_2 y_2^1 \geq c_0^1 + P_1 c_1^1 + P_2 c_2^1.$$

Using the endowments of Consumer 1, the Lagrangian is

$$\mathcal{L}(c_0^1, c_1^1, c_2^1, \lambda) = \frac{1}{2} c_0^1 + 0.9 \left(\frac{1}{3} \log\big(c_1^1\big) + \frac{2}{3} \log\big(c_2^1\big) \right) + \lambda(10 + P_1 + 2P_2 - c_0^1 - P_1 c_1^1 - P_2 c_2^1).$$

The first-order conditions give

$$\frac{\partial \mathcal{L}}{\partial c_0^1} = 0 \Longleftrightarrow \lambda = \frac{1}{2},$$

$$\frac{\partial \mathcal{L}}{\partial c_1^1} = 0 \Longleftrightarrow c_1^1 = \frac{0.6}{P_1} \qquad \text{and} \qquad \frac{\partial \mathcal{L}}{\partial c_2^1} = 0 \Longleftrightarrow c_2^1 = \frac{1.2}{P_2}.$$

These conditions have the general form

$$P_\theta = \frac{\delta^i \pi_\theta M U_\theta^i}{M U_0^i}.$$

Thus an Arrow–Debreu price is a discounted probability weighted by the consumer's marginal rate of substitution between state-θ future consumption and current consumption. Finally,

$$\begin{aligned} c_0^1 &= 10 + P_1 + 2P_2 - P_1 c_1^1 - P_2 c_2^1 \\ &= 10 + P_1 + 2P_2 - P_1 \left(\frac{0.6}{P_1} \right) - P_2 \left(\frac{1.2}{P_2} \right) \\ &= 8.2 + P_1 + 2P_2. \end{aligned}$$

For Consumer 2, the same calculation gives

$$c_1^2 = \frac{0.6}{P_1}, \qquad c_2^2 = \frac{1.2}{P_2}, \qquad \text{and} \qquad c_0^2 = 3.2 + 4P_1 + 6P_2.$$

It remains to solve for the Arrow–Debreu price vector. Market clearing in State Θ_1 requires

$$c_1^1 + c_1^2 = y_1^1 + y_1^2,$$

so

$$\frac{0.6}{P_1} + \frac{0.6}{P_1} = 1 + 4 \qquad \text{and} \qquad P_1^* = 0.24.$$

Similarly, market clearing in State Θ_2 requires

$$c_2^1 + c_2^2 = y_2^1 + y_2^2,$$

so

$$\frac{1.2}{P_2} + \frac{1.2}{P_2} = 2 + 6 \qquad \text{and} \qquad P_2^* = 0.3.$$

At these prices,

$$c_0^1 = 9.04 \qquad \text{and} \qquad c_0^2 = 5.96,$$

and $c_0^1 + c_0^2 = 15 = y_0^1 + y_0^2$, so the Period 0 goods market also clears.

Appendix: Lecture and Tutorial Index