ECON 3800
Lecture Notes on Financial Economics

Instructors: Yongpei Cai and Kuan Xu
Prepared by Professor Kuan Xu

2023

Contents

If you have any feedback about these notes, please send your comments and suggestions to kuan.xu@dal.ca. Thank you!

Chapter 1

Consumption, Saving and Investment

1.1 Learning Objectives

- Utility functions
- Intertemporal savings behavior and portfolio optimization
- Finite future states
- Expected utility and risk aversion
- The determinants of intertemporal behavior
- Three consumption/saving theories
- Investments

1.2 Utility Functions

Are we sure that the individual person indeed has a utility function? If yes, what are the basic characteristics of the utility function?

Winterfeldt and Edwards (1986) show, in their book entitled Decision Analysis and Behavior Research, that it is possible to figure out the utility function if we explore the mind of the individual.

Mary won a prize of \$300 and she is asked by the investigator to play a lottery of the following kind using her \$300:

Win (\$)	Probability
1,000	.20
500	.30
300	.30
0	.20

Table 1.1: Lottery Offered to Mary

To figure out Mary's utility function, the investigator asks Mary to compare a sure thing in dollars ($C = \$100, 300, 500$) with a gamble based on utility with some unknown probabilities. To fix the gamble, we allow the gamble to have two likely outcomes \$ 1000 and \$ 0 and then assign the highest and lowest levels of utility to them:

$$u(\$1000) = 100 \tag{1.1}$$

and

$$u(\$0) = 0. \tag{1.2}$$

(This should not matter as the level of utility is ordinal.) Now the gamble has the probability p of offering utility $u(\$1000) = 100$ and the probability $1 - p$ of offering $u(\$0) = 0$. Please note that $p + (1 - p) = 1$.

The investigator then tries to find out from Mary what probability of winning \$ 1000 in the gamble would make her feel indifferent between the utility of a sure thing C (in our example, $u(C)$ will be $u(\$500), u(\$300), u(\$100)$, respectively) and the expected utility of the gamble?

More specifically, for $u(\$500)$, the utility of a high value of the sure thing, Mary would be indifferent between this sure thing and the gamble that has the 70% chance of winning \$ 1000. This implies

$$u(\$500) = (.70)u(\$1000) + (.30)u(\$0) = 70. \tag{1.3}$$

For $u(\$300)$, the utility of a lower value of the sure thing, Mary would be indifferent between this sure thing and the gamble that has the 50% chance of winning \$ 1000. This implies

$$u(\$300) = (.50)u(\$1000) + (.50)u(\$0) = 50. \tag{1.4}$$

For $u(\$100)$, the utility of an even lower value of the sure thing, Mary would be indifferent between this sure thing and the gamble that has the 20% chance of winning \$ 1000. This implies

$$u(\$100) = (.20)u(\$1000) + (.80)u(\$0) = 20. \tag{1.5}$$

Now we can collect the data from Mary in the following table. We can plot the data in the following figure. Both the table and figure show that, indeed, Mary has a concave utility function.

outcome	0	100	300	500	1000
utility	0	20	50	70	100
linear	0	10	30	50	100

Table 1.2: Mary's Utility Function: Concave Compared with Linear Extrapolation

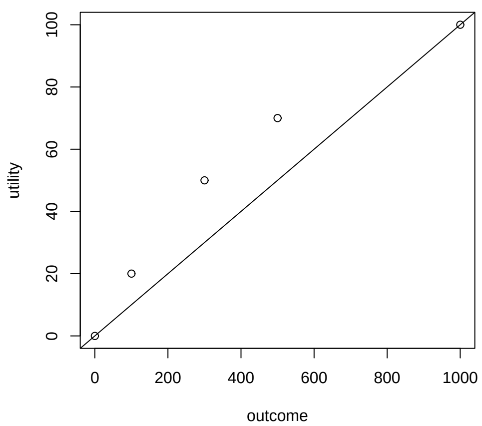


Figure 1.1: Mary's Utility Function: Concave Compared with Linear Extrapolation

Now we can look at the utility function from the theoretical point of view. Let x be the consumption bundle and $u(x)$ the utility function of the consumption bundle, which may take different forms:

- Risk neutral utility function

$$u(x) = x. \tag{1.6}$$

See Figure 1.2.

- Exponential utility function

$$u(x) = 1 - e^{-ax}, \tag{1.7}$$

where $a > 0$.

Let $a = 2$. We show this utility function in Figure 1.2.

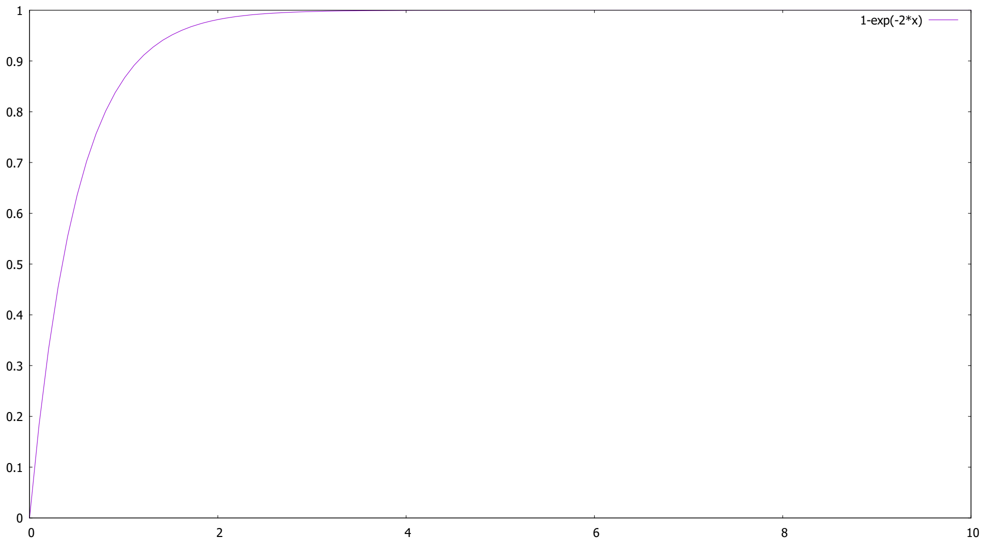


Figure 1.2: Exponential Utility Function: Concave

- Logarithmic utility function

$$u(x) = \ln(x). \tag{1.8}$$

We show this utility function 1.3.

- Power utility function

$$u(x) = bx^b, \tag{1.9}$$

where $b \leq 1$ and $b \neq 0$. When $b = 1$, the utility function is risk neutral.

Let $b = 0.5$. We show this utility function in Figure 1.4.

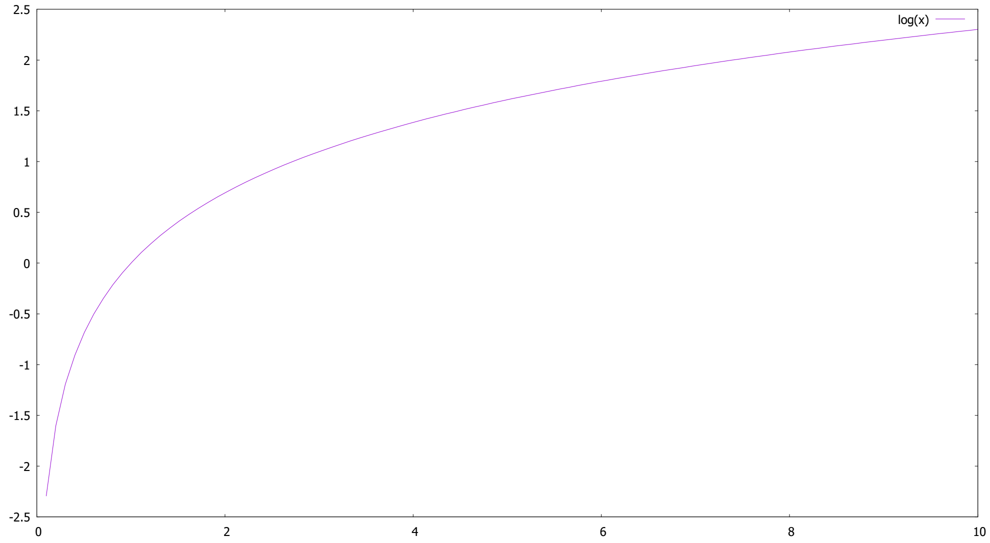


Figure 1.3: Logarithmic Utility Function: Concave

- Quadratic utility function

$$u(x) = ax - \frac{1}{2}bx^2, \tag{1.10}$$

  for some $a > 0$ and $b > 0$. This function is increasing when $x < \frac{a}{b}$.

  Let $a = 10$ and $b = 1$. We show this utility function in Figure 1.5.

Let x and y be two different kinds of consumption bundles and $u(x, y)$ the utility of the two consumption bundles:

- Cobb-Douglas utility function

$$u(x, y) = x^a y^b, \tag{1.11}$$

  where $0 < a \leq 1$ and $0 < b \leq 1$.

  Let $a = 0.5$ and $b = 0.5$. We show this utility function in Figure 1.6.

- CES (constant elasticity of substitution) utility function

$$u(x, y) = (\alpha x^r + (1 - \alpha)y^r)^{1/r}, \tag{1.12}$$

  where α and r are constants.

  Let $a = 0.5$ and $r = 0.5$. We show this utility function in Figure 1.7.

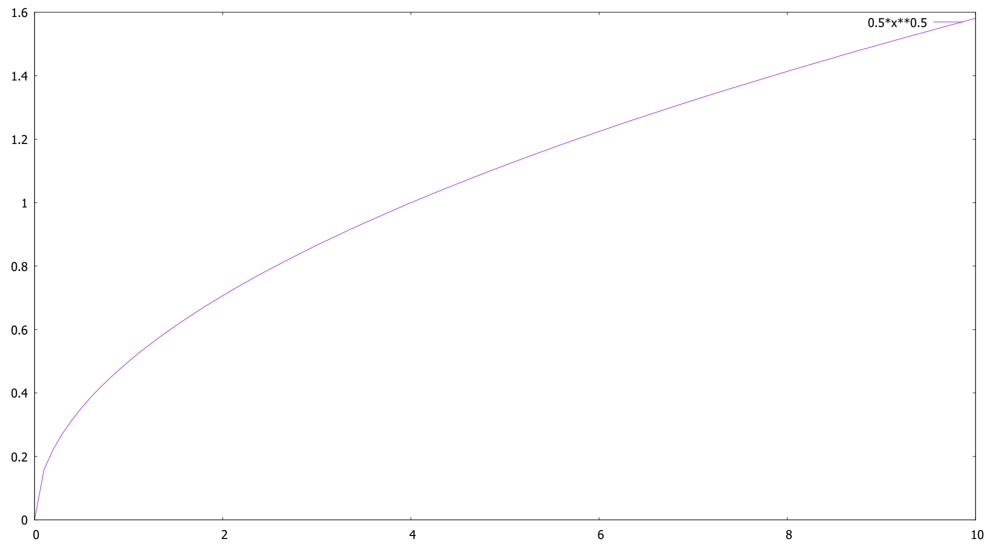


Figure 1.4: Power Utility Function: Concave

Normally, we think of the level of utility as cardinal value. That is, the ranking of utility levels is more important. Therefore, a monotonic affine transformation of $u(x)$ such as

$$v(x) = \alpha u(x) + \beta \tag{1.13}$$

or

$$v(x, y) = \alpha u(x, y) + \beta \tag{1.14}$$

where $\alpha > 0$, will be the equivalent utility function. It should be noted that in Euclidean geometry, an affine transformation is a geometric transformation that preserves lines and parallelism, but not necessarily Euclidean distances and angles.

Example:

It can be shown that $u(x) = 1 - e^{-ax}$ is a monotonic transformation of $v(x) = u(x) - 1 = -e^{-ax}$.

We often use the concave utility function ($u''(x) < 0$) that is increasing ($u' > 0$) to characterize non-satiation and risk aversion.

Example:

For the exponential utility function, $u'(x) = ae^{-ax}$ and $u''(x) = -a^2e^{-ax}$.

Absolute risk aversion coefficient or Arrow-Pratt measure of absolute risk aversion:

$$a(x) = -\frac{u''(x)}{u'(x)} \tag{1.15}$$

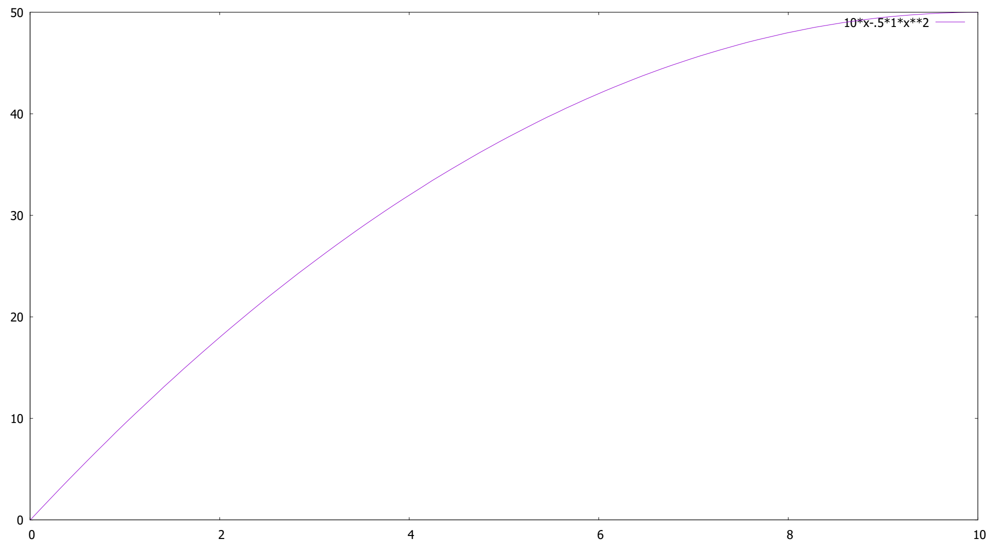


Figure 1.5: Quadratic Utility Function: Concave

Relative risk aversion coefficient or Arrow-Pratt measure of relative risk aversion:

$$r(x) = a(x)x = -\frac{u''(x)x}{u'(x)} \tag{1.16}$$

Example:

$$a(x) = -\frac{-a^2 e^{-ax}}{ae^{-ax}} = a \tag{1.17}$$

and

$$r(x) = a(x)x = -\frac{-a^2 e^{-ax} x}{ae^{-ax}} = ax. \tag{1.18}$$

Example:

Let the new utility function $v(x) = \alpha u(x) + \beta$, where $\alpha > 0$ and $u(x) = 1 - e^{-ax}$. This monotonic affine transformation will not affect the absolute risk aversion coefficient:

$$a(x) = -\frac{u''(x)}{u'(x)} = -\frac{\alpha u''(x)}{\alpha u'(x)} = -\frac{v''(x)}{v'(x)}. \tag{1.19}$$

Behaviour economists believe that people evaluate gains and losses differently and hence the utility function should be kinked. Behaviour economists pay more attention to

1. Framing effects

2. Nonlinear preferences

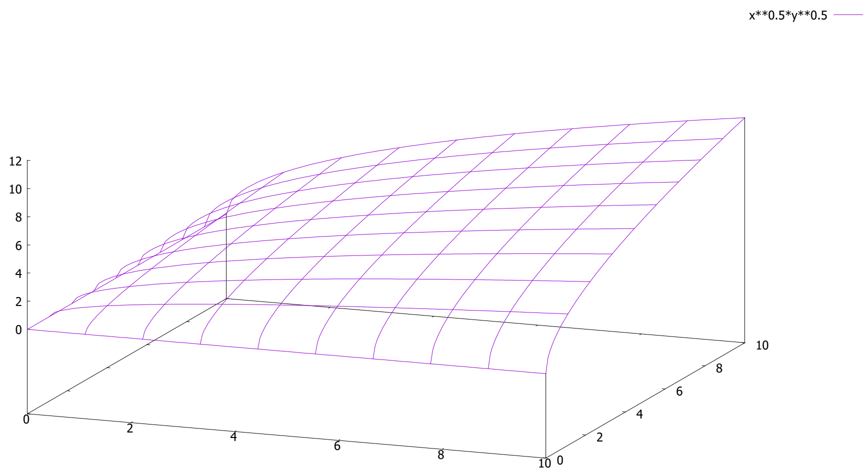


Figure 1.6: Cobb-Douglas Utility Function: Concave

3. Reference dependence
4. Risk seeking
5. Loss aversion

1.3 Intertemporal Savings Behavior

- Saving/borrowing and consumption
- Assumptions:
  (1) an individual earns a fixed income y in each of two periods;
  (2) the individual has no wealth;
  (3) the interest rate does not change.
- Notation:
  y_1 = the income in period 1,
  y_2 = the income in period 2,
  c_1 = the consumption in period 1,
  c_2 = the consumption in period 2,
  $\rho = R - \pi^e$ = the real interest rate, and
  $u(c_1, c_2)$ = the utility derived from consuming c_1 and c_2.

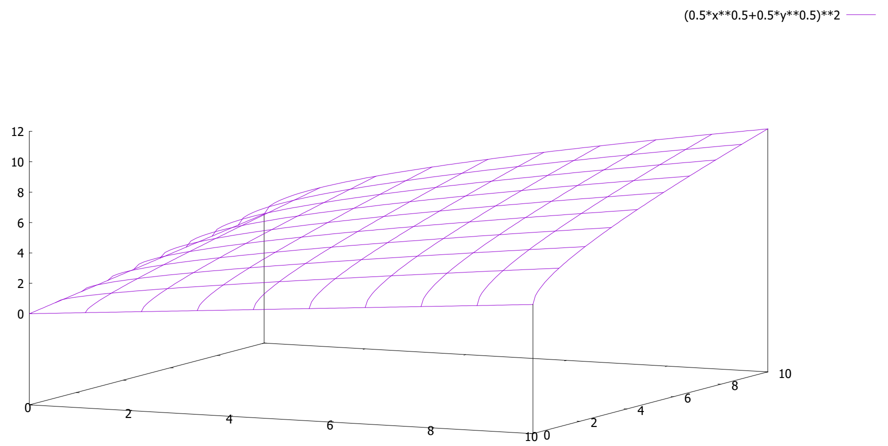


Figure 1.7: CES Utility Function: Concave

- Trade-off between c_1 and c_2: Let the present value of c_1 and c_2, PV_c, is fixed, then

$$PV_c = c_1 + c_2/(1+\rho). \tag{1.20}$$

  From the above equation, we get

$$c_2 = (1+\rho)PV_c - (1+\rho)c_1. \tag{1.21}$$

  The coefficient $-(1+\rho)$ represents the additional future consumption one must sacrifice if one wants to have one more unit of current consumption given that PV_c is fixed.

- What is PV_c ?

$$PV_c = y_1 + y_2/(1+\rho). \tag{1.22}$$

- Budget constraint and its relationship with ρ ($\rho_1 < \rho < \rho_2$)
- Utility function and indifference curves ($u(c_1, c_2) = I_1, I_2, I_3$)
- Time preferences (the slope of indifference curves)
- Optimal consumptions (c_1^*, c_2^*)

$$(1+\rho)(y_1 - c_1^*) + y_2 = c_2^* \tag{1.23}$$

- Two cases:

  (1) if $y_1 > c_1^*$, the individual saves in period 1, and

  (2) if $y_1 < c_1^*$, the individual borrows in period 1.

- An Illustration for Intertemporal Savings Behavior

  Let the utility function be

$$u(c_1, c_2) = c_1^{\alpha} c_2^{1-\alpha} \tag{1.24}$$

and the budget constraint be

$$c_1 + c_2/(1+\rho) = PV_c. \tag{1.25}$$

To find out the optimal choices c_1^* and c_2^*, we can form the Lagrangian function

$$L(c_1, c_2, \lambda) = c_1^{\alpha} c_2^{1-\alpha} + \lambda(PV_c - c_1 - c_2/(1+\rho)). \tag{1.26}$$

This function allows us to maximize $u(c_1, c_2)$ while penalizing the choices of c_1 and c_2 that do not meet the budget constraint.

As the Lagrangian function is concave and hence has a maximum point. Corresponding to the maximum point, we can locate the best choices for c_1 and c_2. To find the maximum point, we need to derive the three first-order conditions of L:

$$\frac{\partial L}{\partial c_1} = \alpha c_1^{\alpha-1} c_2^{1-\alpha} - \lambda = 0, \tag{1.27}$$

$$\frac{\partial L}{\partial c_2} = (1-\alpha) c_1^{\alpha} c_2^{-\alpha} - \lambda/(1+\rho) = 0, \tag{1.28}$$

$$\frac{\partial L}{\partial \lambda} = PV_c - c_1 - c_2/(1+\rho) = 0. \tag{1.29}$$

These first-order conditions can be rewritten as:

$$\alpha \left(\frac{c_1}{c_2}\right)^{\alpha-1} = \lambda \tag{1.30}$$

$$(1-\alpha)(1+\rho)\left(\frac{c_1}{c_2}\right)^{\alpha} = \lambda, \tag{1.31}$$

$$c_1 + c_2/(1+\rho) = PV_c. \tag{1.32}$$

We can use the first two first-order conditions to solve for c_2 as a function of c_1 by eliminating λ and simplifying:

$$\alpha \left(\frac{c_1}{c_2}\right)^{\alpha-1} = (1-\alpha)(1+\rho)\left(\frac{c_1}{c_2}\right)^{\alpha}, \tag{1.33}$$

$\Rightarrow$

$$\alpha \left(\frac{c_2}{c_1}\right) = (1-\alpha)(1+\rho), \tag{1.34}$$

$\Rightarrow$

$$c_2 = \frac{(1-\alpha)(1+\rho)}{\alpha} c_1. \tag{1.35}$$

Substitute equation (1.35) into the third first-order condition:

$$c_1 + \frac{(1-\alpha)(1+\rho)}{\alpha(1+\rho)} c_1 = PV_c \tag{1.36}$$

$\Rightarrow$

$$\frac{\alpha}{\alpha} c_1 + \frac{(1-\alpha)}{\alpha} c_1 = PV_c \tag{1.37}$$

$\Rightarrow$

$$\frac{1}{\alpha} c_1 = PV_c \tag{1.38}$$

$\Rightarrow$

$$c_1^* = \alpha PV_c. \tag{1.39}$$

Substitute equation (1.39) into equation (1.35):

$$c_2^* = \frac{(1-\alpha)(1+\rho)}{\alpha} c_1^* \tag{1.40}$$

$\Rightarrow$

$$c_2^* = \frac{(1-\alpha)(1+\rho)}{\alpha} \alpha PV_c \tag{1.41}$$

$\Rightarrow$

$$c_2^* = (1-\alpha)(1+\rho)PV_c. \tag{1.42}$$

We can verify that c_1^* and c_2^* are optimal solutions by plugging these into the budget constraint. It is apparent that

$$c_1^* + c_2^* \,/(1+\rho) = PV_c. \tag{1.43}$$

The two solutions make a lot of sense. Let $PV_c = \$50,000$ and $\rho = 0.03$. If $\alpha = .20$, the individual cares more about the future consumption. In this case, $c_1^* = (.20)(50,000) = \$10,000$, and $c_2^* = (.80)(1.03)(50000) = \$41,200$. If $\alpha = .80$, the individual cares more about the current consumption. In this case, $c_1^* = (.80)(50000) = \$40,000$, and $c_2^* = (.20)(1.03)(50000) = \$10,300$.

Question: If we modify the utility function into $u(c_1, c_2) = \alpha \ln(c_1) + (1-\alpha)\ln(c_2)$, what are the optimal choices c_1^* and c_2^*? How do these solutions differ from those in our illustration?

The portfolio optimization as a technique can be used to find out the best portfolio for an risk averse investor.

1.4 Finite Future States

In the above, we assume that the future state is certain. This gives us a framework of thinking of now versus the future. In reality, future is uncertain. For simplicity, we can assume that we know a finite number of possible future states, but we do not know which state turns out to be true.

Generally, we can assume that there are S future states and that each state s has its probability p_s. Here, $s = 1, 2, \ldots, S$ and $\sum_{s=1}^{S} p_s = 1$. As an example, assume $S = 2$, then we have two states 1 and 2 with probabilities p_1 and $p_2 = 1 - p_1$, respectively.

A security (for an underlying asset) can be defined as a set of future payoffs such as

$$d = [d^1, d^2, \ldots, d^S]$$

with the current price P. Continue our example, the current price of the security, P, is determined by the future payoffs $[d^1, d^2]$.

If the market has S linearly independent securities, the market is complete. If the market has fewer than S linearly independent securities, the market is incomplete.

Conceptually, it is convenient to consider a special form of security, called the elementary state security, that pays \$ 1 in a particular state but nothing in other states. More specifically, we can assume that there are S elementary state securities:

$$e_1 = [1, 0, 0, \ldots, 0, 0, 0], \tag{1.44}$$

$$e_2 = [0, 1, 0, \ldots, 0, 0, 0], \tag{1.45}$$

$$\vdots$$

$$e_{S-1} = [0, 0, 0, \ldots, 0, 1, 0], \tag{1.46}$$

$$e_S = [0, 0, 0, \ldots, 0, 0, 1]. \tag{1.47}$$

If we stack $e_1, e_2, \ldots, e_S$, we have

$$\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}. \tag{1.48}$$

Apparently, S elementary state securities are linearly independent.

A security's payoffs $d = [d^1, d^2, \ldots, d^S]$ can be viewed as a linear combination of e_s's; that is

$$d = \sum_{s=1}^{S} d^s e_s. \tag{1.49}$$

If we have state prices Ψ_s's, then the price of that security is

$$P = \sum_{s=1}^{S} d^s \Psi_s, \tag{1.50}$$

where $\Psi_s > 0$. It can be proven that a set of positive state prices exists if and only if there are no arbitrage opportunities.

1.5 Expected Utility and Risk Aversion

Let $u(W)$ be the utility of the terminal wealth. Assume that the terminal wealth may take two possible values W_L (low level of wealth) and W_H (high level of wealth). Let p_1 be the probability that W_L occurs so that $0 < p_1 < 1$ and $p_2 = 1 - p_1$ be the probability that W_H occurs. The expected value of the terminal wealth is given by

$$W^e = p_1 W_L + (1 - p_1) W_H. \tag{1.51}$$

Define the expected utility

$$u^e(\widetilde{W}) = p_1 u(W_L) + (1 - p_1) u(W_H) \tag{1.52}$$

where $\widetilde{W}$ is used to denote the randomness of W.

- Risk averse u_1

$$u^e(\widetilde{W}) < u(W^e) \tag{1.53}$$

- Risk neutral u_2

$$u^e(\widetilde{W}) = u(W^e) \tag{1.54}$$

- Risk seeking u_3

$$u^e(\widetilde{W}) > u(W^e) \tag{1.55}$$

- Certainty equivalent C

$$u(C) = u^e(\widetilde{W}). \tag{1.56}$$

For risk averse investors with concave utility functions,

$$C \leq W^e. \tag{1.57}$$

That is, less dollars (C) will attract risk averse investors to go for the risky payoffs W_L and W_H with the expected value of W^e.

Example: In a coin toss investment opportunity, there is the 50-50 chance of winning \$10 or \$ 2. If the investor has a log utility function, $u(W) = \ln(W)$, what is the certainty equivalent?

$$W^e = 0.5(\$10) + 0.5(\$2) = \$6.$$

$$u^e(\widetilde{W}) = 0.5\ln(\$10) + 0.5\ln(\$2) = 1.497866$$

$$C = u^{-1}(1.497866) = e^{1.497866} = \$4.472136 \leq \$6 = W^e.$$

Could you think of the relationship between C and W^e for risk loving investors with convex utility functions?

- Utility function $u(\overline{r}, \sigma^2)$

It should be noted that the return is defined

$$r = \frac{W - W_0}{W_0}. \tag{1.58}$$

where W_0 is the initial wealth. When W is random, so is r. Thus the nature of the return can be characterized by its underlying distribution or by the mean return $\overline{r} = E(r)$ and its standard deviation σ.

To see how $u(\overline{r}, \sigma^2)$ is related to the quadratic utility function, assume that the random variable r is in the feasible range of $r \leq \frac{a}{b}$. Then,

$$\begin{aligned} E(u(r)) &= E(ar - \frac{1}{2}br^2) \\ &= aE(r) - \frac{1}{2}bE(r^2) \\ &= a\overline{r} - \frac{1}{2}b(\overline{r})^2 - \frac{1}{2}b\sigma^2 \end{aligned} \tag{1.59}$$

because $\sigma^2 = E(r^2) - \overline{r}^2$. Hence, $E(u(r))$ can be denoted by $u(\overline{r}, \sigma^2)$. Further, $\partial E(u(r))/\partial \overline{r} = a - b\overline{r}$ and $\partial E(u(r))/\partial \sigma^2 = -b$. If we let $\overline{r} = 1$, then $a - b > 0$.

- Indifference curves I_1, I_2, I_3
- Behaviour economists believe that the outcomes should not be measured in gains and losses. They also believe that the probabilities should not be additive and should be changed to decision weights.

1.6 The Determinants of Intertemporal Behavior

- changes in interest rate: (income effect + substitution effect = total effect) $\leftarrow$depends on the indifference curves.
- changes in income: ($c_1 \uparrow$ and/or $c_2 \uparrow$ if $y_1 + y_2/(1+\rho) \uparrow$).
- liquidity constraint
- changes in preferences
- uncertainty in the future
- taxation and other fiscal policy measures

The above idea can be extended to the situation where future inflation and investment uncertainty exist. The following summary in words might make the points clear.

1. Income flow and spending pattern usually do not coincide
   (a) When income is greater than spending, people tend to invest the surplus.
   (b) When spending is greater than income, people tend to borrow to cover the deficit.
2. People would be willing to give up current consumption only if they expect that reducing current consumption will allow them to have more future consumption.
3. The ratio of future consumption (future dollars) to present consumption (current dollars) is the gross real interest rate. Market forces determine this ratio.
4. Investment is the current commitment of dollars today to obtain future payments that will compensate the investor for the time the funds are committed (real interest rate), for the expected rate of inflation (inflationary adjustment), and for the uncertainty of the future payments (risk premium).

1.7 Three Consumption/Saving Theories

- John Maynard Keynes' consumption theory.

  An average man will spend a part of his disposable income for consumption:

  $$c = \alpha + \beta y \tag{1.60}$$

  where $0 < \beta < 1$ is the marginal propensity to consume (the MPC). Because the saving s can be represented by $y - c$, the above equation can be changed to

  $$s = y - c = -\alpha + (1 - \beta)y \tag{1.61}$$

  where $0 < 1 - \beta < 1$ is the marginal propensity to save (MPS).

  Problems:

  (1) This model does not allow one to differentiate the long term income from the temporary income.

  (2) This model also ignores the borrowing and saving in one's life cycle.

- Milton Friedman's permanent income hypothesis.

  He noted two different income components in actual income — permanent income and transitory income.

  He hypothesized that the MPC out of permanent income is high while the MPC out of transitory income is low.

  Some empirical evidence supports this hypothesis but it does not deal with consumer behaviors' that are related to different stages of the life cycle.

- Franco Modigliani's life cycle hypothesis. He noted that the consumption is related to the individual's life cycle:

  (1) Before having labor income, the individual borrows or has negative wealth to maintain a certain level of consumption.

  (2) During the working ages, the individual maintain the consumption level. He or she saves and accumulates his or her wealth as his or her labor income is greater than the consumption.

  (3) After retirement, the individual receives no labor income. He or she needs to maintain his or her consumption by using the wealth.

  Let c_t be the consumption in period t and y_t the income in period t. The saving in period t is $s_t = y_t - c_t$. This could be viewed as the change of the wealth in period t; that is $\Delta w_t = s_t$. Please note that $\Delta w_t = s_t$ can be negative if one is to borrow rather than save. The total weal in period t can be expressed as $w_t = w_0 + \sum_{t=1}^{t} \Delta w_t$.

  Some empirical evidence supports this hypothesis

1.8 Investments

- Current savings are known while future returns of the current savings are uncertain and inherently risky. The natural question is how to price the future uncertain returns.

- Real investments refer to the activities of using current dollars to increase physical assets, improve human resources, and innovate in production process. Real investments are the driving forces of a growing economy. Real investments and their returns will influence market forces that will determine returns of corresponding financial investments.

- Financial investments

  Financial investments are the activities of putting current dollars into the financial *claims* (stock ownership, bond ownership, and so on) to real investments so that investors will be compensated with future payments for the time the funds are committed (real interest rate), for the expected rate of inflation (inflationary adjustment), and for the uncertainty of the future payments (risk premium)

  Prior to a decision to invest financially, it is prudent to consider other "investments" to hedge unexpected events that may cause financial hardship first. (1) Insurance—life insurance, health insurance, disability insurance, auto/home insurance, and liability insurance; (2) cash reserves— cash or cash equivalent investments (e.g., money market funds) that can cover six months living expenses.

  The five steps for implementing financial investments: (1) investment policy (investment goals, liquidity, time horizon, tax considerations, acceptable risk levels, and benchmark portfolios), (2) security analysis (economic conditions, future trends, and investments that are consistent with investment goals), (3) portfolio construction (fund allocation into various asset classes to meet investment goals via stock selection, market timing, and diversification), (4) portfolio revision (repetition of (1)-(3)), and (5) portfolio performance evaluation (evaluation of investment performance—risk, return, and relative ranking of an asset class in relation to its benchmark portfolio and its universe over time and in up and down markets).

- Other non-financial investments: Real estate, precious metals, diamonds, fine arts, antiques, and other collectibles

- Inflation adjustments

  For both real and financial investments, the investor in general would expect that the future payments will have inflation adjustments so returns will be greater in real terms.

- Risk premium

  For both real and financial investments, the investor in general would expect that the expected future payments can compensate his/her risk taking activities as these activities are uncertain and risky.

  The fundamental risk in investments can take the following forms:

  - Business risk (a) uncertainty of cash flow due to the nature of a firm's business and (b) unstable sales and operating leverage
  - Financial risk (a) debt financing, (b) debt payments, and (c) stockholder's interest as a residual firm value after debt holder's interest
  - Liquidity risk uncertainty in the secondary market for an investment: (a) time taken to liquidate an asset and (b) the price at which the asset is liquidated
  - Exchange risk uncertainty in exchange rates affect (a) payments and (b) investment returns in home currency

  - Country risk (a) uncertainty and (b) instability in the political or economic environment in a country

  All of these forms of risk will affect (increase) the risk premium of an investment.

- Let the initial wealth be W_0 at time $t = 0$ and the terminal wealth be W_T at time $t = T$, where T is the number of periods (say, years). The rate of return on the initial wealth W_0 over T periods (say, years) is given by

$$r = \frac{W_T - W_0}{W_0}. \tag{1.62}$$

  r is also called the holding period yield (HPY);

$$HPY = r. \tag{1.63}$$

  Sometimes, we also use the concept of the holding period return (HPR) over T periods (say, years):

$$HPR = \frac{W_T}{W_0} = \frac{W_0}{W_0} + \frac{W_T - W_0}{W_0} = 1 + r. \tag{1.64}$$

  As can be seen, $HPR = 1 + HPY$ for $T > 0$ periods (say, years).

  Example: Let $W_0 = 100$ and $W_T = 150$, where $T = 5$ years. Then, for T years, $HPY = \frac{150-100}{100} = .50$ and $HPR = 1 + HPY = 1 + .50 = 1.50$.

  If HPR and HPY are used for T years, we can find the *annual* HPR and HPY using the following equations

$$Annual \;\; HPR = HPR^{\frac{1}{T}} \tag{1.65}$$

  and

$$Annual \;\; HPY = Annual \;\; HPR - 1. \tag{1.66}$$

  Example: Using the data from the previous example, we can find out the annual HPR and annual HPY as follows.

$$Annual \;\; HPR = HPR^{\frac{1}{T}} = (1.50)^{\frac{1}{5}} = 1.0845 \tag{1.67}$$

  and

$$Annual \;\; HPY = Annual \;\; HPR - 1 = 1.0845 - 1 = .0845. \tag{1.68}$$

  It is possible to calculate HPR for 5 years from Annual HPR: $1.0845^5 = 1.50$.

- Historical mean returns

  There are two different ways to compute historical mean returns ($r_t, t = 1, 2, \ldots, T$). One is the arithmetic mean return (AM)

$$AM = \frac{\sum_{t=1}^{T} HPY_t}{T} = \frac{\sum_{t=1}^{T} r_t}{T}. \tag{1.69}$$

  The other is the geometric mean return (GM)

$$GM = \left(\prod_{t=1}^{T} HPR_t\right)^{\frac{1}{T}} - 1 = \left(\prod_{t=1}^{T} (1 + r_t)\right)^{\frac{1}{T}} - 1. \tag{1.70}$$

Table 1.3: Which Is Better? AM or GM

Time	$t=0$	$t=1$	$t=T=2$	AM/GM
Wealth	\$100	\$50	\$100	
r_t	NA	-.50	1.00	$AM=(-.50+1.00)/2=.25$
$1+r_t$	NA	.50	2.00	$GM=\sqrt{(.50)(2.00)}-1.00=0$

First, we use one example to show $AM \geq GM$. Please see Table 1.3.

From time $t=0$ to time $t=1$, $r_1 = \frac{\$50-\$100}{\$100} = -.50$ From time $t=1$ to $t=T=2$, $r_2 = \frac{\$100-\$50}{\$50} = 1.00$. Therefore, $AM = \frac{HPY_1+HPY_2}{2} = \frac{r_1+r_2}{2} = \frac{-.50+1.00}{2} = .25$. $GM = \sqrt{(1+r_1)(1+r_2)} - = \sqrt{(.50)(2.00)} - 1 = 0$.

Second, we can show a rigorous mathematical proof for $AM \geq GM$. We can show that GM is always smaller than or equal to AM as follows. This is not apparent and difficult to show. Therefore, if we can show $AM+1 \geq GM+1$, we can infer $AM \geq GM$. Let $T=2$ so that $t=1,2$ and hence we have r_1 and r_2. We can let $X=1+r_1$ and $Y=1+r_2$. Then we can write

$$AM+1 = \frac{X+Y}{2} = \frac{(1+r_1)+(1+r_2)}{2} = \frac{r_1+r_2}{2}+1 \tag{1.71}$$

and

$$GM+1 = \sqrt{XY} = \sqrt{(1+r_1)(1+r_2)}. \tag{1.72}$$

Recall $(a-b)^2 \geq 0$ for any constants a and b. It is known that

$$(\sqrt{X}-\sqrt{Y})^2 \geq 0, \tag{1.73}$$

which implies

$$(\sqrt{X})^2 - 2\sqrt{XY} + (\sqrt{Y})^2 \geq 0 \tag{1.74}$$

or

$$\frac{X+Y}{2} \geq \sqrt{XY}. \tag{1.75}$$

The last inequality implies

$$AM+1 \geq GM+1 \tag{1.76}$$

or

$$AM \geq GM. \tag{1.77}$$

- Portfolio return (holding period yield).

  The mean historical rate of return for a portfolio of investments is measured as the weighted average of the HPYs for the individual investments in the portfolio, or the overall change in value of the original portfolio. The weights used in computing the averages are the relative beginning market values for each investment; this is referred to as dollar-weighted or value-weighted mean rate of return (see Table 1.4).

  Alternatively, one can compute the portfolio return based on the total beginning and ending values of the portfolio (see Table 1.4) as

$$\frac{\$21,900,000}{\$20,000,000} - 1 = 0.095. \tag{1.78}$$

Table 1.4: Portfolio return (holding period yield)

Stock	number of Shares	Begin Price	Beginning Mkt. Value	Ending Price	Ending Mkt. Value	HPR	HPY	Market Wt.	Wtd. HPY
A	100,000	\$10	\$1,000,000	\$12	\$1,200,000	1.20	20%	0.05	0.010
B	200,000	\$20	\$4,000,000	\$21	\$4,200,000	1.05	5%	0.20	0.010
C	500,000	\$30	\$15,000,000	\$33	\$16,500,000	1.10	10%	0.75	0.075
Total			\$20,000,000		\$21,900,000				0.095

- Historical inflation and returns on Treasury bills and bonds, corporate bonds, and stocks

  It is generally observed that returns on Treasury bill and bonds are low on average and have less variability while returns on corporate bonds and stocks are higher on average and have higher variability. Over the long run, returns on stocks can beat inflation but not in some years.

1.9 Review Questions

1. What is a utility function? What are the commonly used utility functions? What are the key features of the utility function for non-satiable and risk averse consumers/investors?
2. What is the absolute risk aversion coefficient? What is the relative risk aversion coefficient?
3. Please describe the single-person two-period model for intertemporal saving behavior discussed in our lecture
4. What are the critical assumptions for a single-person two-period model for intertemporal saving behavior discussed in our lecture?
5. What are the determinants of intertemporal consumption/saving behavior?
6. Please differentiate a complete market from an incomplete market. What are elementary state securities?
7. What is the expected utility? Please explain risk aversion, risk neutral, and risk seeking/loving based on expected utility functions.
8. If you invest today, what are the components of the future return that you look for?
9. Please explain three consumption theories and how they differ.
10. What are the differences between real investments and financial investments?
11. How do stocks differ from bonds as financial investment instruments?
12. What are the two major considerations before you start to invest in either real or financial assets?
13. What are the five steps for implementing financial investments?
14. What are the likely sources of risk in financial investments?
15. What is the return on investments? How to calculate it?
16. What are holding period return and holding period yield?
17. Is the arithmetic mean always at least as large as the geometric mean? Can you show why or why not?
18. Assume that the consumer/investor has the utility function:

$$u(c_1, c_2) = c_1^{\alpha} c_2^{1-\alpha} \tag{1.79}$$

and the budget constraint

$$c_1 + c_2/(1+\rho) = PV_c, \tag{1.80}$$

where c_1 and c_2 are the consumption bundles in periods 1 and 2, respectively, α is the parameter, ρ is the real interest rate, and PV_c is the total present value of all consumption bundles.

Please find the optimal consumption bundles in periods 1 and 2, respectively.

If the consumer/investor changes his/her utility function to $u(c_1, c_2) = \alpha \ln(c_1) + (1-\alpha)\ln(c_2)$, what are the optimal bundles c_1^* and c_2^*? How do the new optimal consumption bundles differ from the previous ones?

1.10 Review on the Lagrangian Method

Let the objective function be $f(x_1, x_2)$, which is continuous and, in this case, concave. Let $g(x_1, x_2)$ be the constrain function which is also continuous and defined over a convex set. Generally, economists require $x_1 > 0$ and $x_2 > 0$ since nonnegative quantities are often involved in economics. The maximization problem can be written as

$$\max_{x_1, x_2} f(x_1, x_2) \tag{1.81}$$

$$s.t. g(x_1, x_2) \leq 0. \tag{1.82}$$

To solve the optimal x_1^* and x_2^*, we can form the Lagrangian function and introduce the Lagrangian multiplier λ:

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda g(x_1, x_2). \tag{1.83}$$

The Lagrangian function is used later to find the optimal x_1^* and x_2^*.

For those who are somewhat rusty about partial derivatives, let us review some basic rules about partial derivatives. Note that, for $y = f(x_1, x_2) = x_1^n x_2^m$ where n and m are constants, the partial derivatives are given as follows:

$$\frac{\partial f(x_1, x_2)}{\partial x_1} = n x_1^{n-1} x_2^m$$

and

$$\frac{\partial f(x_1, x_2)}{\partial x_2} = m x_1^n x_2^{m-1}.$$

For $L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda g(x_1, x_2)$,

$$\frac{\partial L(x_1, x_2, \lambda)}{\partial x_1} = \frac{\partial f(x_1, x_2)}{\partial x_1} + \lambda \frac{\partial g(x_1, x_2)}{\partial x_1},$$

$$\frac{\partial L(x_1, x_2, \lambda)}{\partial x_2} = \frac{\partial f(x_1, x_2)}{\partial x_2} + \lambda \frac{\partial g(x_1, x_2)}{\partial x_2},$$

and

$$\frac{\partial L(x_1, x_2, \lambda)}{\partial \lambda} = g(x_1, x_2).$$

Now use the above rules to solve for the optimal values x_1^*, x_2^*, and λ^*.Taking derivative of $L(x_1, x_2, \lambda)$ with respect to x_1, x_2, and λ yields $\frac{\partial L(x_1, x_2, \lambda)}{\partial x_1}$, $\frac{\partial L(x_1, x_2, \lambda)}{\partial x_2}$, and $\frac{\partial L(x_1, x_2, \lambda)}{\partial \lambda}$, respectively. Setting these partial derivatives to zero gives the first-order conditions (F.O.C.):

$$\frac{\partial L(x_1^*, x_2^*, \lambda^*)}{\partial x_1^*} = 0, \tag{1.84}$$

$$\frac{\partial L(x_1^*, x_2^*, \lambda^*)}{\partial x_2^*} = 0, \tag{1.85}$$

and

$$\frac{\partial L(x_1^*, x_2^*, \lambda^*)}{\partial \lambda^*} = 0. \tag{1.86}$$

These equations can be used to solve x_1^*, x_2^* and λ^*.

Now let us see an example. Let us take an example. Let the choice variables be the consumption in period 1 $C_1 > 0$ and that in period 2 $C_2 > 0$. The objective function—the utility function in this case—is

$$U(C_1, C_2) = C_1^{\alpha} C_2^{1-\alpha}. \tag{1.87}$$

Let the constrain function—the intertemporal budget constrain in this case—be

$$C_1 + \frac{C_2}{1+r} = W, \tag{1.88}$$

where W is the present value of the total wealth and r is the interest rate.

In order to solve for the optimal level of the consumption in periods 1 and 2, form the Lagrangian function

$$L(C_1, C_2, \lambda) = C_1^{\alpha} C_2^{1-\alpha} + \lambda \left(W - C_1 - \frac{C_2}{1+r} \right). \tag{1.89}$$

Now find the F.O.C.:

$$\frac{\partial L(C_1^*, C_2^*, \lambda^*)}{\partial C_1^*} = \alpha C_1^{*\alpha-1} C_2^{*1-\alpha} - \lambda^* = 0, \tag{1.90}$$

$$\frac{\partial L(C_1^*, C_2^*, \lambda^*)}{\partial C_2^*} = (1-\alpha) C_1^{*\alpha} C_2^{*-\alpha} - \lambda^* \frac{1}{1+r} = 0, \tag{1.91}$$

and

$$\frac{\partial L(C_1^*, C_2^*, \lambda^*)}{\partial \lambda^*} = W - C_1^* - \frac{C_2^*}{1+r} = 0. \tag{1.92}$$

The above equations can be used to solve for λ^*, C_1^* and C_2^*.

Equations (1.90) and (1.91) can be used to eliminate λ^*; that is,

$$\alpha C_1^{*\alpha-1} C_2^{*1-\alpha} = \lambda^*$$

$$(1-\alpha) C_1^{*\alpha} C_2^{*-\alpha} = \lambda^* \frac{1}{1+r}$$

$\Rightarrow$Combine the above two equations

$$\frac{\alpha C_1^{*\alpha-1} C_2^{*1-\alpha}}{(1-\alpha) C_1^{*\alpha} C_2^{*-\alpha}} = 1+r$$

$\Rightarrow$

$$\frac{\alpha C_2^*}{(1-\alpha) C_1^*} = 1+r$$

$\Rightarrow$

$$C_2^* = \frac{(1-\alpha)(1+r)}{\alpha} C_1^* \tag{1.93}$$

Substituting equation (1.93) into equation (1.92) yields

$$C_1^* = W - \frac{C_2^*}{1+r} = W - \frac{1}{(1+r)} \left[\frac{(1-\alpha)(1+r)}{\alpha} C_1^* \right]$$

$\Rightarrow$

$$C_1^* = W - \left[\frac{(1-\alpha)}{\alpha} C_1^*\right]$$

$\Rightarrow$

$$C_1^* + \left[\frac{(1-\alpha)}{\alpha} C_1^*\right] = W$$

$\Rightarrow$

$$\frac{1}{\alpha}\left[\alpha C_1^* + (1-\alpha) C_1^*\right] = W$$

$\Rightarrow$

$$\frac{1}{\alpha} C_1^* = W$$

$\Rightarrow$

$$C_1^* = \alpha W. \tag{1.94}$$

Substituting equation (1.94) into (1.93) gives

$$C_2^* = \frac{(1-\alpha)(1+r)}{\alpha}(\alpha W) = (1-\alpha)(1+r)W. \tag{1.95}$$

As can be seen that C_1^* in equation (1.94) and C_2^* in equation (1.95) are functions of α, r, and W, which are known, and hence these are the optimal solutions for C_1 and C_2.

Now let us solve for λ^*. Substituting equations (1.94) and (1.95) into equation (1.90) gives

$$\lambda^* = \alpha[\alpha W]^{\alpha-1}[(1-\alpha)(1+r)W]^{1-\alpha}$$

$\Rightarrow$

$$\lambda^* = \alpha\left[\frac{\alpha W}{(1-\alpha)(1+r)W}\right]^{\alpha-1} = \alpha\left[\frac{\alpha}{(1-\alpha)(1+r)}\right]^{\alpha-1} = \alpha\left[\frac{(1-\alpha)(1+r)}{\alpha}\right]^{1-\alpha}. \tag{1.96}$$

The above is the optimal solution for λ.

Now turn to the question of why the optimal choices are consistent with the condition of equality between the marginal rate of substitution and that of transformation, $MRS = MRT$. Equations (1.90) and (1.91) can be written as the marginal utility form

$$MU_{C_1} = \alpha C_1^{*\alpha-1} C_2^{*1-\alpha} = \lambda^* \tag{1.97}$$

and

$$MU_{C_2} = (1-\alpha) C_1^{*\alpha} C_2^{*-\alpha} = \lambda^* \frac{1}{1+r}. \tag{1.98}$$

Note that when the utility level is fixed, we have

$$MU_{C_1}\Delta C_1 + MU_{C_2}\Delta C_2 = 0. \tag{1.99}$$

where $MU_{C_1} > 0$, $MU_{C_2} > 0$, and $\Delta C_1$and ΔC_2 can be positive, zero or negative. From equation (1.99), one can get the marginal rate of substitution

$$\frac{\Delta C_2}{\Delta C_1} = -\frac{MU_{C_1}}{MU_{C_2}} = MRS_{C_1}^{C_2}. \tag{1.100}$$

Substituting equations (1.97) and (1.98) into equation (1.100) gives the optimal condition where the marginal rate of substitution equals the marginal rate of transformation

$$MRS_{C_1}^{C_2} = -(1+r) = MRT.$$

Chapter 2

Financial Markets and Security Trading

2.1 Learning Objectives

- Brokerage Firms, Brokers, Trading and Commissions
- Types of orders
- Margin transactions
- Short sales
- Call and continuous markets
- Organization of the stock exchanges
- Transaction costs
- Kelly's criterion

2.2 Brokerage Firms, Brokers, Trading and Commissions

- Brokerage firms: full service and discount service
- Brokers - registered representatives who facilitate the buying and selling securities
- Commissions - an amount charged to investors for the services; self-interest in generating commissions

2.3 Types of Orders

Order specifications: (1) name of the company, (2) buy or sell, (3) size of the order, (4) how long the order is to be outstanding, and (5) what type of order is to be used.

When one is going to buy and sell a security, he or she needs to place an order (instruction for buying or selling the security) to a (full service or discount/electronic) broker who provides the necessary service for a fee called commission.

The quote provided by the broker may looks like:

XYZ	Bid	Ask	Last Trade
Price	50.125	50.250	50.125
Lot	300	320	200

Order size: board lot (e.g. 100 shares) or odd lot (e.g. 36 shares)

Time limit: On May 21, 200X, buy 100 shares of XYZ at 50.125 good for today (an day order) or good till May 25, 200X (a good through order).

Open order: On May 21, 200X, buy 200 shares of XYZ at 50.125 good-till-cancelled order (an open order).

At the market order: If one buys or sells 200 shares of XYZ at the market, the best price between 50.125 and 50.250 is normally obtained. This is also a day order.

Limit order: If one buys 100 shares of XYZ at 50 (or sell 100 shares of XYZ at 51), this order is a limit order.

Stop order: These come with different types.

a Hard stop order

An order that allows you to decide on a hard/firm price at which to sell your stock.

b Trailing stop order

Follows the movement of a stock price as the price climbs then sells at the specified stop limit price or percentage you set.

c Sell on stop order

The sell on stop order is a sell order that specifies a stop price. The stop price must be below the market price at the time when the order is placed. If the security's price reaches or passes the stop price, then a market order is created.

d Buy on stop order

The buy on stop order is a buy order that specifies a stop price. The stop price must be above the market price at the time when the order is placed. If the security's price reaches or passes the stop price, then a market order is created.

2.4 Margin Transactions

A cash account with a brokerage firm is like a regular chequing account: deposits (cash and the proceeds from selling securities) must cover withdrawals (cash and the costs of purchasing securities).

A margin account is like a chequing account that has overdraft privileges within a certain limit.

Margin purchase occurs when the investor buys securities with a percentage of his or her own cash and the rest borrowed from the broker normally at 2% plus the prime rate. The broker borrows from a bank at the call loan rate. The securities purchased are used as collateral on the loan from the bank by the broker.

The minimum percentage of the purchase price that must come from the investor's own funds is known as the minimum margin requirement. The long position requires a minimum margin requirement from 30% to 100% of the market depending on the riskiness of securities. The minimum margin requirement for short positions will be discussed after the concept of short sales is introduced.

Actual margin can be, for the margin purchase, computed as

$$\text{Actual Margin} = (\text{Market Value of Assets - Loan})/(\text{Market Value of Assets}).$$

The daily computation of the actual margin is known as having the account marked to the market. When the actual margin (say 20%) falls below the minimum margin requirement (say 30%), the investor would be subject to a margin call from the broker to add more funds into the account to meet the minimum margin requirement.

The use of margin purchases is a form of financial leverage. When the stock price is rising, it generates higher returns; when the stock price is falling, it causes larger losses. Therefore, the use of margin purchases often occurs when the investor expects that the stock price will rise in the near future.

Example: Let the margin be 60% and the interest rate 8%. If one buys 100 shares of XYZ at the price of \$50 and sells the shares at the price of \$65, his or her return will be

$$\frac{(\$65 - 50) \times 100}{\$50 \times 100} = 30\%$$

without using any margin; his or her return will be

$$\frac{\$1500 - (1 - 0.60)\$50 \times 100 \times .08}{0.60 \times \$50 \times 100} = 44.7\%$$

with the 60% margin. Similarly, if he or she sells the shares at the price of \$35, his or her return will be

$$\frac{(\$35 - 50) \times 100}{\$50 \times 100} = -30\%$$

without using any margin; his or her return will be

$$\frac{-\$1500 - (1 - 0.60)\$50 \times 100 \times .08}{0.60 \times \$50 \times 100} = -55.3\%$$

with the 60% margin.

Here we note that the impact of using margin is not symmetrical—the downside percentage is great than the upside percentage. If one considers the trading commission, this uneven patten will be more prominent. Assuming that the commission for the trade is \$100, then we have the following results. If one buys 100 shares of XYZ at the price of \$50 and sells the shares at the price of \$65, his or her return will be

$$\frac{\$1500 - (1 - 0.60)\$50 \times 100 \times .08 - \$100}{0.60 \times \$50 \times 100} = 41.3\%$$

with the 60% margin. If he or she sells the shares at the price of \$35, his or her return will be

$$\frac{-\$1500 - (1 - 0.60)\$50 \times 100 \times .08 - \$100}{0.60 \times \$50 \times 100} = -58.7\%$$

with the 60% margin.

2.5 Short Sales

An old saying is to "buy low, sell high." For short sales, it becomes "sell high, buy low." Short sales are accomplished by selling the shares borrowed from the lender through the broker and buying them back returning them to the lender.

There are various rules governing short sales. For example, in Canada, the short sales may not be made when the market price of the security is falling. More specifically, the Toronto Stock Exchange rules that a short sale must not be made on the Exchange below the price of the last sale of a board lot. This implies that a short sale may be effected on an uptick (for a price higher than that of the previous trade) or on a zero tick (for a price equal to that of the previous trade).

Due to the complexity of short sales, dividends, annual reports, and voting rights need to be discussed. The lender of the securities still gets dividends and annual reports but no longer have the voting rights. The purchaser of the securities from short sales gets dividends, annual reports and voting rights. The short seller has to pay the dividends to the lender. The broker has to provide an additional copy of annual reports to the lender.

The short position typically has a minimum margin requirement from 30% to 100% in the United States (or 130% to 200% in Canada) depending on the riskiness of securities. Please note that the minimum margin requirement 30% in the United States (or 130% in Canada) means that the investor must have his or her own funds worth 30% of the stock and that he or she must also keep 100% of the stock price in cash from short selling (this is why one must add 100% to 30% to get 130% in Canada). In the United States, the actual margin for the short sales is computed as

$$\text{Actual Margin} = [\text{Short Sale Proceeds} + \text{Initial Margin}) - \text{Loan}]/\text{Loan},$$

where [(Short Sale Proceeds + Initial Margin) - Loan] represents the investor's equity while Loan represents the current value of the loan of the short seller. For example, an investor uses the initial margin of \$30 to short sell 100 shares of stock XYZ at \$100. After the short sales, the price of stock XYZ rises to \$120. Then the actual margin becomes

$$[(\$100 + \$30)(100) - (\$120)(100)]/(\$120)(100) = (\$10)/(\$120) = 8.3\%,$$

which would be

$$1 + 0.083 = 108.3\%$$

in Canada. If the actual margin is lower than the minimum margin of 30% (or 130% in Canada), a margin call will be made to the short seller.

2.6 Call and Continuous Markets

In call markets, trading is allowed only at certain specified time. Typically, a call market has an auction at a specific location at a specific time. In many stock exchanges (such as the Toronto Stock Exchange), a visible single price is formed in the call market at the market opening. This price maximizes order matching through a fair and transparent process for all market participants.

In continuous markets, trades occur at any time. The organized exchanges (such as the Toronto Stock Exchange) which have their physical locations and trading hours, represent continuous markets beyond the market opening. In continuous markets, the demand and supply schedules are updated instantly and when a bid order and an ask order match in price and quantity, trades are executed and prices are discovered.

2.7 Organization of the Toronto Stock Exchange

Trading on organized exchanges is conducted only by members who have seats and specialize in a specific group of securities.

Exchange members fall into one of two categories, depending on the type of permitted trading activity. Order traders take order that the public has placed with brokerage firms and see that they are executed on the exchange; the brokerage firms that they work for are paid commissions by the customers for their services.

Registered traders are allowed to handle customer's orders and also to trade on their firm's behalf. A special sub-group of these traders are known as registered traders responsible for a stock or designated market markers (DMM).

When orders are placed, one may observe bid and ask prices and lots. The difference between the bid and ask price is called the bid-ask spread. When the bid-ask spread is large enough, the registered trader may buy or sell so that a trade is possible.

The designated market makers are in charge of maintaining an orderly market in their assigned stocks. They play two roles; they maintain a limit order book for unexecuted trades, and they act as a dealer, trading in their stocks for their own accounts.

Foreign security markets each have their own particular operating procedures, rules, and customs. Dealers typically make money trading with liquidity-motivated traders and lose money with adept information-motivated traders. A dealer must set a bid-ask spread that attracts sufficient revenue from the former group but avoids excessive losses to the latter.

Clearinghouses facilitate the transfer of securities and cash between buyers and sellers. Most Canadian securities transactions are cleared electronically.

2.8 Transaction Costs

- The bid-ask spread
- Commissions
- Interest rates of borrowing
- Time
- Other resources

2.9 Kelly's Criterion

Suppose that the investor has the initial wealth W_0 and faces the opportunity of *doubling* the investment with probability p or *zero return* with probability $1-p$. What proportion, α, of the initial wealth, W_0, should the investor optimally invest for this opportunity? The amount of αW_0 is called the wager, or the amount to be risked in a game.

It is clear that if the investor wins, his/her initial wealth will change with the factor of $1 - \alpha + 2\alpha = 1 + \alpha > 1$. If he/she loses, his/her initial wealth will change with the factor of $1 - \alpha < 1$.

If we use the log utility function $u(x) = \ln(x)$, the expected utility function will be

$$u = p\ln(1+\alpha) + (1-p)\ln(1-\alpha). \tag{2.1}$$

Note that the above expected utility function is a monotonic affine transformation from the following expected utility function.

$$u = p\ln[W_0(1+\alpha)] + (1-p)\ln[W_0(1-\alpha)] = \ln(W_0) + p\ln(1+\alpha) + (1-p)\ln(1-\alpha), \tag{2.2}$$

where $\ln(W_0)$ is a constant and does not affect the maximization of the expected utility function.

The investor wishes to maximize the expected utility u by selecting the optimal α, α^*.

To find α^*, take the derivative of u with respect to α and set the derivative to zero:

$$\frac{p}{1+\alpha^*} - \frac{1-p}{1-\alpha^*} = 0. \tag{2.3}$$

Rearrange the above expression:

$$p(1-\alpha^*) - (1-p)(1+\alpha^*) = 0. \tag{2.4}$$

Then we have

$$\alpha^* = 2p - 1. \tag{2.5}$$

It is clear that when $p < 0.5$, we will end up with a negative α^*. If the investor is not allowed to offer this opportunity but taking this opportunity offered by others, his/her investment is restricted to zero, $\alpha^* = 0$ if $p \leq 0.5$. We can have a positive α^* only if $p > 0.5$, See Table 2.9.

Substitute $\alpha^* = 2p - 1$ given by equation (2.12) into the utility function given by equation (2.1):

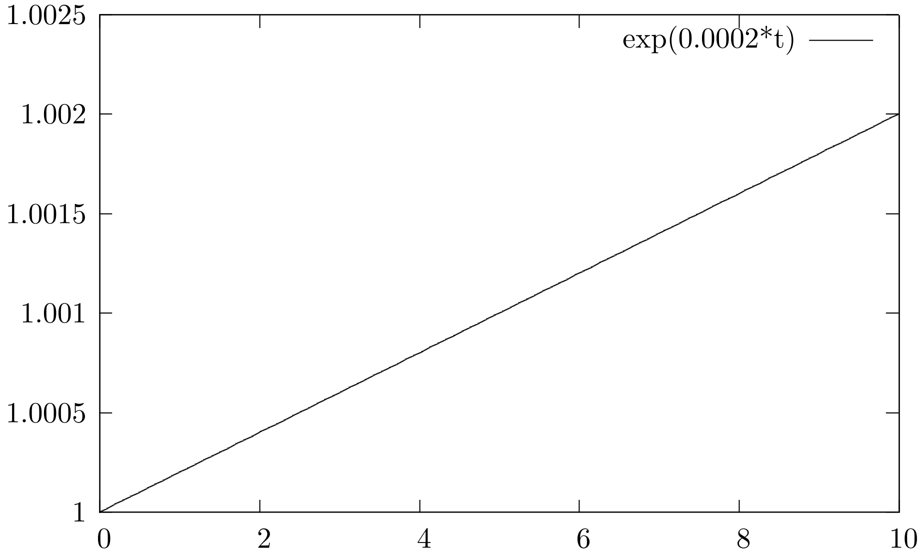
$$u = p\ln(2p) + (1-p)\ln(2-2p) = p\ln(p) + (1-p)\ln(1-p) + \ln(2). \tag{2.6}$$

Table 2.1: p versus α^* with $p \geq 0.5$.

p	$\alpha^* = 2p - 1$
0.5	0.0
0.6	0.2
0.7	0.4
0.8	0.6
0.9	0.8
1.0	1.0

If $p = 0.51$, $u = 0.0002000133$ and $W = e^u = 1.0002$. This is a gain of 0.02 % of W_0.

How do we grow the investment over time t (more than one period)? We can calculate the wealth over t; that is, $W_t = e^{ut}$.



If p is much greater than 0.50 (say 0.60 or 0.70), then the wealth can grow much faster. In practice, selecting investments that have greater future gains ensures a higher p.

Now we further introduce the net fractional odds, b, received on the wager αW_0, which makes Kelly's criterion more realistic. In our previous example, if we bet αW_0 as the wager, we will either double the wager to $2\alpha W_0$ (net αW_0 after subtracting the wager) with probability p or get nothing (net $-\alpha W_0$ after subtracting the wager) with probability $1 - p$. In this case, the net fractional odds received on the wager, $b = \frac{\alpha W_0}{\alpha W_0}$, is 1. Now we change the payoffs. If we bet αW_0, we will either win 160% of the wager (net $0.6\alpha W_0$ after subtracting the wager) with probability p or get nothing (net $-\alpha W_0$ after subtracting the wager) with probability $1 - p$. In this case, the net fractional odds received on the wager, $b = \frac{0.6\alpha W_0}{\alpha W_0}$, is 0.60. Now we can rewrite the expected utility function by incorporating b:

$$u = p \ln(1 + b\alpha) + (1 - p) \ln(1 - \alpha). \tag{2.7}$$

Take the derivative of u with respect to α and set the derivative to zero:

$$\frac{bp}{1 + b\alpha^*} - \frac{1 - p}{1 - \alpha^*} = 0. \tag{2.8}$$

Rearrange the above expression:

$$bp(1-\alpha^*) - (1-p)(1+b\alpha^*) = 0. \tag{2.9}$$

$$bp - bp\alpha^* - (1-p) - (1-p)b\alpha^* = 0. \tag{2.10}$$

$$bp + p - 1 = b\alpha^*. \tag{2.11}$$

Then we have

$$\alpha^* = \frac{bp+p-1}{b}. \tag{2.12}$$

To ensure $\alpha^* > 0$, we need $b > 0$ and $(b+1)p > 1$. For example, if $p = 0.5$, then $b > \frac{1-p}{p} = 1$. If $p = 0.4$, then $b > \frac{1-p}{p} = 1.5$. If $p = 0.6$, then $b > \frac{1-p}{p} \approx 0.66$. Further, when $b = 1$, $\alpha^* = 2p - 1$.

2.10 Review Questions

1. What are the differences between the full service and discount brokerage firms?

2. What are the key pieces of information in an order for buying or selling a stock?

3. Please define the following concepts: an open order, a fill-or-kill order, an all or none order, an at the market order a limit order, a stop loss order and a stop buy.

4. What is a margin investment account? What is the minimum margin requirement?

5. What is the actual margin for long positions? How to calculate it?

6. Please use an example to show the asymmetric impact of using margin on returns between downs and ups in prices when having long positions in stocks.

7. What is a short sale? Please explain the parties involved and their roles in a short sale transaction?

8. What is the actual margin for short positions? How to calculate it?

9. What is the Kelly's criterion? Is it equivalent to the maximization of the expected log-utility of wealth?

Chapter 3

Portfolio Theory

3.1 Learning Objectives

- Evaluation of risky assets
- Portfolio returns
- Efficient frontier
- Optimal portfolios
- The market model

3.2 Evaluation of Risky Securities

The valuation of risky assets requires the explicit or implicit analysis of the events on which the assets' payments are contingent.

When evaluating a risk-free asset, it is straightforward. The present value of the future payment \$1 at time t is $PV(\$1,t) = \frac{1}{(1+r)^t}$, where r is the interest rate. If the payments are $D(t)$, $t = 1, 2, \ldots, T$, they the price of the asset is

$$P = \sum_{t=1}^{T} PV(\$1,t) \cdot D(t). \tag{3.1}$$

When evaluating a risky asset, we need to identify the present value of the future payment \$1 at time say t in a particular state s $PV(\$1,t,s) = \frac{1}{(1+r_{t,s})^t}$, where $r_{t,s}$ is the interest rate for state s at time t. If the payments are $D(t,s)$, $t = 1, 2, \ldots, T$ and $s = 1, 2, \ldots, S_t$, and the probability for state s at time t is $p(t,s)$. S_t is the maximum number of states at time t. When it is fixed for all t, we can write $S_t = S$ for all t. Then the expected price of the asset is

$$E(P) = \sum_{t=1}^{T} \sum_{s=1}^{S_t} p(t,s) \cdot PV(\$1,t,s) \cdot D(t,s). \tag{3.2}$$

A contingent payment is a guaranteed cash flow that will be made if, and only if, a particular event or state of the world occurs.

Identifying all possible future events or states of the world would be difficult because many factors are and will be at play.

Therefore there are many different ways of defining risk. One of the commonly used ones is that risk is the possibility of potential loss.

Sources of risk, as discussed previously, are: (1) business risk, (2) financial risk, (3) liquidity risk, (4) exchange risk, (5) country risk, and so on.

Measures of risk: (1) distribution of returns, (2) expected value and variance of returns, and (3) correlations among returns. Other measures are semi-variance, probability of loss, average loss, and value-at-risk.

Distribution of returns: parametric versus nonparametric, discrete versus continuous distributions

The expected value of return or the mean return on asset i is

$$E(r_i) = \bar{r}_i \tag{3.3}$$

and the variance of returns on asset i is

$$E(r_i - \bar{r}_i)^2 = \sigma_i^2 = \sigma_{ii}, \tag{3.4}$$

where σ_i^2 and σ_{ii} are the same but are used interchangeably in different contexts. The standard deviation is therefore

$$\sqrt{E(r_i - \bar{r}_i)^2} = \sqrt{\sigma_i^2} = \sqrt{\sigma_{ii}} = \sigma_i. \tag{3.5}$$

The covariance between returns on assets i and j is

$$E[(r_i - \bar{r}_i)(r_j - \bar{r}_j)] = \sigma_{ij}. \tag{3.6}$$

The correlation between returns on assets i and j is

$$\frac{\sigma_{ij}}{\sigma_i \sigma_j} = \rho_{ij}. \tag{3.7}$$

Please note

$$\sigma_{ij} = \rho_{ij} \sigma_i \sigma_j. \tag{3.8}$$

If asset i's return r_i only takes discrete values r_{is}'s with state probabilities p_s's in state $s = 1, 2, \ldots, S$, then we can compute the mean return on asset i, $E(r_i)$, as follows:

$$E(r_i) = \bar{r}_i = \sum_{s=1}^{S} p_s r_{is}. \tag{3.9}$$

Similarly, we can compute the variance of asset i's return σ_i^2 as follows:

$$\sigma_i^2 = E(r_i - \bar{r}_i)^2 = \sum_{s=1}^{S} p_s (r_{is} - \bar{r}_i)^2. \tag{3.10}$$

Example:

Table 3.1: Mean and variance of a random return

p_s	20%	30%	30%	20%
r_s	-10.000%	6.000%	9.000%	15.000%
$p_s \times r_s$	-2.000%	1.800%	2.700%	3.000%
$E(r)$	5.500%			
$r_s - E(r)$	-15.500%	0.500%	3.500%	9.500%
$(r_s - E(r))^2$	2.403%	0.003%	0.123%	0.903%
$p_s \times (r_s - E(r))^2$	0.481%	0.001%	0.037%	0.181%
$Var(r)$	0.699%			
$Std.dev(r)$	8.358%			

Now discuss the covariance and correlation coefficients between two returns r_i and r_j when both returns are discrete random variables. First, we need to define the state dependent return on asset i (j) in state s as r_{s_i} (r_{s_j}) and the joint probability of r_i and r_j in state s as p_{s_i,s_j}. Then, we can compute the covariance as

$$\sigma_{ij} = Cov(r_i, r_j) = \sum_{\forall s_i} \sum_{\forall s_j} p_{s_i,s_j} (r_{s_i} - E(r_i))(r_{s_j} - E(r_j)). \tag{3.11}$$

Table 3.2: Covariance and correlation between two random returns

States (s)			1	2	3	
	Returns (r_i)	r_1	-10%	-5%	10%	Marginal Prob.
	r_2					$p_{r_2,s}$
1	-5%		15%	2%	4%	21%
2	-1%		5%	30%	10%	45%
3	6%		7%	7%	20%	34%
	Marginal Prob.	$p_{r_1,s}$	27%	39%	34%	100%

Further, the correlation coefficient is given by

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j}. \tag{3.12}$$

Example:

We can then calculate means, variances, standard deviations, covariance and correlation between the two random returns as follows:

Table 3.3: Moments of two random returns

Moment	Value
$E(r_1) = \bar{r}_1$	-0.012500
$E(r_2) = \bar{r}_2$	0.005400
$Var(r_1) = \sigma_1^2$	0.006919
$Var(r_2) = \sigma_2^2$	0.001765
$Std.dev(r_1) = \sigma_1$	0.083179
$Std.dev(r_2) = \sigma_2$	0.042010
$Cov(r_1, r_2) = \sigma_{12}$	0.001338
$Corr(r_1, r_2) = \rho_{12} = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$	0.382761

In practice, we may not be able to observe either the density or probability functions of random variables. We can only use the sample data to estimate. Let $r_i = \{r_{i1}, r_{i1}, \ldots, r_{iT}\}$ for $i = 1, 2$ be the sample data of the return on asset i from time 1 to time T. Then we can compute the mean return and return variance for each asset and the covariance and the correlation between the returns on assets 1 and 2.

$$\bar{r}_i = \frac{1}{T} \sum_{t=1}^{T} r_{it}, \tag{3.13}$$

for $i = 1, 2$.

$$\hat{\sigma}_{ii} = \hat{\sigma}_i^2 = \frac{1}{T-1} \sum_{t=1}^{T} (r_{it} - \bar{r}_i)^2, \tag{3.14}$$

for $i = 1, 2$.

$$\hat{\sigma}_i = \sqrt{\hat{\sigma}_i^2} = \sqrt{\frac{1}{T-1} \sum_{t=1}^{T} (r_{it} - \bar{r}_i)^2}, \tag{3.15}$$

for $i = 1, 2$.

$$\hat{\sigma}_{12} = \frac{1}{T-1} \sum_{t=1}^{T} (r_{1t} - \bar{r}_1)(r_{2t} - \bar{r}_2). \tag{3.16}$$

$$\hat{\rho}_{12} = \frac{\hat{\sigma}_{12}}{\hat{\sigma}_1 \hat{\sigma}_2}. \tag{3.17}$$

3.3 Portfolio's Returns

- Let the portfolio p contains assets 1 and 2 with, respectively, weights w_1 and w_2, where $w_1 + w_2 = 1$.

- Let ρ_{12} be the correlation coefficient between the returns on assets 1 and 2.

$$-1 \leq \rho_{12} \leq 1 \tag{3.18}$$

- The mean return of the portfolio is:

$$\bar{r}_p = w_1 \bar{r}_1 + w_2 \bar{r}_2. \tag{3.19}$$

- We can do a thought-experiment to see, under which condition for rho_{12} among three scenarios, σ_p^2 is the lowest.

  The variance of returns of the portfolio is:

$$\sigma_p^2 = w_1^2 \sigma_{11} + w_2^2 \sigma_{22} + 2 w_1 w_2 \sigma_{12}. \tag{3.20}$$

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

$$\sigma_p^2 = w_1^2 \sigma_{11} + w_2^2 \sigma_{22} + 2 w_1 w_2 \rho_{12} \sigma_1 \sigma_2. \tag{3.21}$$

  It is apparent that when $rho_{1,2}$ σ_p^2 is the lowest among these three scenarios.

$$\begin{array}{rclcrcl}
\rho_{12} & = & 0 & \Rightarrow & \sigma_p^2 & = & w_1^2 \sigma_{11} + w_2^2 \sigma_{22} \\
\rho_{12} & = & +1 & \Rightarrow & \sigma_p^2 & = & w_1^2 \sigma_{11} + w_2^2 \sigma_{22} + 2 w_1 w_2 \sigma_1 \sigma_2 \\
 & & & & \sigma_p^2 & = & (w_1 \sigma_1 + w_2 \sigma_2)^2 \\
\rho_{12} & = & -1 & \Rightarrow & \sigma_p^2 & = & w_1^2 \sigma_{11} + w_2^2 \sigma_{22} - 2 w_1 w_2 \sigma_1 \sigma_2 \\
 & & & & \sigma_p^2 & = & (w_1 \sigma_1 - w_2 \sigma_2)^2
\end{array} \tag{3.22}$$

- Example: When $\bar{r}_1 = 10$, $\bar{r}_1 = 20$, $w_1 = 0.5$, and $w_2 = 0.5$,

$$\bar{r}_p = (0.5)(10) + (0.5)(20) = 15. \tag{3.23}$$

- Example: When $\rho_{12} = -1$, $\sigma_{11} = 4$, $\sigma_{22} = 9$, $w_1 = 0.5$, and $w_2 = 0.5$,

$$\begin{gathered}
\sigma_p^2 = (0.5)^2(4) + (0.5)^2(9) + 2(0.5)(0.5)(-1)(2)(3) \\
\sigma_p^2 = 1 + 2.25 - 3 \\
\sigma_p^2 = 0.25.
\end{gathered} \tag{3.24}$$

- Example: When $\bar{r}_1 = \bar{r}_2$, $\sigma_1 = \sigma_2$, and $\rho_{12} = 0$,we can find the minimum variance portfolio and its return variance and standard deviation. It is clear that $\bar{r}_p = \bar{r}_1 = \bar{r}_2$. Because $\rho_{12} = 0$, $\sigma_1 = \sigma_2$, and $w_2 = 1 - w_1$,

$$\begin{aligned} \sigma_p^2 &= [w_1^2 + (1-w_1)^2]\sigma_i^2 \qquad (3.25) \\ &= (2w_1^2 + 1 - 2w_1)\sigma_i^2, \end{aligned}$$

for $i = 1, 2$.

$$\frac{\partial \sigma_p^2}{\partial w_1} = (4w_1 - 2)\sigma_1^2 = 0. \qquad (3.26)$$

Solving w_1 from equation (3.26), we have

$$w_1^* = 1/2 \quad w_2^* = 1/2. \qquad (3.27)$$

Substituting the above solutions into equation (3.25) yields

$$\sigma_p^2 = \frac{\sigma_1^2}{2} = \frac{\sigma_2^2}{2} \qquad (3.28)$$

and

$$\sigma_p = \frac{\sigma_1}{\sqrt{2}} = \frac{\sigma_2}{\sqrt{2}}. \qquad (3.29)$$

Why do we see a smaller portfolio return variance (standard deviation) even if the portfolio is composed of two assets that have the identical mean return and identical variance (standard deviation)? Because two asset returns are not correlated, pooling them together will reduce the overall risk (return variance and standard deviation) of the portfolio.

- Example: When $\bar{r}_1 < \bar{r}_2$, $\sigma_1 = 0$, $\sigma_2 > 0$, and $\rho_{12} = 0$, we note that asset 2 is risk-free (e.g. T-bill) and, hence, its return variance is zero. We find the portfolios consisting of assets 1 and 2 in the $\bar{r}_p$–σ_p space. First, the mean return of the portfolio is given by

$$\bar{r}_p = w_1\bar{r}_1 + (1 - w_1)\bar{r}_2. \qquad (3.30)$$

Second, because asset 1 has no risk, the covariance between the returns of assets 1 and 2 is zero ($\sigma_{12} = 0$) and the correlation between the returns of assets 1 and 2 is, therefore, zero ($\rho_{12} = 0$). Third, the variance of the portfolio return is given by

$$\sigma_p^2 = (1 - w_1)^2\sigma_2^2. \qquad (3.31)$$

The standard deviation of the portfolio return is given by

$$\sigma_p = (1 - w_1)\sigma_2. \qquad (3.32)$$

Fourth, we note that when $w_1 = 1$ and $w_2 = 0$, then $\bar{r}_p = \bar{r}_1$ and $\sigma_p = 0$. When $w_1 = 0$ and $w_2 = 1$, then $\bar{r}_p = \bar{r}_2$ and $\sigma_p = \sigma_2$. When $0 < w_1 < 1$ and $0 < w_2 < 1$, $\bar{r}_p = w_1\bar{r}_1 + (1-w_1)\bar{r}_2$ and $\sigma_p = (1-w_1)\sigma_2$. Therefore, there is a linear relationship between $\bar{r}_p$ and σ_p. This linear relationship can be extended to the situations where $w_1 < 0$ as long as $w_1 + w_2 = 1$. An example is setting $w_1 = -1$, $w_2 = 2$.

- The Markowitz model

Assume that there are n different assets with their returns $\bar{r}_i$, $i = 1, 2, \ldots, n$ and their variances and covariances σ_{ij}, $i, j = 1, 2, \ldots, n$. We wish to find the efficient portfolio w_i, $i = 1, 2, \ldots, n$ by solving the following problem:

$$\min_{\{w_i\}} \quad \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} \tag{3.33}$$

subject to

$$\sum_{i=1}^{n} w_i \bar{r}_i = \bar{r}_p, \tag{3.34}$$

$$\sum_{i=1}^{n} w_i = 1. \tag{3.35}$$

To solve the above problem, we first form the Lagrangian function

$$L = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} + \lambda\left(\bar{r}_p - \sum_{i=1}^{n} w_i \bar{r}_i\right) + \mu\left(1 - \sum_{i=1}^{n} w_i\right), \tag{3.36}$$

where λ and μ are the Lagrange multipliers. Then, we differentiate the Lagrangian function L w.r.t. each w_i, λ, and μ and set these derivatives to zero, obtaining the $n+2$ first-order conditions:

$$\sum_{j=1}^{n} w_j \sigma_{ij} - \lambda \bar{r}_i - \mu = 0, \qquad i = 1, 2, \ldots, n \tag{3.37}$$

$$\sum_{i=1}^{n} w_i \bar{r}_i = \bar{r}_p, \tag{3.38}$$

and

$$\sum_{i=1}^{n} w_i = 1. \tag{3.39}$$

We solve n w_i's, λ and μ using these $n+2$ first-order conditions. The solutions w_i^*, $i = 1, 2, \ldots, n$, represent the efficient portfolio choice.

Remarks: Please note the efficient portfolios can be chosen for various values for $\bar{r}_p$. Hence, the Markowitz model can be used to find the efficient portfolio set which contains many efficient portfolios for all given values of $\bar{r}_p^{\prime}s$.

- Example: Let the number of assets be $n = 2$. Please find the optimal portfolio w_1^* and w_2^* which minimizes the portfolio variance σ_p^2 given that we choose the mean return for the portfolio $\bar{r}_p$. For this problem, the Lagrangian function is

$$L = \frac{1}{2}(w_1^2\sigma_{11} + 2w_1w_2\sigma_{12} + w_2^2\sigma_{22}) + \lambda(\bar{r}_p - w_1\bar{r}_1 - w_2\bar{r}_2) + \mu(1 - w_1 - w_2). \tag{3.40}$$

The two first-order conditions are:

$$\frac{\partial L}{\partial w_1} = w_1\sigma_{11} + w_2\sigma_{12} - \lambda\bar{r}_1 - \mu = 0 \tag{3.41}$$

and

$$\frac{\partial L}{\partial w_2} = w_1\sigma_{12} + w_2\sigma_{22} - \lambda\bar{r}_2 - \mu = 0. \tag{3.42}$$

The two constraints are:

$$w_1\bar{r}_1 + w_2\bar{r}_2 = \bar{r}_p \tag{3.43}$$

and

$$w_1 + w_2 = 1. \tag{3.44}$$

We can solve for w_1, w_2, λ and μ from the above four equations.
Subtracting equation (3.42) from equation (3.41) yields

$$w_1(\sigma_{11} - \sigma_{12}) + w_2(\sigma_{12} - \sigma_{22}) + \lambda(\bar{r}_2 - \bar{r}_1) = 0. \tag{3.45}$$

From equation (3.44), we obtain

$$w_2 = 1 - w_1. \tag{3.46}$$

Then, substituting equation (3.46) into equations (3.45) and (3.43) yields

$$\begin{aligned} w_1(\sigma_{11} - \sigma_{12}) + (1 - w_1)(\sigma_{12} - \sigma_{22}) + \lambda(\bar{r}_2 - \bar{r}_1) &= 0; \\ w_1\bar{r}_1 + (1 - w_1)\bar{r}_2 &= \bar{r}_p, \end{aligned} \tag{3.47}$$

which can be simplified into

$$\begin{aligned} w_1(\sigma_{11} - 2\sigma_{12} + \sigma_{22}) + \lambda(\bar{r}_2 - \bar{r}_1) &= (\sigma_{22} - \sigma_{12}); \\ w_1(\bar{r}_1 - \bar{r}_2) &= \bar{r}_p - \bar{r}_2. \end{aligned} \tag{3.48}$$

We solve for w_1 directly from equations (3.48):

$$w_1^* = \frac{\bar{r}_p - \bar{r}_2}{\bar{r}_1 - \bar{r}_2}. \tag{3.49}$$

Using equations (3.44) and (3.49) we can solve for w_2:

$$w_2^* = 1 - w_1^* = 1 - \frac{\bar{r}_p - \bar{r}_2}{\bar{r}_1 - \bar{r}_2} = \frac{\bar{r}_1 - \bar{r}_2 - \bar{r}_p + \bar{r}_2}{\bar{r}_1 - \bar{r}_2} = \frac{\bar{r}_1 - \bar{r}_p}{\bar{r}_1 - \bar{r}_2}. \tag{3.50}$$

We can verify that equations (3.49) and (3.50) are correct by verifying if the sum of w_1^* and w_2^* equals to 1:

$$w_1^* + w_2^* = \frac{\bar{r}_p - \bar{r}_2}{\bar{r}_1 - \bar{r}_2} + \frac{\bar{r}_1 - \bar{r}_p}{\bar{r}_1 - \bar{r}_2} = 1. \tag{3.51}$$

We can also solve for λ^*[1] and μ^*[2] but these are of secondary importance in our context.

[1]We can substitute equation (3.49) into equation (3.48) to solve for λ. That is,

$$\frac{(\bar{r}_p - \bar{r}_2)(\sigma_{11} - 2\sigma_{12} + \sigma_{22})}{\bar{r}_1 - \bar{r}_2} + \lambda(\bar{r}_2 - \bar{r}_1) = (\sigma_{22} - \sigma_{12}). \tag{3.52}$$

$$\lambda(\bar{r}_2 - \bar{r}_1) = (\sigma_{22} - \sigma_{12}) + \frac{(\bar{r}_p - \bar{r}_2)(\sigma_{11} - 2\sigma_{12} + \sigma_{22})}{\bar{r}_2 - \bar{r}_1}. \tag{3.53}$$

$$\lambda^* = \frac{(\sigma_{22} - \sigma_{12})}{(\bar{r}_2 - \bar{r}_1)} + \frac{(\bar{r}_p - \bar{r}_2)(\sigma_{11} - 2\sigma_{12} + \sigma_{22})}{(\bar{r}_2 - \bar{r}_1)^2}. \tag{3.54}$$

$$\lambda^* = \frac{(\bar{r}_p - \bar{r}_1)\sigma_{22} + (-2\bar{r}_p + \bar{r}_1 + \bar{r}_2)\sigma_{12} + (\bar{r}_p - \bar{r}_2)\sigma_{11}}{\bar{r}_2^2 - 2\bar{r}_1\bar{r}_2 + \bar{r}_1^2}. \tag{3.55}$$

$$\lambda^* = \frac{(\bar{r}_p - \bar{r}_1)\sigma_{22} + (-2\bar{r}_p + \bar{r}_1 + \bar{r}_2)\sigma_{12} + (\bar{r}_p - \bar{r}_2)\sigma_{11}}{(\bar{r}_2 - \bar{r}_1)^2}. \tag{3.56}$$

[2]Substituting equations (3.49), (3.50) and (3.54) into equation (3.41), we can solve for μ. That is,

- Minimization problem revisited

  Let $\mathbf{w}$ be an $n \times 1$ column vector of weights for n assets ($w_i, i = 1, 2, \ldots, n$), $\bar{\mathbf{r}}$ an $n \times 1$ column vector of mean returns for n assets ($r_i, i = 1, 2, \ldots, n$), $\mathbf{1}$ an $n \times 1$ vector of unity, and $\mathbf{\Sigma}$ an $n \times n$ variance-covariance matrix of returns for n assets ($\sigma_{ij},\ i = 1, 2, \ldots, n, j = 1, 2, \ldots, n$). Then the mean return of the portfolio can be written as

$$\mathbf{w}'\bar{\mathbf{r}} = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} \begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \\ \vdots \\ \bar{r}_n \end{bmatrix} = \sum_{i=1}^{n} w_i \bar{r}_i = \bar{r}_p. \tag{3.62}$$

  If $n = 2$, then

$$w_1 \bar{r}_1 + w_2 \bar{r}_2 = \bar{r}_p. \tag{3.63}$$

  The sum of the portfolio weights is equal to one:

$$\mathbf{w}'\mathbf{1} = 1. \tag{3.64}$$

  If $n = 2$, then

$$w_1 + w_2 = 1. \tag{3.65}$$

  The variance of returns of the portfolios can be written as

$$\sigma_p^2 = \mathbf{w}'\mathbf{\Sigma}\mathbf{w} \tag{3.66}$$

  or

$$\sigma_p^2 = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}. \tag{3.67}$$

  Note that $\sigma_{ij} = \sigma_{ji}$ for all i and j and, therefore, Σ is a symmetric matrix.

  If $n = 2$,

$$\sigma_p^2 = \begin{bmatrix} w_1 & w_2 \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}. \tag{3.68}$$

---

equation (3.41) becomes

$$\mu^* = -\left(\frac{\bar{r}_p - \bar{r}_2}{\bar{r}_2 - \bar{r}_1}\right)\sigma_{11} - \left(\frac{\bar{r}_1 - \bar{r}_p}{\bar{r}_2 - \bar{r}_1}\right)\sigma_{12} \tag{3.57}$$

$$- \left(\frac{(\bar{r}_p - \bar{r}_1)\sigma_{22} + (-2\bar{r}_p + \bar{r}_1 + \bar{r}_2)\sigma_{12} + (\bar{r}_p - \bar{r}_2)\sigma_{11}}{\bar{r}_2^2 - 2\bar{r}_1\bar{r}_2 + \bar{r}_1^2}\right)\bar{r}_1.$$

$$\mu^* = \left(\frac{-(\bar{r}_p - \bar{r}_2)(\bar{r}_2 - \bar{r}_1)}{(\bar{r}_2 - \bar{r}_1)^2}\right)\sigma_{11} + \left(\frac{-(\bar{r}_1 - \bar{r}_p)(\bar{r}_2 - \bar{r}_1)}{(\bar{r}_2 - \bar{r}_1)^2}\right)\sigma_{12} \tag{3.58}$$

$$- \left(\frac{(\bar{r}_p - \bar{r}_1)\sigma_{22} + (-2\bar{r}_p + \bar{r}_1 + \bar{r}_2)\sigma_{12} + (\bar{r}_p - \bar{r}_2)\sigma_{11}}{\bar{r}_2^2 - 2\bar{r}_1\bar{r}_2 + \bar{r}_1^2}\right)\bar{r}_1.$$

$$\mu^* = \left(\frac{(\bar{r}_1\bar{r}_p - \bar{r}_2\bar{r}_p - \bar{r}_1\bar{r}_2 + \bar{r}_2^2)\sigma_{11} + (\bar{r}_1^2 - \bar{r}_1\bar{r}_2 - \bar{r}_1\bar{r}_p + \bar{r}_2\bar{r}_p)\sigma_{12}}{\bar{r}_2^2 - 2\bar{r}_1\bar{r}_2 + \bar{r}_1^2}\right) \tag{3.59}$$

$$- \left(\frac{(\bar{r}_1\bar{r}_p - \bar{r}_1^2)\sigma_{22} + (-2\bar{r}_1\bar{r}_p + \bar{r}_1^2 + \bar{r}_1\bar{r}_2)\sigma_{12} + (\bar{r}_1\bar{r}_p - \bar{r}_1\bar{r}_2)\sigma_{11}}{\bar{r}_2^2 - 2\bar{r}_1\bar{r}_2 + \bar{r}_1^2}\right).$$

$$\mu^* = -\frac{(\bar{r}_1\bar{r}_p - \bar{r}_1^2)\sigma_{22} + ((-\bar{r}_2 - \bar{r}_1)\bar{r}_p + 2\bar{r}_1\bar{r}_2)\sigma_{12} + (\bar{r}_2\bar{r}_p - \bar{r}_2^2)\sigma_{11}}{\bar{r}_2^2 - 2\bar{r}_1\bar{r}_2 + \bar{r}_1^2}. \tag{3.60}$$

$$\mu^* = -\frac{(\bar{r}_1\bar{r}_p - \bar{r}_1^2)\sigma_{22} + ((-\bar{r}_2 - \bar{r}_1)\bar{r}_p + 2\bar{r}_1\bar{r}_2)\sigma_{12} + (\bar{r}_2\bar{r}_p - \bar{r}_2^2)\sigma_{11}}{(\bar{r}_2 - \bar{r}_1)^2}. \tag{3.61}$$

Because $\sigma_{12} = \sigma_{21}$ and $\sigma_{ii} = \sigma_i^2$, we can write

$$\sigma_p^2 = w_1^2\sigma_{11} + 2w_1w_2\sigma_{12} + w_2^2\sigma_{22}. \tag{3.69}$$

The above discussion shows that the compact representation is consistent with the previous two-asset case but it is more general. The Markowitz model can be expressed more compactly as

$$\min_{\mathbf{w}} \quad \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} \tag{3.70}$$

subject to

$$\mathbf{w}'\bar{\mathbf{r}} = \bar{r}_p, \tag{3.71}$$

$$\mathbf{w}'\mathbf{1} = 1. \tag{3.72}$$

Here, $\bar{\mathbf{r}}$ is an $n \times 1$ vector of n asset mean returns and $\mathbf{\Sigma}$ is an $n \times n$ variance-covariance matrix of n asset returns. $\bar{r}_p = \mathbf{w}'\bar{\mathbf{r}}$ is a specific mean return of the portfolio consisting of n assets. We usually specify the value of $\bar{r}_p$ in one constraint. $\sigma_p^2 = \mathbf{w}'\mathbf{\Sigma}\mathbf{w}$ is the return variance of the portfolio, which is to be minimized in the portfolio optimization problem. There is another constraint on the asset weights in the portfolio—the sum of portfolio weights equals to 1; i.e., $\mathbf{w}'\mathbf{1} = 1$. To solve this portfolio optimization problem, we set up the Lagrangian function as

$$L(\mathbf{w}, \lambda, \mu) = \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} + \lambda(\bar{r}_p - \mathbf{w}'\bar{\mathbf{r}}) + \mu(1 - \mathbf{w}'\mathbf{1}), \tag{3.73}$$

where λ and μ are the Lagrangian multipliers.

Recall $\frac{\partial \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w}}{\partial \mathbf{w}} = \mathbf{\Sigma}\mathbf{w}$, $\frac{\partial \mathbf{w}'\bar{\mathbf{r}}}{\partial \mathbf{w}} = \bar{\mathbf{r}}$, and $\frac{\partial \mathbf{w}'\mathbf{1}}{\partial \mathbf{w}} = \mathbf{1}$. To make sense of these derivatives, we let $n = 2$. Then we have

$$\frac{\partial \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w}}{\partial \mathbf{w}} = \mathbf{\Sigma}\mathbf{w} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}. \tag{3.74}$$

$$\frac{\partial \mathbf{w}'\bar{\mathbf{r}}}{\partial \mathbf{w}} = \bar{\mathbf{r}} = \begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \end{bmatrix}. \tag{3.75}$$

$$\frac{\partial \mathbf{w}'\mathbf{1}}{\partial \mathbf{w}} = \mathbf{1} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \tag{3.76}$$

We can solve the above portfolio optimization problem as follows. We obtain the fist-order conditions and set them to zero:

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{w}} &= \Sigma\mathbf{w}_p - \lambda\bar{\mathbf{r}} - \mu\mathbf{1} = \mathbf{0}, \\ \frac{\partial L}{\partial \lambda} &= \bar{r}_p - \mathbf{w}_p^{'}\bar{\mathbf{r}} = \mathbf{0}, \\ \frac{\partial L}{\partial \mu} &= 1 - \mathbf{w}_{\mathbf{p}}^{'}\mathbf{1} = \mathbf{0}. \end{aligned} \tag{3.77}$$

Please note that the optimal portfolio weights $\mathbf{w} = \mathbf{w}_p$ will satisfy the above equations and can be solved from these equations.

Since the objective function is a quadratic form, a unique solution exists. We can then use the first-order conditions to find $\mathbf{w}_p$. Multiplying the first first-order equation by $\mathbf{\Sigma^{-1}}$ and rearranging terms, we have

$$\mathbf{w_p} = \lambda(\mathbf{\Sigma}^{-1}\bar{\mathbf{r}}) + \mu(\mathbf{\Sigma}^{-1}\mathbf{1}). \tag{3.78}$$

Now we digress to discuss two observations:

(1) The constraint $\mathbf{w'1} = 1$ may be dropped from the Markowwitz model. If we drop this constraint by setting $\mu = 0$, the resulting raw portfolio weights may not be summed to 1; that is, $\mathbf{w}_p'\mathbf{1} \neq 1$. In this case, we can always normalize the raw portfolio weights so that we get the new portfolio weights $\mathbf{v}_p = \frac{\mathbf{w}_p}{\mathbf{w}_p'\mathbf{1}}$. In this case, $\mu = 0$ and we can set $\lambda = 1$ (we are going to normalize the raw portfolio weights anyway). Then equation (3.78) becomes

$$\mathbf{w}_p = \mathbf{\Sigma}^{-1}\bar{\mathbf{r}} \tag{3.79}$$

for the raw portfolio weights. We can normalize the raw portfolio weights using $\mathbf{v}_p = \frac{\mathbf{w}_p}{\mathbf{w}_p'\mathbf{1}}$.

(2) If we attempt to find the minimum variance portfolio regardless what the mean return of the portfolio may be, we can drop the constraint $\mathbf{w'\bar{r}} = \bar{r}_p$ by setting $\lambda = 0$. We can further set $\mu = 1$ (we are going to normalize the raw portfolio weights anyway). Then equation (3.78) becomes

$$\mathbf{w}_p = \mathbf{\Sigma}^{-1}\mathbf{1} \tag{3.80}$$

for the raw minimum variance portfolio weights. We can normalize the raw minimum variance portfolio weights using $\mathbf{v}_p = \frac{\mathbf{w}_p}{\mathbf{w}_p'\mathbf{1}}$.

Now we leave the digression assuming that we do not drop any of the two constraints. That is, we do not set $\lambda = 0$ and $\mu = 0$. In this case, substituting equation (3.78) in to the second and third first-order expressions in equations (3.77) yields

$$\begin{aligned} \lambda(\bar{\mathbf{r}}'\mathbf{\Sigma}^{-1}\bar{\mathbf{r}}) + \mu(\mathbf{1}'\mathbf{\Sigma}^{-1}\bar{\mathbf{r}}) &= \bar{r}_p, \\ \lambda(\bar{\mathbf{r}}'\mathbf{\Sigma}^{-1}\mathbf{1}) + \mu(\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}) &= 1. \end{aligned} \tag{3.81}$$

or

$$\begin{bmatrix} (\bar{\mathbf{r}}'\mathbf{\Sigma}^{-1}\bar{\mathbf{r}}) & (\mathbf{1}'\mathbf{\Sigma}^{-1}\bar{\mathbf{r}}) \\ (\bar{\mathbf{r}}'\mathbf{\Sigma}^{-1}\mathbf{1}) & (\mathbf{1}'\mathbf{\Sigma}^{-1}\mathbf{1}) \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{r}_p \\ 1 \end{bmatrix}. \tag{3.82}$$

The above equation system can be used to solve for λ and μ and then the values of λ and μ can be used to plug into the equation (3.78) for $\mathbf{w_p}$. Let

$$\begin{aligned} A &= (\mathbf{1}'\mathbf{\Sigma}^{-1}\bar{\mathbf{r}}) = (\bar{\mathbf{r}}'\mathbf{\Sigma}^{-1}\mathbf{1}) \\ B &= (\bar{\mathbf{r}}'\mathbf{\Sigma}^{-1}\bar{\mathbf{r}}) \\ C &= (\mathbf{1}'\Sigma^{-1}\mathbf{1}) \\ D &= BC - A^2 \end{aligned} \tag{3.83}$$

Now the system can be rewritten as

$$\begin{bmatrix} B & A \\ A & C \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{r}_p \\ 1 \end{bmatrix}.$$

Thus,

$$\lambda = \frac{C\bar{r}_p - A}{D} \tag{3.84}$$

and

$$\mu = \frac{B - A\bar{r}_p}{D}. \tag{3.85}$$

Substituting λ and μ into

$$\mathbf{w_p} = \lambda(\mathbf{\Sigma^{-1}\bar{r}}) + \mu(\mathbf{\Sigma}^{-1}\mathbf{1}), \tag{3.86}$$

we have

$$\mathbf{w_p} = \frac{C\bar{r}_p - A}{D}(\mathbf{\Sigma^{-1}\bar{r}}) + \frac{B - A\bar{r}_p}{D}(\mathbf{\Sigma^{-1}1}). \tag{3.87}$$

Rearranging terms, we have

$$\mathbf{w_p} = \frac{1}{D}[B(\mathbf{\Sigma^{-1}1}) - A(\mathbf{\Sigma}^{-1}\mathbf{\bar{r}})] + \frac{1}{D}[C(\mathbf{\Sigma^{-1}\bar{r}}) - A(\mathbf{\Sigma^{-1}1})]\bar{r}_p. \tag{3.88}$$

Note that $\mathbf{\Sigma}^{-1}$ is symmetric and positive definite. Thus, $B > 0, C > 0$, and $D > 0$. Let

$$g = \frac{1}{D}[B(\mathbf{\Sigma^{-1}1}) - A(\mathbf{\Sigma}^{-1}\mathbf{r})] \tag{3.89}$$

and

$$h = \frac{1}{D}[C(\mathbf{\Sigma^{-1}}\bar{r}) - A(\mathbf{\Sigma^{-1}1})]. \tag{3.90}$$

We have

$$\mathbf{w_p} = g + h\bar{r}_p. \tag{3.91}$$

- Dual problem

  The Markowitz model can be also expressed more compactly as

$$\max_{\mathbf{w}} \quad \mathbf{w'\bar{r}} \tag{3.92}$$

subject to

$$\begin{aligned} \mathbf{w'\Sigma w} &= \sigma_p^2, \\ \mathbf{w'1} &= 1. \end{aligned} \tag{3.93}$$

This maximization problem is a dual of the original minimization problem. If the minimum variance is set to σ_p^2 in this maximization problem corresponding to $\bar{r}_p$ specified in the original minimization problem, the solution of the maximization problem will be $\bar{r}_p$.

- Minimum variance portfolio

  To get the minimum variance portfolio, we solve the following problem.

$$\min_{\mathbf{w}} \quad \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} \tag{3.94}$$

  subject to

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

  The minimum variance portfolio is the solution $\mathbf{w}_p$ as given in equation (3.80).

- Tangent portfolio

  Let the risk-free rate be r_f. We can use the Sharpe ratio to measure the risk-adjusted return as

$$\frac{r_p - r_f}{\sigma_p} = \frac{\mathbf{w}'\bar{\mathbf{r}} - r_f}{\sqrt{\mathbf{w}'\mathbf{\Sigma}\mathbf{w}}}. \tag{3.95}$$

  The tangent portfolio can be found by solving the following maximization problem.

$$\max_{\mathbf{w}} \quad \frac{r_p - r_f}{\sigma_p} \tag{3.96}$$

  subject to

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

  The resulting portfolio $\mathbf{w}_p$ is the tangent portfolio.

  The tangent portfolio can also be found equivalently by solving the following maximization problem.

$$\max_{\mathbf{w}} \quad \mathbf{w}'(\bar{\mathbf{r}} - r_f\mathbf{1}) - \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} \tag{3.97}$$

  subject to

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

  The resulting portfolio $\mathbf{w}_p$ is the tangent portfolio. The Lagrangian function L is

$$L = \mathbf{w}'(\bar{\mathbf{r}} - r_f\mathbf{1}) - \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w} + \mu(1 - \mathbf{w}'\mathbf{1}). \tag{3.98}$$

  Given that the raw portfolio weights can be normalized anyway, we can simplify L by setting $\mu = 0$:

$$L = \mathbf{w}'(\bar{\mathbf{r}} - r_f\mathbf{1}) - \frac{1}{2}\mathbf{w}'\mathbf{\Sigma}\mathbf{w}. \tag{3.99}$$

  The first-order condition is given by:

$$\frac{\partial L}{\partial \mathbf{w}} = \bar{\mathbf{r}} - r_f\mathbf{1} - \mathbf{\Sigma}\mathbf{w}_p = \mathbf{0}. \tag{3.100}$$

This leads to the raw tangent portfolio weights:

$$\mathbf{w}_p = \mathbf{\Sigma}^{-1}\bar{\mathbf{r}} - r_f \mathbf{\Sigma}^{-1}\mathbf{1}, \tag{3.101}$$

where $\mathbf{\Sigma}^{-1}\bar{\mathbf{r}}$ is given as the raw optimal portfolio weights (before normalization) with the portfolio mean return $\bar{r}_p$ in equation (3.79) while $\mathbf{\Sigma}^{-1}\mathbf{1}$ is the raw minimum variance portfolio weights (before normalization) in equation (3.80). That is, the raw tangent portfolio weights can be calculated from the raw optimal portfolio weights with a specific portfolio mean return, the raw minimum variance portfolio weights, and the risk-free rate r_f. The raw tangent portfolio weights can be normalized using $\mathbf{v}_p = \frac{\mathbf{w}_p}{\mathbf{w}_p'\mathbf{1}}$.

3.4 The Efficient Frontier

- Feasible portfolio set

  By selecting a set of values for $\mathbf{w}' = [w_1, w_2, \ldots, w_n]$ (no short sales implies no negative weights) one can obtain a portfolio. All the possible portfolios form a feasible portfolio set. When a risk-free asset is added, one can form a new portfolio by combining the risk-free asset with any one of the feasible portfolios.

- The efficient set

  For a given mean return $\bar{r}_p = \bar{r}$ and $\mathbf{w}'\mathbf{1} = \sum_{i=1}^n w_i = 1$, select the portfolio which gives a least variance

  $$\sigma_p^2 = \min(\sigma_{p_1}^2, \sigma_{p_2}^2, \ldots, \sigma_{p_n}^2). \tag{3.102}$$

  Similarly, for a given portfolio variance $\sigma_p^2 = \bar{\sigma}^2$ and $\mathbf{w}'\mathbf{1} = \sum_{i=1}^n w_i = 1$, select the portfolio which gives a highest mean return

  $$\bar{r}_p = \max(\bar{r}_{p_1}, \bar{r}_{p_2}, \ldots, \bar{r}_{p_n}). \tag{3.103}$$

- The efficient set theorem: An investor will choose his or her optimal portfolio from the set of portfolios that (1) offer maximum mean return for varying levels of risk, and (2) offer minimum risk for varying levels of the mean return.

- Concavity of the efficient set

  $$\begin{array}{lclclcl} \rho_{12} & = & 0 & \Rightarrow & \sigma_p^2 & = & w_1^2\sigma_{11} + w_2^2\sigma_{22} \\ \rho_{12} & = & +1 & \Rightarrow & \sigma_p^2 & = & w_1^2\sigma_{11} + w_2^2\sigma_{22} + 2w_1w_2\sigma_1\sigma_2 \\ & & & & \sigma_p^2 & = & (w_1\sigma_1 + w_2\sigma_2)^2 \\ \rho_{12} & = & -1 & \Rightarrow & \sigma_p^2 & = & w_1^2\sigma_{11} + w_2^2\sigma_{22} - 2w_1w_2\sigma_1\sigma_2 \\ & & & & \sigma_p^2 & = & (w_1\sigma_1 - w_2\sigma_2)^2 \end{array} \tag{3.104}$$

  $\bar{r}_p$ is a linear function of w_i while σ_p is a quadratic function of w_i.

  Note: For two assets, one can also consider the cases when $\rho_{12} = 1$, $\rho_{12} = -1$ and $-1 < \rho_{12} < 1$.

- Graphical illustrations

3.5 Optimal Portfolios

- Preferences revisited
- Optimal portfolios with reference to the efficient portfolios and preferences
- Limitations

3.6 Diversification and Rebalancing

To see the role of diversification, we consider a specific case where there are asset A and asset B.

At time zero, both assets A and B are priced at $ 100. In period 1, asset A has a 100% return but asset B has a -50% return. In period 2, asset A has a -50% return but asset B has a 100% return.

We evaluate returns and risk over periods 1 and 2. The arithmetic mean of asset A's return over time is 25% [$= (100 - 50)/2$]. The variance of asset A's return is 11250 [$=[((100 - 25)^2 + (-50 - 25)^2)/(2 - 1)]$]. The standard deviation of asset A's return is 106.066. By the same calculation, the arithmetic mean of asset B's return over time is 25%. The variance of asset B's return is 11250. The standard deviation of asset B's return is 106.066. The correlation between asset A's return and asset B's return is -1.

Now we can look at the return of the 50-50 portfolio that invests 50% into asset A and 50% into asset B. Assume the initial investment at time 0 is $200, of which $100 is invested in asset A and $ 100 is invested in asset B. At the end of period 1, the price of asset A rises from $ 100 to $ 200 while the price of asset B falls from $ 100 to $ 50. The price of the 50-50 porfolio, therefore, rises from $ 200 to $ 250. The portfolio's return in period 1 is 25% [$= (200 + 50 - 200)/200$]. See Table 3.4 for details.

To maintain the 50-50 split, the portfolio manager will sell $75 ($= 200 - 250/2$) worth of asset A so that the dollar amount invested in asset A will fall from $ 200 to $125. The manager uses the $ 75 cash to buy additional asset B so that the dollar amount invested in asset B will rise from $ 50 to $ 125 ($= 50 + 75$). See Table 3.4 for details.

During period 2, asset A has a -50 % return so that the value of asset A drops from $ 125 to $ 62.50 while asset B has a 100 % return so that the value of asset B rises from $ 125 to $ 250. The price of the porfolio rises from $ 250 to $ 312.50. The portfolio's return during period 2 is also 25% ($= (250 + 62.50 - 250)/250$). See Table 3.4 for details.

If we assume that asset A and asset B maintain their return correlation (-1) and their return patterns (asset A's return is 100 % in even-numbered periods and -50 % in odd-numbered periods while asset B's return is -50 % in even-numbered periods and 100 % in odd-numbered periods) and that the portfolio manager will keep rebalancing their portfolio to the 50-50 split between asset A and asset B at the end of each period, this portfolio can maintain a 25% return per period into the future and, over time, this return will have zero variance and standard deviation. The portfolio performance and its sources of returns over time are described in Table 3.4.

This specific case shows the power of diversification and the power of portfolio rebalancing.

3.7 The Market Model

Let r_i be the return on asset i, r_I the return on market index I, α_{iI} the intercept term, β_{iI} the slope term, and ϵ_{iI} the random error term. The market model relating r_i and r_I

Table 3.4: Portfolio Diversification and Rebalancing

Period	0	Period 1	Rebalancing	Period 2	Rebalancing	Period 3	⋯
Asset A							
Price	\$ 100.00	\$ 200.00	\$ 125.00	\$ 62.50	\$ 156.25	\$ 312.50	⋯
Return		100.00 %		-50.00 %		100.00 %	⋯
Asset B							
Price	\$ 100.00	\$ 50.00	\$ 125.00	\$ 250.00	\$ 156.25	\$ 78.13	…
Return		-50.00%		100.00%		-50.00%	⋯
50-50 Porfolio							
Price	\$ 200.00	\$ 250.00	\$ 250.00	\$ 312.50	\$ 312.50	\$ 390.63	⋯
Return		25.00 %		25.00 %		25.00 %	⋯

is given by

$$r_i = \alpha_{iI} + \beta_{iI} r_I + \epsilon_{iI}. \tag{3.105}$$

Interpretations and illustrations

Beta: sensitivity to the market index

Diversification: The asset i has two different risk (1) market or systematic risk and (2) unique or unsystematic risk as illustrated in the variance of returns on the asset:

$$\sigma_i^2 = \beta_{iI}^2 \sigma_I^2 + \sigma_{\epsilon_{iI}}^2. \tag{3.106}$$

The portfolio return will be

$$r_p = \sum_{i=1}^{n} w_i r_i, \tag{3.107}$$

or

$$r_p = \sum_{i=1}^{n} w_i(\alpha_{iI} + \beta_{iI} r_I + \epsilon_{iI}), \tag{3.108}$$

or

$$r_p = \sum_{i=1}^{n} w_i \alpha_{iI} + \sum_{i=1}^{n} w_i \beta_{iI} r_I + \sum_{i=1}^{n} w_i \epsilon_{iI}, \tag{3.109}$$

or

$$r_p = \alpha_{pI} + \beta_{pI} r_I + \epsilon_{pI}. \tag{3.110}$$

If the random error components of asset returns are uncorrelated, then

$$\sigma_p^2 = \beta_{pI}^2 \sigma_I^2 + \sigma_{\epsilon_{pI}}^2. \tag{3.111}$$

Thus, the market risk can not be reduced while the unique risk ($\sum_{i=1}^{n} w_i^2 \sigma_{\epsilon_I}^2$) can through diversification. This is because as $n \to \infty$, $w_i^2 \to 0$ and $\sum_{i=1}^{n} w_i^2 \sigma_{\epsilon_I}^2 \to 0$.

Graphical illustration

3.8 Review Questions

1. If we have a cash flow from period 1 to period T, how to compute its present value?
2. If we have a cash flow from period 1 to period T and there are S possible states in time period, how to compute the present value of this cash flow?
3. What is the mean return of a asset? What is the mean return of a portfolio?
4. What is the return variance of a asset? What is the return variance of a portfolio?
5. What are the measures of risk in stock returns?
6. What is the covariance between the returns on two assets?
7. What is the correlation between the returns on two assets?
8. How to use the sample data to estimate the mean return, return variance, and the covariance and correlation between the returns on two assets?
9. Which portfolio has the lowest return variance: the portfolio consists of two asset returns that are (a) perfectly positively correlated, (b) perfectly negatively correlated, or (c) completely uncorrelated?
10. What is a feasible set of portfolios for given n stocks? Please use a figure to show the feasible set?
11. What is the minimum variance portfolio? Please show it in a figure.
12. How to find the efficient portfolio frontier? What are the two methods for achieving the efficient portfolio frontier given the feasible set of portfolios? Please show the frontier in a figure.
13. Why is the efficient portfolio frontier concave?
14. What is the two fund separation?
15. You are given the mean returns and return variances of two risky assets and the covariance between returns of the two risky assets, could you find the portfolio consisting of these two risky assets with the mean portfolio return $\bar{r}_p$ and minimum portfolio variance σ_p?
16. What are the two types of risk in asset returns according to the market model? Please explain why diversification cannot eliminate risk completely and what type of risk will remain in a large well diversified portfolio.

3.9 Review on Mean-Variance Optimization

3.9.1 General Setup

When we attempt to find the optimal portfolio, $w_1, w_2, \ldots, w_n$, based on the mean-variance criterion, we set up the follow optimization problem:

$$\min_{w_1,w_2,\ldots,w_n} \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} \tag{3.112}$$

$$\begin{aligned} s.t. \quad \textstyle\sum_{i=1}^{n} w_i \bar{r}_i = &\quad \bar{r}_p, & (3.113)\\ \textstyle\sum_{i=1}^{n} w_i = &\quad 1. & (3.114)\end{aligned}$$

To solve the optimization problem, we form the Lagragian function

$$L = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_{ij} + \lambda(\bar{r}_p - \sum_{i=1}^{n} w_i \bar{r}_i) + \mu(1 - \sum_{i=1}^{n} w_i). \tag{3.115}$$

Taking the derivatives of L w.r.t. w_k, $k = 1, 2, \ldots, n$, and setting these derivatives to zero yield the first-order conditions:

$$\sum_{j=1}^{n} w_j \sigma_{kj} - \lambda \bar{r}_k - \mu = 0, \tag{3.116}$$

where $k = 1, 2, \ldots, n$.

Taking the derivative of L w.r.t. λ and setting the derivative to zero yield

$$\bar{r}_p = \sum_{i=1}^{n} w_i \bar{r}_i. \tag{3.117}$$

Taking the derivative of L w.r.t. μ and setting the derivative to zero yield

$$1 = \sum_{i=1}^{n} w_i. \tag{3.118}$$

To find the optimal portfolio weights, $w_1, w_2, \ldots, w_n$, we simplify the first-order conditions by relaxing the constraint in equation (3.118) (setting $\mu = 0$). That is, we can replace the true portfolio weights w_i with the raw portfolio weights v_i for all $i = 1, 2, \ldots, n$. These raw portfolio weights $v_1, v_2, \ldots, v_n$ do not have to sum up to 1. We can always normalize the raw portfolio weights to obtain the true portfolio weights by using the following formula:

$$w_i = \frac{v_i}{\sum_{i=1}^{n} v_i}, \tag{3.119}$$

for $i = 1, 2, \ldots, n$. With this relaxation, the first order conditions in equation (3.116) become

$$\sum_{j=1}^{n} v_j \sigma_{kj} = \bar{r}_k, \tag{3.120}$$

where $k = 1, 2, \ldots, n$ and $\lambda = 1$.

3.9.2 Minimum Variance Portfolio with No Restriction on Portfolio Return $\bar{r}_p$

Now we attempt to find the minimum variance portfolio without imposing any restriction on the portfolio mean return $\bar{r}_p$. Effectively, in the first-order conditions shown in equation (3.116), we have relaxed the first constraint $\sum_{i=1}^{n} w_i \bar{r}_i = \bar{r}_p$ (setting $\lambda = 0$). We could arbitrarily set $\mu = 1$ given that we can always normalize the raw portfolio weights. With w_i being replaced by v_i, Equation (3.116) becomes

$$\sum_{j=1}^{n} v_j \sigma_{kj} = 1, \tag{3.121}$$

where $k = 1, 2, \ldots, n$. If we know σ_{ij} for $i, j = 1, 2, \ldots, n$ or $\mathbf{\Sigma}$, we can use equation (3.121) to solve for $v_1, v_2, \ldots, v_n$ and normalize them into $w_1, w_2, \ldots, w_n$ using equation (3.119).

We can show how to use equation (3.121) to solve for $w_1, w_2, \ldots, w_n$ using the following numerical example. When $\gamma = 0$ and $\mu = 1$,

$$\bar{\mathbf{r}} = \begin{bmatrix} \bar{r}_1 \\ \bar{r}_2 \\ \bar{r}_3 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix},$$

$$\mathbf{\Sigma} = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_{22} & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_{33} \end{bmatrix} = \begin{bmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix},$$

$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix},$$

and

$$\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

We can write equation (3.121) as

$$\mathbf{\Sigma v} = \mathbf{1}. \tag{3.122}$$

To solve for $\mathbf{v}$, we need $\mathbf{\Sigma}^{-1}$ so that the solution $\mathbf{v}^1$ is

$$\mathbf{v}^1 = \mathbf{\Sigma}^{-1}\mathbf{1}. \tag{3.123}$$

$$= \begin{bmatrix} \frac{3}{4} & -\frac{1}{2} & \frac{1}{4} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & \frac{3}{4} \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 \\ 0.0 \\ 0.5 \end{bmatrix}.$$

Because the raw portfolio weights are added up to 1; $\sum_{i=1}^{3} v_i^1 = 1$. They are the true portfolio weights as well.

$$\mathbf{w}^1 = \frac{\mathbf{v}^1}{\mathbf{v}^{1\prime}\mathbf{1}} = \begin{bmatrix} 0.50/1.00 \\ 0.00/1.00 \\ 0.50/1.00 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.0 \\ 0.5 \end{bmatrix}.$$

It is also possible to check if v_1^1, v_2^1, v_3^1 satisfy equation (3.122):

$$\begin{aligned} 2v_1^1 + v_2^1 + 0 &= 1.0 + 0.0 + 0.0 = 1.0 \\ v_1^1 + 2v_2^1 + v_3^1 &= 0.5 + 0.0 + 0.5 = 1.0 \\ 0 + v_2^1 + 2v_3^1 &= 0.0 + 0.0 + 1.0 = 1.0 \end{aligned}$$

Therefore, the true minimum variance portfolio is $\{w_1^1, w_2^1, w_3^1\} = \{0.5, 0.0, 0.5\}$.

3.9.3 Minimum Variance Portfolio with a Restriction on Portfolio Return $\bar{r}_p$

Now we wish to find the minimum variance portfolio that has the portfolio mean return of $\bar{r}_p$. Here we do not remove the first constraint $\bar{r}_p = \sum_i w_i \bar{r}_i$ in the first-order conditions in equation (3.116) but we can arbitrarily set $\lambda = 1$. As in the previous example, we continue to relax the second constraint $1 = \sum_{i=1} w_i$ or set $\mu = 0$. That is, we replace the true portfolio weights w_i with the raw portfolio weights v_i, which do not have to sum to 1. With these changes, the first-order conditions in equation (3.116) change to

$$\sum_{j=1}^{n} v_j \sigma_{kj} = \bar{r}_k, \tag{3.124}$$

where $k = 1, 2, \ldots, n$. We can write equation (3.124) as

$$\mathbf{\Sigma v} = \bar{\mathbf{r}}. \tag{3.125}$$

We still use the previous numerical example. To solve for $\mathbf{v}$, we need $\mathbf{\Sigma}^{-1}$ so that the solution $\mathbf{v}^2$ is

$$\mathbf{v}^2 = \mathbf{\Sigma}^{-1}\bar{\mathbf{r}}. \tag{3.126}$$

$$\begin{aligned} \mathbf{v}^2 &= \begin{bmatrix} \frac{3}{4} & -\frac{1}{2} & \frac{1}{4} \\ -\frac{1}{2} & 1 & -\frac{1}{2} \\ \frac{1}{4} & -\frac{1}{2} & \frac{3}{4} \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \end{bmatrix} \\ &= \begin{bmatrix} 1.0 \\ 0.0 \\ 2.0 \end{bmatrix}. \end{aligned}$$

Because $\sum_i v_i^2 = 3$,

$$\mathbf{w}^2 = \frac{\mathbf{v}^2}{\mathbf{v}^{2\prime}\mathbf{1}} = \begin{bmatrix} 1.00/3.00 \\ 0.00/3.00 \\ 2.00/3.00 \end{bmatrix} = \begin{bmatrix} 0.33 \\ 0.00 \\ 0.66 \end{bmatrix}.$$

It is also possible to check if v_1^2, v_2^2, v_3^2 satisfy equation (3.125):

$$\begin{aligned}
2v_1^2 + v_2^2 + 0 &= 2.0 + 0.0 + 0.0 = 2.0 \\
v_1^2 + 2v_2^2 + v_3^2 &= 1.0 + 0.0 + 2.0 = 3.0 \\
0 + v_2^2 + 2v_3^2 &= 0.0 + 0.0 + 4.0 = 4.0
\end{aligned}$$

Therefore, the minimum variance portfolio with the specific mean portfolio return $\bar{r}_p = 2/3 + 4(2)/3 = 10/3 = 3.33$ is $\{w_1^2, w_2^2, w_3^2\} = \{0.33, 0.0, 0.66\}$.

3.9.4 Efficient Portfolio

Under the assumptions that the investor is insatiable and risk-averse and that there are n risky assets and one risk-free asset with the return r_f, the investor always wish to form a portfolio $w_1, w_2, \ldots, w_n$ that maximizes the excess return over the risk-free rate $\bar{r}_p - r_f$ and minimizes the portfolio risk σ_p.

We can formulate the optimization problem of this insatiable and risk-averse investor in, at least, the following two ways:

First, we can imagine that you can draw a straight line starting from r_f on the vertical axis in the $\bar{r}_p$-σ_p coordinate system and going through the tangent point on the efficient portfolio frontier. The angle θ between this straight line and the horizontal line both starting from r_f. The optimal portfolio maximizes θ. The tangent function of θ, $\tan\theta$, is defined as the "rise" $(\bar{r}_p - r_f)$ over "run" (σ_p). Therefore, maximizing θ is equivalent to maximizing $\tan\theta$. We formulate the optimization as

$$\max_{\{w_1, w_2, \ldots, w_n\}} \tan\theta, \tag{3.127}$$

where $\tan\theta = \frac{\sum_i w_i(\bar{r}_i - r_f)}{\left(\sum_{i,j} w_i w_j \sigma_{ij}\right)^{1/2}}$ and $\sum_i w_i = 1$. We can rewrite the above as

$$\max_{\{w_1, w_2, \ldots, w_n\}} \frac{\sum_i w_i(\bar{r}_i - r_f)}{\left(\sum_{i,j} w_i w_j \sigma_{ij}\right)^{1/2}}. \tag{3.128}$$

Now find the first-order conditions of equation (3.128) noting $(x^n)' = nx^{n-1}$ and $\{\frac{u}{v}\}' = \frac{u'v - v'u}{v^2}$.

$$\begin{aligned}
\frac{\partial \tan\theta}{\partial w_k} &= \frac{(\bar{r}_k - r_f)(\sum_{i,j} w_i w_j \sigma_{ij})^{1/2} - (\sum_i w_i(\bar{r}_i - r_f))\frac{\partial(\sum_{i,j} w_i w_j \sigma_{ij})^{1/2}}{\partial w_k}}{(\sum_{i,j} w_i w_j \sigma_{ij})} \\
&= \frac{(\bar{r}_k - r_f)(\sum_{i,j} w_i w_j \sigma_{ij})^{1/2} - (\sum_i w_i(\bar{r}_i - r_f))\frac{1}{2}(\sum_{i,j} w_i w_j \sigma_{ij})^{-1/2}\sum_j \sigma_{kj} w_j}{(\sum_{i,j} w_i w_j \sigma_{ij})} \\
&= \frac{(\bar{r}_k - r_f)}{(\sum_{i,j} w_i w_j \sigma_{ij})^{1/2}} - \frac{(\sum_i w_i(\bar{r}_i - r_f))\sum_j \sigma_{kj} w_j}{2(\sum_{i,j} w_i w_j \sigma_{ij})^{3/2}} \\
&= 0.
\end{aligned} \tag{3.129}$$

Simplifying the above first-order condition, we get

$$\bar{r}_k - r_f = \delta \sum_j \sigma_{kj} w_i, \tag{3.130}$$

where

$$\delta = \left(\frac{\sum_i w_i(\bar{r}_i - r_f)}{2(\sum_{i,j} w_i w_j \sigma_{ij})} \right)$$

is a positive unknown constant. Because δ is a positive constant, we can set $\delta = 1$ and write the above as

$$\bar{r}_k - r_f = \sum_j \sigma_{kj} v_i, \tag{3.131}$$

where the raw portfolio weights are functions of the true portfolio weights; that is, $v_i = \delta w_i$, for $i = 1, 2, \ldots, n$. When we obtain all raw portfolio weights $v_1, v_2, \ldots, v_n$, we can always use equation (3.119) to obtain the true portfolio weights $w_1, w_2, \ldots, w_n$. We can write equation (3.131) as

$$\bar{\mathbf{r}} - r_f \mathbf{1} = \mathbf{\Sigma v}. \tag{3.132}$$

Second, we can formulate the maximization of $\tan\theta$ as the simultaneous minimization of $\bar{r}_p - r_f$ and minimizing of σ_p^2. That is,

$$\max_{\mathbf{w}} \quad \mathbf{v}'(\bar{\mathbf{r}} - r_f\mathbf{1}) - \frac{1}{2}\mathbf{v}'\mathbf{\Sigma v}. \tag{3.133}$$

The first-order condition is given by:

$$\frac{\partial L}{\partial \mathbf{v}} = \bar{\mathbf{r}} - r_f\mathbf{1} - \mathbf{\Sigma v}_p = \mathbf{0}. \tag{3.134}$$

This leads to the raw tangent portfolio weights:

$$\mathbf{v}_p = \mathbf{\Sigma}^{-1}(\bar{\mathbf{r}} - r_f\mathbf{1}). \tag{3.135}$$

We will continue to use the previous numerical example with $r_f = 1$. To find the optimal $\mathbf{v}$, we use

$$\mathbf{v}^3 = \mathbf{\Sigma}^{-1}(\bar{\mathbf{r}} - r_f\mathbf{1}) = \mathbf{v}^2 - r_f\mathbf{v}^1. \tag{3.136}$$

$$\mathbf{v}^3 = \begin{bmatrix} 1.0 \\ 0.0 \\ 2.0 \end{bmatrix} - 1 \times \begin{bmatrix} 0.5 \\ 0.0 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.0 \\ 1.5 \end{bmatrix}.$$

Because $\sum_i v_i^3 = 2$,

$$\mathbf{w}^3 = \frac{\mathbf{v}^3}{\mathbf{v}^{3\prime}\mathbf{1}} = \begin{bmatrix} 0.50/2 \\ 0.00/2 \\ 1.50/2 \end{bmatrix} = \begin{bmatrix} 0.25 \\ 0.00 \\ 0.75 \end{bmatrix}.$$

It is also possible to check if v_1^3, v_2^3, v_3^3 satisfy equation (3.132):

$$\begin{aligned} \bar{r}_1 - r_f = 2v_1^3 + v_2^3 + 0 &= 1.0 + 0.0 + 0.0 = 1.0 \\ \bar{r}_1 - r_f = v_1^3 + 2v_2^3 + v_3^3 &= 0.5 + 0.0 + 1.5 = 2.0 \\ \bar{r}_3 - r_f = 0 + v_2^3 + 2v_3^3 &= 0.0 + 0.0 + 3.0 = 3.0 \end{aligned}$$

Therefore, the minimum variance portfolio is $\{w_1^3, w_2^3, w_3^3\} = \{0.25, 0.0, 0.75\}$.

3.9.5 The relationship between $\mathbf{v}^3$ and $\mathbf{v}^1$ & $\mathbf{v}^2$

We note that equation (3.131) can be written as

$$\boldsymbol{\Sigma}^{-1}\bar{\mathbf{r}} - r_f\boldsymbol{\Sigma}^{-1}\mathbf{1} = \mathbf{v}^3, \tag{3.137}$$

where $\boldsymbol{\Sigma}^{-1}\mathbf{1} = \mathbf{v}^1$ according to equation (3.123) and $\boldsymbol{\Sigma}^{-1}\bar{\mathbf{r}} = \mathbf{v}^2$ according to equation (3.126). Hence, we have

$$\mathbf{v}^3 = \mathbf{v}^2 - r_f\mathbf{v}^1. \tag{3.138}$$

Now we can use the previous example to verify equation (3.138).

$$\mathbf{v}^3 = \mathbf{v}^2 - r_f\mathbf{v}^1 = \begin{bmatrix} 1.0 \\ 0.0 \\ 2.0 \end{bmatrix} - 1 \times \begin{bmatrix} 0.5 \\ 0.0 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 0.0 \\ 1.5 \end{bmatrix}. \tag{3.139}$$

Chapter 4

Asset Pricing Models

4.1 Learning Objectives

- The Capital Market Line
- Components of risk
- The Security Market Line
- Beta and the market model revisited
- Factor models
- Arbitrage Pricing Theory

4.2 The Capital Market Line

To understand how assets or securities are priced, we need to construct a model or propose a theory. A model or a theory must be simple and abstract to some degree. Some assumptions are needed for simplifications.

About the assumption, Milton Friedman said in his famous essay (Essays in the Theory of Positive Economics, Chicago: The University of Chicago Press, 1953):

> [T]he relevant question to ask about the "assumptions" of a theory is not whether they are descriptively "realistic," for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields accurate predictions.

Assumptions:

1. Investors evaluate portfolios by looking at the expected returns and standard deviation of the portfolios over a one-period horizon.
2. Investors are non-satiable.
3. Investors are risk-averse.
4. Individual assets are infinitely divisible.
5. There is a risk-free rate for borrowing or lending.
6. There are no taxes and transaction costs.
7. All investors have the identical investment horizon.
8. The risk-free rate is the same for all investors.
9. Information is freely and instantly available to all investors.
10. Investors have homogeneous expectations.

Assumptions 1–6 are used for both the Capital Asset Pricing Model (CAPM) and the portfolio theory while Assumptions 7–10 are used for the CAPM only.

One of the features of the CAPM is that each investor would spread his or her funds among risky securities in the same relative proportions, adding risk-free borrowing or lending to varying degrees according to individuals' preferences. This feature is captured by the separation theorem.

The separation theorem:

> The optimal combination of risky assets for an investor can be determined without any knowledge about the investor's preferences toward risk and return.

That is, the determination of the optimal combination of risky assets can be made separately from the determination of the shape of an investor's indifference curves. The reason for such a theorem to hold is that the linear efficient set represents a combination of the tangency portfolio and varying degrees of risk-free borrowing or lending.

With the CAPM, all individual investors are identical and faces the same efficient set. Therefore their risky portion of their portfolios are the same.

Another important feature of the CAPM is that, in equilibrium, each security must have a non-zero proportion in the composition of the tangency portfolio.

The argument for this is: when a security has a zero proportion its prices falls and hence its expected return will be higher. The higher expected return of this security attracts investors to purchase it and thus the security obtains a non-zero proportion in the composition of the tangency portfolio.

How about the demand for a security is too large for the shares outstanding? A large demand will drive the price high and the expected return low so that the demand will fall.

Eventually, everything will balance out and the market will reach equilibrium. In the equilibrium, the market prices and holding volumes of securities will not move away from the state. The borrowers and lenders finalize their borrowing and lending at the risk-free rate. Thus, the proportions of the tangency portfolio will correspond to the proportions of the market portfolio.

The market portfolio is a portfolio consisting of an investment in all securities where the proportion to be invested in each security corresponds to its relative market value. The relative market value of a security is simply equal to the aggregate market value of the security divided by the sum of the aggregate market of all securities. The market portfolio plays an important role in the CAPM.

Let M be the market portfolio and its return r_M. Let r_f be the risk-free rate. In the return $(\bar{r}_p)$–risk(σ_p) space of the portfolio (p), we can draw a straight line from r_f to M $(\bar{r}_p,\sigma_p)$. This linear efficient set of the CAPM is called the Capital Market Line (CML).

The slope of the CML is equal to the difference between expected returns of the market portfolio and that of the risk-free rate ($\bar{r}_M - r_f$) divided by the difference in their standard deviations $(\sigma_M - 0)$. The vertical intercept is r_f. Therefore, the expected return of an efficient portfolio p $(\bar{r}_p)$ can be computed as a function of the standard deviation of the portfolio (σ_p) by

$$\bar{r}_p = r_f + \left(\frac{\bar{r}_M - r_f}{\sigma_M}\right)\sigma_p. \tag{4.1}$$

4.3 Components of Risk

Recall the standard deviation of the efficient portfolio can be computed by

$$\sigma_p = \sqrt{\sigma_p^2} = \sqrt{\mathbf{w}'\mathbf{\Sigma}\mathbf{w}} \tag{4.2}$$

or

$$\sigma_p = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_{ij} \right)^{\frac{1}{2}} \tag{4.3}$$

where $\mathbf{w}$ is an $n \times 1$ column vector of weights $(w_i, i = 1, 2, \ldots, n)$ and $\mathbf{\Sigma}$ is an $n \times n$ variance-covariance matrix of returns of n securities $(\sigma_{ij}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, n)$.

Consider the standard deviation of the market portfolio:

$$\sigma_M = \sqrt{\sigma_M^2} = \sqrt{\mathbf{w}_M' \mathbf{\Sigma}_M \mathbf{w}_M} \tag{4.4}$$

or

$$\sigma_M = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} w_{iM} w_{jM} \sigma_{ij} \right)^{\frac{1}{2}} \tag{4.5}$$

where $\mathbf{w}_M$ is an $n \times 1$ column vector of security weights of n securities in the market portfolio $(w_{iM}, i = 1, 2, \ldots, n)$ and $\mathbf{\Sigma}_M$ is an $n \times n$ variance-covariance matrix of returns of n securities $(\sigma_{ij}, i = 1, 2, \ldots, n, j = 1, 2, \ldots, n)$ in the market securities.

It is also possible to express the standard deviation of the market portfolio in the following way

$$\sigma_M = \left(w_{1M} \sum_{j=1}^{n} w_{jM} \sigma_{1j} + w_{2M} \sum_{j=1}^{n} w_{jM} \sigma_{2j} + \cdots + w_{nM} \sum_{j=1}^{n} w_{jM} \sigma_{nj} \right)^{\frac{1}{2}} \tag{4.6}$$

where each term $\sum_{j=1}^{n} w_{jM} \sigma_{ij}$ $(i = 1, 2, \ldots, n)$ can be viewed as the covariance of returns of security i and the market portfolio; that is

$$\sum_{j=1}^{n} w_{jM} \sigma_{ij} = \sigma_{iM}, i = 1, 2, \ldots, n. \tag{4.7}$$

Thus, the standard deviation of the market portfolio is

$$\sigma_M = \left(w_{1M} \sigma_{1M} + w_{2M} \sigma_{2M} + \cdots + w_{nM} \sigma_{nM} \right)^{\frac{1}{2}}. \tag{4.8}$$

Thus, the standard deviation of the market portfolio is equal to the squared root of a weighted average of the covariances of all the securities within it with the weights being the proportions of the respective securities in the market portfolio.

As can be seen from the above, the relevant measure of risk for a security is its covariance with the market portfolio; that is, $\sigma_{iM}, i = 1, 2, \ldots, n$.

To show how to use the above expressions, assume that there are two securities in the market portfolio, A and B. Their random returns are r_A and r_B, respectively. The mean returns of A and B are $\bar{r}_A$ and $\bar{r}_B$, respectively. The variances of A's return and B's return are σ_A^2 and σ_B^2, respectively. The covariance between A's return and B's return is σ_{AB}. Now let the weight of A (B) in the market portfolio be w_A (w_B). Assume that $w_A + w_B = 1$. What are the mean and variance of the market portfolio's return, r_M? They are, respectively,

$$\bar{r}_M = w_A \bar{r}_A + w_B \bar{r}_B \tag{4.9}$$

and

$$\sigma_M = (w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2 w_A w_B \sigma_{AB})^{\frac{1}{2}}. \tag{4.10}$$

The terms on the right-hand-side of the variance equation can be regrouped according to the shares of contributions from A and B, respectively. That is,

$$\begin{aligned} \sigma_M^2 &= w_A^2 \sigma_A^2 + w_B^2 \sigma_B^2 + 2 w_A w_B \sigma_{AB} \\ &= w_A (w_A \sigma_A^2 + w_B \sigma_{AB}) + w_B (w_B \sigma_B^2 + w_A \sigma_{AB}) \\ &= w_A \sigma_{AM} + w_B \sigma_{BM}, \end{aligned} \tag{4.11}$$

where $\sigma_{AM} = w_A \sigma_A^2 + w_B \sigma_{AB}$ and $\sigma_{AM} = w_B \sigma_B^2 + w_A \sigma_{AB}$. Intuitively, the covariance between security A's (B's) return, r_A (r_B), and the market portfolio return, r_M, includes the weighted variance of security A's (B's) return as well as the weighted covariance between security A's (B's) return and the other security in the market portfolio.

4.4 The Security Market Line

Now we can explain the Security Market Line (SML). The SML describes the exact form of the equilibrium relationship between risk and return of a security, say security i:

$$\bar{r}_i = r_f + \left(\frac{\bar{r}_M - r_f}{\sigma_M^2} \right) \sigma_{iM}. \tag{4.12}$$

Graphically, the SML can be shown in the return ($\bar{r}_i$)-risk (σ_{iM}) space of security i: the vertical intercept is r_f. The slope of the straight line is $\frac{\bar{r}_M - r_f}{\sigma_M^2}$. The market portfolio M is located at $(\bar{r}_M, \sigma_M^2)$. When $\sigma_{iM} = 0$, $\bar{r}_i = r_f$.

To show that slope is correct. Let w_i be the weight of security i in portfolio p and $1 - w_i$ the weight of the market portfolio M in portfolio p. The expected return of portfolio p is

$$\bar{r}_p = w_i \bar{r}_i + (1 - w_i) \bar{r}_M. \tag{4.13}$$

The standard deviation of portfolio p is

$$\sigma_p = (w_i^2 \sigma_i^2 + (1 - w_i)^2 \sigma_M^2 + 2 w_i (1 - w_i) \sigma_{iM})^{\frac{1}{2}}. \tag{4.14}$$

In the return ($\bar{r}_p$)-risk (σ_p) space, the curved line linking the market portfolio M (tangent portfolio) and security i in the feasible portfolio set represents all possible portfolios satisfying the above two equations.

To understand the slope of the SML, it is necessary to figure out the slope of the curved line at a given point. It is known that this slope can be written as

$$\frac{d\bar{r}_p}{d\sigma_p} = \frac{d\bar{r}_p / dw_i}{d\sigma_p / dw_i}. \tag{4.15}$$

Now we can find $d\bar{r}_p / dw_i$ and $d\sigma_p / dw_i$ to compute $\frac{d\bar{r}_p}{d\sigma_p}$.

$$\frac{d\bar{r}_p}{dw_i} = \bar{r}_i - \bar{r}_M \tag{4.16}$$

and

$$\frac{d\sigma_p}{dw_i} = \frac{w_i\sigma_i^2 - \sigma_M^2 + w_i\sigma_M^2 + \sigma_{iM} - 2w_i\sigma_{iM}}{(w_i^2\sigma_i^2 + (1-w_i)^2\sigma_M^2 + 2w_i(1-w_i)\sigma_{iM})^{\frac{1}{2}}}. \tag{4.17}$$

Thus

$$\frac{d\bar{r}_p}{d\sigma_p} = \frac{(\bar{r}_i - \bar{r}_M)(w_i^2\sigma_i^2 + (1-w_i)^2\sigma_M^2 + 2w_i(1-w_i)\sigma_{iM})^{\frac{1}{2}}}{w_i\sigma_i^2 - \sigma_M^2 + w_i\sigma_M^2 + \sigma_{iM} - 2w_i\sigma_{iM}}. \tag{4.18}$$

The slope of the curved line linking the market portfolio M and security i at the endpoint M can be found by letting $w_i = 0$; that is

$$\frac{d\bar{r}_p}{d\sigma_p} = \frac{(\bar{r}_i - \bar{r}_M)\sigma_M}{\sigma_{iM} - \sigma_M^2}. \tag{4.19}$$

At the point M, the slope of the CML, $(\bar{r}_M - r_f)/\sigma_M$, must equal the slope of the curved line linking M and i. Thus, we have

$$\frac{(\bar{r}_i - \bar{r}_M)\sigma_M}{\sigma_{iM} - \sigma_M^2} = \frac{(\bar{r}_M - r_f)}{\sigma_M}. \tag{4.20}$$

Solving for $\bar{r}_i$ results the SML:

$$\bar{r}_i = r_f + \left(\frac{\bar{r}_M - r_f}{\sigma_M^2}\right)\sigma_{iM}. \tag{4.21}$$

Please note that in the above form of the SML the risk premium is normalized or standardized $(\frac{\bar{r}_M - r_f}{\sigma_M^2})$ but the level of risk is not (σ_{iM}).

According to the above expression, the SML can be graphed in the return $(\bar{r}_i)$-risk (σ_{iM}) space with the vertical intercept being r_f. The market portfolio M is located at $(\bar{r}_M, \sigma_M^2)$ or $(\bar{r}_M, \sigma_{MM})$.

Another way to express the SML is

$$\bar{r}_i = r_f + (\bar{r}_M - r_f)\,\frac{\sigma_{iM}}{\sigma_M^2} = r_f + (\bar{r}_M - r_f)\,\beta_i, \tag{4.22}$$

where $\beta_i = \frac{\sigma_{iM}}{\sigma_M^2}$ is called the beta coefficient or beta for security i. Please note that in the above form of the SML the level of risk is normalized or standardized $(\frac{\sigma_{iM}}{\sigma_M^2})$ but the risk premium is not $(\bar{r}_M - r_f)$.

According to the above expression, the SML can be graphed in the return $(\bar{r}_i)$-risk (β_i) space with the vertical intercept being r_f. The market portfolio M is located at $(\bar{r}_M, 1)$, where $\beta_i = \frac{\sigma_{MM}}{\sigma_M^2} = \frac{\sigma_M^2}{\sigma_M^2}$.

In practice, financial analysts often use the variant of equation (4.22) to find out the excess return, α_i, and the normalized security risk in the market portfolio, β_i,. They estimate the empirical version of the security market line for security i using the data:

$$r_i - r_f = \alpha_i + \beta_i(r_M - r_f) + e_i, \tag{4.23}$$

where $r_i - r_f$ is the excess return of security i, $r_M - r_f$ is the excess return of the market portfolio, and e_i is the random error. Theoretically, α_i should be zero. However, when security i has an additional excess return, α_i will be positive. Searching for Alpha is, therefore, what investors are working on.

One may wish to know the consistency between the CML and SML. It would be interesting to explore this from a slightly different angle. Let replace $\overline{r}_i$ with $\overline{r}_p$ and σ_{iM} with σ_{pM} in the SML to get

$$\bar{r}_p = r_f + \left(\frac{\bar{r}_M - r_f}{\sigma_M^2} \right) \sigma_{pM}, \tag{4.24}$$

which must be consistent with

$$\bar{r}_p = r_f + \left(\frac{\bar{r}_M - r_f}{\sigma_M} \right) \sigma_p. \tag{4.25}$$

We wish to verify if it is the case. The above two equations imply that

$$\frac{\sigma_{pM}}{\sigma_M^2} = \frac{\sigma_p}{\sigma_M}, \tag{4.26}$$

which further implies that

$$\frac{\sigma_{pM}}{\sigma_p \sigma_M} = \rho_{pM} = 1. \tag{4.27}$$

That is, given the existence of the risk-free rate r_f and the expected market portfolio return $\overline{r}_M$, the two fund separation result will hold—the portfolio p will be a linear combination of the risk-free asset and the market portfolio. The expected return and risk of the portfolio p will be linearly related to the expected return and risk of the market portfolio. Let the portfolio p contains $1 - \alpha$ % of the total asset value in the risk-free asset and α % in the market portfolio. Then it can be shown that

$$\overline{r}_p = (1 - \alpha) r_f + \alpha \overline{r}_M \tag{4.28}$$

and

$$\sigma_p = \alpha \sigma_M. \tag{4.29}$$

The last equation is because that the variance of the risk-free asset return is zero and that the covariance between the risk-free asset return and market portfolio return is zero. That is,

$$Var(r_f) = 0 \tag{4.30}$$

and

$$Cov(r_f, r_M) = 0. \tag{4.31}$$

With the above knowledge, now we can explain why $\frac{\sigma_{pM}}{\sigma_p \sigma_M} = \rho_{pM} = 1$. Rewrite

$$\rho_{pM} = \frac{\sigma_{pM}}{\sigma_p \sigma_M} = \frac{Cov(r_p, r_M)}{[Var(r_p)]^{\frac{1}{2}} \sigma_M} = \frac{E(r_p - \overline{r}_p)(r_M - \overline{r}_M)}{[E(r_p - \overline{r}_p)^2]^{\frac{1}{2}} \sigma_M}. \tag{4.32}$$

If the portfolio p contains α % of the total asset into the market portfolio, then

$$\frac{E(r_p - \overline{r}_p)(r_M - \overline{r}_M)}{[E(r_p - \overline{r}_p)^2]^{\frac{1}{2}} \sigma_M} = \frac{E((1-\alpha) r_f + \alpha r_M - (1-\alpha) r_f - \alpha \overline{r}_M)(r_M - \overline{r}_M)}{[E((1-\alpha) r_f + \alpha r_M - (1-\alpha) r_f - \alpha \overline{r}_M)^2]^{\frac{1}{2}} \sigma_M}, \tag{4.33}$$

which can be further simplified into

$$\frac{\alpha E(r_M - \overline{r}_M)(r_M - \overline{r}_M)}{[\alpha^2 E(r_M - \overline{r}_M)^2]^{\frac{1}{2}} \sigma_M} = \frac{\alpha \sigma_M^2}{\alpha \sigma_M^2} = 1 = \rho_{pM}. \tag{4.34}$$

The above analysis shows that the consistency between the SML and CML from a new angle.

4.5 Beta and the Market Model Revisited

In the previous section, we have discussed the beta for security i. Now let us examine the beta for portfolio p. Recall the SML is

$$\bar{r}_i = r_f + (\bar{r}_M - r_f)\,\beta_i. \tag{4.35}$$

The expected return of portfolio p is

$$\bar{r}_p = \sum_{i=1}^{n} w_i \bar{r}_i. \tag{4.36}$$

Substituting the equation for $\bar{r}_i$ into the equation for $\bar{r}_p$ yields:

$$\bar{r}_p = \sum_{i=1}^{n} w_i \left(r_f + (\bar{r}_M - r_f)\,\beta_i\right) = r_f + (\bar{r}_M - r_f) \sum_{i=1}^{n} w_i \beta_i. \tag{4.37}$$

Thus, the beta for portfolio p is

$$\beta_p = \sum_{i=1}^{n} w_i \beta_i. \tag{4.38}$$

The the SML for portfolio p is

$$\bar{r}_p = r_f + (\bar{r}_M - r_f)\,\beta_p. \tag{4.39}$$

It is important to note that the beta appears in both the SML

$$\bar{r}_i = r_f + (\bar{r}_M - r_f)\beta_i \tag{4.40}$$

or

$$\bar{r}_i = r_f(1 - \beta_i) + \bar{r}_M \beta_i \tag{4.41}$$

and the market model

$$r_i = \alpha_{iI} + \beta_{iI} r_I + \epsilon_{iI}. \tag{4.42}$$

The differences between the two models are:

1. The market model is a factor model. More specifically, it is a single-factor model with the market index being the factor while the CAPM is an equilibrium model.
2. The market model uses a market index while the CAPM uses the market portfolio.
3. The market model captures the randomness of the return while the CAPM describes the expected return.

However both models offer a way for decomposing the security risk into the market risk and non-market risk:

$$\sigma_i^2 = \beta_{iI}^2 \sigma_I^2 + \sigma_{\epsilon_i}^2 \tag{4.43}$$

from the market model and

$$\sigma_i^2 = \beta_{iM}^2 \sigma_M^2 + \sigma_{\epsilon_i}^2 \tag{4.44}$$

from the underlying process of the CAPM. In each of the above variance decomposition, the first term on the right-hand side of equation represents the market (index or portfolio) risk and the second term represents the non-market risk.

According to the CAPM, the investor is awarded by bearing the market risk but not the non-market risk in equilibrium. Note that this is an equilibrium result. In reality, the investor may be awarded or punished by bearing non-market risk depending on the realization of the underlying return.

It should be noted that there are some debates as to whether or not the CAPM is a good model (Roll's critique).

4.6 Factor Models

The objective of modern portfolio theory is to help the investor to identify his or her optimal portfolio. Using the information on expected return and standard deviation of each security and the covariance between each and every pair of securities, one may be able to compute the curved efficient set—the efficient portfolio frontier.

The proposition of factor models makes the identification of the curved efficient set much more easier. Under necessary assumptions, the return process can be described by a factor model which relates returns on securities to the movement in one or more common factors. Sometimes the common factors are common indices. Thus factor models are also called index models.

Any aspect of a security's return unexplained by the factor model is assumed to be unique to the security and therefore uncorrelated with the unique element of returns on other securities. Thus, we can consider a linear regression model as a factor model.

The general form of the single-factor model can be used to illustrate the properties of the factor model in general. Let F be the value of the factor. The return of security i, r_i, can be viewed as a function of F :

$$r_i = a_i + b_i F + e_{i.} \tag{4.45}$$

The term a_i is a fixed part of the return that is not affected by F. The term b_i measures the sensitivity of the return of security i to F. The random term e_i captures the random component of the return of security i. The mean of e_i is zero and the standard deviation is σ_{ei}. If security i is insensitive to F ($b_i = 0$) or if $F = 0$, the return of security i is $r_i = a_i + e_{i.}$

The mean return of security i can be written as

$$\bar{r}_i = a_i + b_i \bar{F}. \tag{4.46}$$

Let the variance of F be σ_F^2. The variance of security i can be expressed as

$$\sigma_i^2 = b_i^2 \sigma_F^2 + \sigma_{ei}^2. \tag{4.47}$$

When securities i and j are considered, it should noted that e_i and e_j are assumed to be uncorrelated and that F and e_i (e_j) are assumed to be uncorrelated. The covariance of returns of securities i and j are

$$\sigma_{ij} = b_i b_j \sigma_F^2. \tag{4.48}$$

The market model is a specific example of a factor model where the factor is the return on a market index. Recall the model can be written as

$$r_i = \alpha_{iI} + \beta_{iI} r_I + \epsilon_{iI}. \tag{4.49}$$

Two important features of the one-factor model are:

(1) The assumption that the returns on securities respond to common factors greatly simplifies the task of calculating the curved Markowitz efficient set. More specifically, we can compute all the expected returns ($\bar{r}_i$'s), variances (σ_i^2's), and covariances (σ_{ij}'s) by estimating a_i, b_i, σ_{ei}^2, $\bar{F}$ and σ_F^2. Using the values of $\bar{r}_i$'s, σ_i^2's, and σ_{ij}'s, we can compute the curved Markowitz efficient set of portfolios.

(2) The sensitivity of a portfolio to a factor is the weighted average of the sensitivities of the component securities, with the securities' proportions in the portfolio serving as weights. By using the one-factor model, the variance of a portfolio p is given by

$$\sigma_p^2 = b_p^2 \sigma_F^2 + \sigma_{ep}^2, \tag{4.50}$$

where

$$b_p = \sum_{i=1}^{n} w_i b_i \tag{4.51}$$

and

$$\sigma_{ep}^2 = \sum_{i=1}^{n} w_i^2 \sigma_{ei}^2. \tag{4.52}$$

This expression shows that the total risk can be decomposed into the factor risk and the non-factor risk of the portfolio. This expression also shows that diversification leads to an averaging of factor risk ($b_p = \sum_{i=1}^{n} w_i b_i$) and that diversification leads to a decreasing non-factor risk ($\sigma_{ep}^2 = \sum_{i=1}^{n} w_i^2 \sigma_{ei}^2$).

The idea of the single-factor model can be extended to the case for the multiple-factor model. The factors considered could be the growth rate of GDP, the level of interest rates, the inflation rate, the level of oil prices and so on. This extension can be illustrated by the two-factor model.

Let F_1 and F_2 be two factors and b_{i1} and b_{i2} be the sensitivities of security i to these two factors. Let e_i be an random error term and a_i be the intercept term. The two-factor model is

$$r_i = a_i + b_{i1} F_1 + b_{i2} F_2 + e_i. \tag{4.53}$$

The expected return of security i is

$$\bar{r}_i = a_i + b_{i1} \bar{F}_1 + b_{i2} \bar{F}_2. \tag{4.54}$$

The variance of returns of security i is

$$\sigma_i^2 = b_{i1}^2 \sigma_{F1}^2 + b_{i2}^2 \sigma_{F2}^2 + 2 b_{i1} b_{i2} COV(F_1, F_2) + \sigma_{ei}^2, \tag{4.55}$$

where $COV(F_1, F_2)$ is the covariance between F_1 and F_2. The covariance of returns between securities i and j is

$$\sigma_{ij} = b_{i1} b_{j1} \sigma_{F1}^2 + b_{i2} b_{j2} \sigma_{F2}^2 + (b_{i1} b_{j2} + b_{i2} b_{j1}) COV(F_1, F_2). \tag{4.56}$$

Using the above expected returns, variances, and covariances, we can computed the curved Markowitz efficient set of portfolio with great ease.

We can also show (1) that diversification leads to an averaging of factor risk and (2) that diversification can substantially reduce non-factor risk. These conclusions can be explained as follows:

The return of portfolio p is given by

$$r_p = \sum_{i=1}^{n} w_i r_i. \tag{4.57}$$

Since security i has a return following the two-factor model, the above expression can be written as

$$r_p = \sum_{i=1}^{n} w_i (a_i + b_{i1} F_1 + b_{i2} F_2 + e_i), \tag{4.58}$$

which can be simplified into

$$r_p = \sum_{i=1}^{n} w_i a_i + \sum_{i=1}^{n} w_i b_{i1} F_1 + \sum_{i=1}^{n} w_i b_{i2} F_2 + \sum_{i=1}^{n} w_i e_i. \tag{4.59}$$

Let $a_p = \sum_{i=1}^{n} w_i a_i$, $b_{p1} = \sum_{i=1}^{n} w_i b_{i1}$, $b_{p2} = \sum_{i=1}^{n} w_i b_{i2}$, and $e_p = \sum_{i=1}^{n} w_i e_i$. The above model can be written as a two-factor model for the portfolio:

$$r_p = a_p + b_{p1}F_1 + b_{p2}F_2 + e_p. \tag{4.60}$$

Hence,

$$\sigma_p^2 = b_{p1}^2\sigma_{F_1}^2 + b_{p2}^2\sigma_{F_2}^2 + 2b_{p1}b_{p2}COV(F_1, F_2) + \sigma_{ep}^2 \tag{4.61}$$

where $\sigma_{ep}^2 = \sum_{i=1}^{n} w_i^2\sigma_{ei}^2$. As $n \to \infty$, $w_i^2 \to 0$ and $\sigma_{ep}^2 \to 0$.

The single- and two-factor models can be generalized even further into the k-factor model:

$$r_i = a_i + b_{i1}F_{1t} + b_{i2}F_2 + \cdots + b_{ik}F_k + e_i. \tag{4.62}$$

The k-factor model, although general, does not give us any guidance as to how many factors, and what factors, should be included. In practice, economists/financial analysts often fit the sector specific models. That is, the factors used for mining industry are different from the factors used for airline industry.

The factor models are often used in empirical research. The factor models can be estimated by using the time-series approach, the cross-sectional approach, and the factor-analytic approach.

It should be noted that a factor model is not an equilibrium model of asset prices as is the CAPM. However, if equilibrium exists, certain relationships will hold between the factor model and the equilibrium asset-pricing model.

4.7 Arbitrage Pricing Theory

Arbitrage Pricing Theory (APT) is an equilibrium model of security prices as is the CAPM. But APT makes fewer assumptions about investor preferences than does the CAPM.

One primary APT assumption is that each investor will seize every opportunity to increase the return of his or her portfolio so that no arbitrage opportunity will exist in equilibrium. Another primary APT assumption is that security returns are generated by a factor model.

The principle of arbitrage can be explained as follows: Arbitrage is the earning of riskless profit by taking advantage of differential pricing for the same physical asset or security. Arbitrage, as a widely-used investment tactic, typically entails the sale of a security at a relatively high price and the simultaneous purchase of the same security (or its functional equivalent) at a relatively low price.

This principle can be explained as follows. Assume that we have three securities and we know their expected returns and betas:

i	$\bar{r}_i$	b_i
1	15%	.9
2	21%	3.0
3	12%	1.8

(4.63)

We know that the investor wants to form an arbitrage portfolio so that he or she can increase the expected return of the portfolio substantially without increasing its risk. There are three requirements for an arbitrage portfolio to exist: (1) An arbitrage portfolio does not require any additional funds from the investor. This requirement can be written as

$$w_1 + w_2 + w_3 = 0. \tag{4.64}$$

The above expression can hold when an arbitrage portfolio includes long and short positions in securities. (2) An arbitrage portfolio must have no sensitivity to any factor. This requirement can be written as

$$w_1 b_1 + w_2 b_2 + w_3 b_3 = 0. \tag{4.65}$$

(3) An arbitrage portfolio must have a positive expected return. That is

$$w_1 \bar{r}_1 + w_2 \bar{r}_2 + w_3 \bar{r}_3 > 0 \tag{4.66}$$

Now identify an arbitrage portfolio using the information given to securities 1, 2, and 3. Let $w_1 = .1$, then we have

$$.1 + w_2 + w_3 = 0 \tag{4.67}$$

and

$$.1(.9) + w_2(3.0) + w_3(1.8) = 0. \tag{4.68}$$

Solving the above two equations yields $w_2 = .075$ and $w_3 = -0.175$. Now we check if an arbitrage portfolio has an arbitrage profit so that

$$w_1(15) + w_2(21) + w_3(12) > 0. \tag{4.69}$$

That is, $(.1 \times 15\%) + (.075 \times 21\%) + (-.175 \times 12\%) = .975\% > 0$. Indeed, we have found an arbitrage portfolio.

What is the consequence of buying securities 1 and 2 and selling security 3? Since everyone is doing this, their market prices will be affected. The prices of securities 1 and 2 will rise but their expected return will fall. The price of security 3 will fall but its expected returns will rise. Recall that the expected return can be expressed as

$$\bar{r} = \frac{\bar{P}_1}{P_0} - 1 \tag{4.70}$$

where $\bar{P}_1$ is the security's expected price at the end of the period and P_0 is the current price. This expression shows that the current price is inversely related to the expected return.

Investors will invest in arbitrage portfolios, driving up the prices of the securities held in long positions and driving down the prices of securities held in short positions, until all arbitrage possibilities are eliminated. When all arbitrage possibilities are eliminated, the equilibrium-expected return on a security will be a linear function of its sensitivities to the factors. Using the single-factor model the expected return of security i can be written as

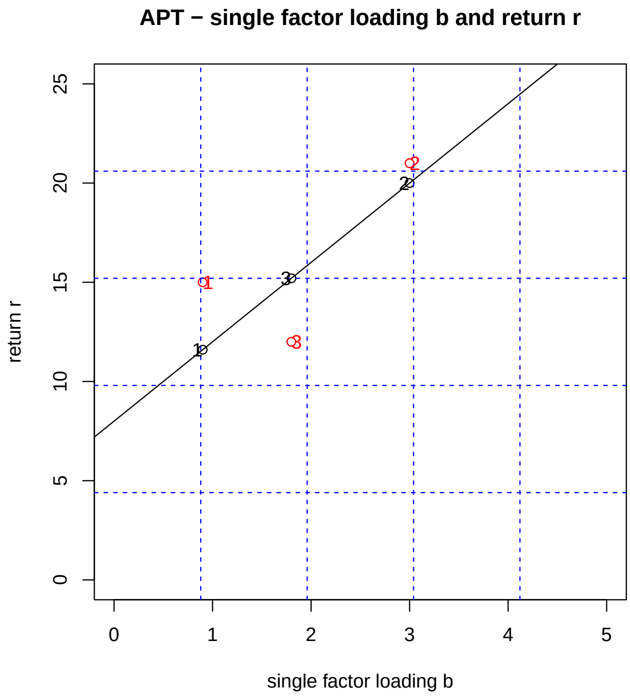
$$\bar{r}_i = \lambda_0 + \lambda_1 b_i \tag{4.71}$$

where λ_0 and λ_1 are constants. This is the asset pricing equation of the APT when returns are generated by one factor. In our example, $\lambda_0 = 8$ and $\lambda_1 = 4$ so that

$$\begin{aligned} \bar{r}_1 &= 8 + 4 \times .9 &= 11.6\% \\ \bar{r}_2 &= 8 + 4 \times 3.0 &= 20.0\% \\ \bar{r}_3 &= 8 + 4 \times 1.8 &= 15.2\% \end{aligned} \tag{4.72}$$

In equilibrium, the expected returns of securities 1 and 2 will fall from 15% and 21 %, respectively, to 11.6% and 20.0%, because of increased buying pressure. The expected return of security 3 will rise from 12% to 15.2%.

Figure 4.1: APT - single factor model



We can plot the data of the returns with and without arbitrage opportunities and the single factor model loadings to show the APT model as follows.

Note that if the reader can appreciate matrix algebra and the linear space. The non-arbitrage argument can lead to the APT model as follows: a non-arbitrage portfolio must satisfy (1) $\mathbf{w}'\mathbf{1} = 0 \Rightarrow \mathbf{w} \perp \mathbf{1}$ (no money down), (2) $\mathbf{w}'\mathbf{b} = 0 \Rightarrow \mathbf{w} \perp \mathbf{b}$ (no risk), and (3) $\mathbf{w}'\bar{\mathbf{r}} = 0 \Rightarrow \mathbf{w} \perp \bar{\mathbf{r}}$ (no profit) where $\mathbf{w} = [w_1, w_2]$, $\mathbf{1} = [1, 1]$, $\mathbf{b} = [b_1, b_2]$, and $\bar{\mathbf{r}} = [\bar{r}_1, \bar{r}_2]$. Thus, the vector of expected returns, $\bar{\mathbf{r}}$, must be a linear combination of $\mathbf{1}$ and $\mathbf{b}$:

$$\bar{\mathbf{r}} = \lambda_0 \mathbf{1} + \lambda_1 \mathbf{b} \tag{4.73}$$

or

$$\begin{aligned} \bar{r}_1 &= \lambda_0 + \lambda_1 b_1 \\ \bar{r}_2 &= \lambda_0 + \lambda_1 b_2 \end{aligned} \tag{4.74}$$

How can we interpret λ_0 and λ_1? In the world in which there is a risk-free asset, its return will be a constant and hence is not sensitive to the factor. The expected return of this risk-free asset can be expressed as $\bar{r}_i = \lambda_0$. More specifically we can name this return $r_f = \lambda_0$. Now we can consider another risk asset whose expected return is determined by

$$\bar{r}_i = r_f + \lambda_1 b_i. \tag{4.75}$$

Now we want to interpret λ_1. Assume we know the expected return of a portfolio that has the unit sensitivity to the factor; that is, the portfolio has $b_i = 1$. This portfolio is often called a pure factor portfolio denoted by p^*. The expected return of p^* is

$$\bar{r}_{p^*} = r_f + \lambda_1. \tag{4.76}$$

By rearranging the above equation, we have

$$\lambda_1 = \bar{r}_{p^*} - r_f. \tag{4.77}$$

Thus, λ_1 is interpreted as the risk premium on a portfolio that has unit sensitivity to the factor. It is known as a factor risk premium. Let $\delta = \bar{r}_{p^*}$. Thus, the APT pricing equation for the one-factor model is

$$\bar{r}_i = r_f + (\delta - r_f)b_i. \tag{4.78}$$

The above idea can be generalized to the multi-factor models. Let k factors be $F_1, F_2, \ldots, F_k$. Each security has k sensitivities $b_1, b_2, \ldots, b_k$. The k-factor model is

$$r_{it} = a_i + b_{i1}F_{1t} + b_{i2}F_{2t} + \cdots + b_{ik}F_{kt} + e_{it}. \tag{4.79}$$

We can drop the subscript t from the above equation for simplicity.

$$r_i = a_i + b_{i1}F_1 + b_{i2}F_2 + \cdots + b_{ik}F_k + e_i. \tag{4.80}$$

Using the approach we have learned, we can find the expected return of security i is a function of $\lambda_0, \lambda_1, \ldots, \lambda_k$ and $b_{i1}, b_{i2}, \ldots, b_{ik}$.

$$\bar{r}_i = \lambda_0 + \lambda_1 b_{1i} + \lambda_2 b_{2i} + \cdots + \lambda_k b_{ki}. \tag{4.81}$$

As discussed before we can define a factor-risk premium is the equilibrium return over the risk-free rate expected to be generated by a portfolio with a unit sensitivity to the factor and no sensitivity to any other factor. That is, $\lambda_0 = r_f$, $\lambda_1 = \delta_1 - r_f, \lambda_2 = \delta_2 - r_f, \ldots, \lambda_k = \delta_k - r_f$. Note δ_i is the a pure factor portfolio that has a unit sensitivity to factor i and zero sensitivity to the rest of the k factors. Thus, the k-factor APT pricing equation is

$$\bar{r}_i = r_f + (\delta_1 - r_f)b_{1i} + (\delta_2 - r_f)b_{2i} + \cdots + (\delta_k - r_f)b_{ki}. \tag{4.82}$$

APT and the CAPM are not necessarily inconsistent with each other. Now we can discuss the relationship between the CAPM and the single-factor APT models:

$$\bar{r}_i = r_f + (\bar{r}_M - r_f)\beta_{iM} \tag{4.83}$$

and

$$\bar{r}_i = r_f + (\delta - r_f)b_i. \tag{4.84}$$

Note that $COV(r_i, r_M) = COV(a_i + b_iF_1 + e_i, r_M)$. This can be simplified into

$$COV(r_i, r_M) = b_i COV(F_1, r_M) + COV(e_i, r_M). \tag{4.85}$$

The second term on the right-hand-side of the above equation is close to zero. Hence

$$COV(r_i, r_M) \simeq b_i COV(F_1, r_M). \tag{4.86}$$

Substituting the above into

$$\beta_{iM} = \frac{COV(r_i, r_M)}{\sigma_M^2} \tag{4.87}$$

yields

$$\beta_{iM} \simeq \frac{COV(F_1, r_M)}{\sigma_M^2} b_i. \tag{4.88}$$

Hence, if security returns are generated by a factor model and the CAPM holds, then a security's beta will depend on the security's sensitivity to the factor and the covariances between the factor and the market portfolio.

Finally, APT does not specify the number or identity of the factors that affect expected returns. Most research into factors focuses on indicators of aggregate economic activity, inflation, and interest rates.

4.8 Review Questions

1. What is the fund separation theorem?
2. What is the Capital Market Line (CML)?
3. Please explain why the market portfolio risk (in terms of the return variance) is the weighted sum of the covariances between individual security returns and the market portfolio return.
4. What is the Security Market Line (SML)? How is the SML related to the CML?
5. What are the two forms of the SML?
6. Can you establish the SML for a portfolio?
7. What is the Roll's critique? (you may need to do some research.)
8. What is a factor model? What can you learn from a factor model?
9. What are the likely factors for a typical factor model in finance? (you may need to do some research.)
10. What are the three conditions of no arbitrage? Or what is an arbitrage opportunity?
11. What is the Arbitrage Pricing Theory (APT)?
12. The risk-free interest rate is 8%. The risk premium on the market risk factor is 3%. The factor sensitivities (or factor loadings) of stocks A, B and C are 1, 2 and 3, respectively. The returns of stocks A, B and C are 11%, 15% and 17%, respectively. If there is one mispriced stock according to the APT, which is the stock? Should you buy more of this stock or short more of this stock?

Chapter 5

Bond Analysis

5.1 Learning Objectives

- Money market instruments
- Government bonds
- Corporate bonds
- Debt instruments trading
- Valuation of riskless securities
- Risk premium and term premium
- The term structure of interest rates
- Bond attributes and determinations of yield spreads
- Bond market efficiency
- Bond-pricing theorems
- Convexity
- Duration
- Active management

5.2 Money Market Instruments

Highly marketable, short-term securities are referred to as money market instruments, which include Treasury bills, finance and commercial paper, large-denomination certificates of deposit, bankers' acceptances, certificates of deposits (CDs), Eurodollars CDs, Eurodollar deposits, and so on.

Treasury bills are issued on a discount basis by the Federal governments with maturities of up to 52 weeks.

Finance paper is an unsecured short-term promissory note issued by sales finance and consumer loan companies. Commercial paper is similar to finance paper but is issued by commercial and industrial corporations.

A bankers' acceptance is a pure-discount security that is a substitute for commercial paper as a source of financing.

Certificates of deposit are the short-term instruments issued by the chartered banks.

Eurodollar CDs are large short-term CDs denominated in dollars and issued by banks outside North America. Eurodollars are dollar-denominated time deposits in banks outside North America.

5.3 Government Bonds

The central bank issues debt securities to finance the government's borrowing needs. These include Treasury bills and bonds.

We have mentioned Treasury bills. Treasury bonds could be short-term (less than 3 year maturities), medium-term (3–10 year maturities), and long-term (greater than 10 year maturities).

Lower level of governments also issue debt securities to finance their borrowing needs.

5.4 Corporate Bonds

Corporate bonds are issued by corporations for various financing needs. Corporate bonds carry the promise of specified payments at specified times and provide legal remedies in the event of default.

From the bond issuers' point of view, debt financing differs from equity financing in the following two ways: (1) the principal and interest payments of a bond are obligatory while the equity of a stock represents an ownership. (2) the interest payments are considered expenses to the corporation from the taxation point of view while the dividend payments are not.

The Indenture

An issue of bonds is generally covered by a trust indenture or trust deed, which contains a promise to a specific trustee that the issuer will comply with a number of provisions. The trustee of a bond issue acts on behalf of the bondholders. The information in the indenture is summarized briefly in the prospectus for new bond issuing.

Types of Bonds

Mortgage bonds represent debt that is secured by the pledge of specific property.

Collateral trust bonds are the bonds backed by other securities that usually held by the trustee.

Equipment obligations are backed by specific pieces of equipment.

Debentures are general obligations of the issuer and thus present unsecured credit.

Subordinated debentures refer those debentures which are junior to unsubordinated debentures. In the event of bankruptcy, junior clams are to be considered only after senior claim have been fully satisfied.

Other Types of Bonds

Income bonds are the bonds which may not pay interest in full and on schedule.

Guaranteed bonds are the bonds issued by the corporation but guaranteed by another corporation.

Convertible bonds are the bonds which can be converted into other securities (say common stocks) at the option of the bondholder.

Extendible bonds are the bonds which may have an extended term-to-maturity.

Retractable bonds are the bonds which give the bondholder the option to reducing the term-to-maturity on a specific date.

Call Provisions

Some bonds have the call provisions which allow the issuer to call the bonds back at par prior to the maturity. This feature will affect the pricing of the bonds.

Sinking Funds

A sinking fund is a fund into which annual payments must be made by a bond indenture issuer. The existence of a sinking fund will reduce the amount outstanding at maturity.

Foreign Bonds and Eurobonds

Foreign bonds are those that are issued and denominated in the currency of a country other than that of the issuer.

Eurobonds are those that are issued in a country other than that of the issuer and in a currency different from that of the country where they are offered.

5.5 Debt Instruments Trading

Unlike the stock trading that is conducted at various stock exchanges with their physical locations, debt instruments trading till recently do not have designated exchanges with their physical locations. Debt instruments trading occurs among dealers reported to some information system (such as TRACE in the U.S.) or through some kinds of platform (such as CanDeal in Canada).

5.6 Valuation of Riskless Securities

The yield-to-maturity on any fixed-income security is the discount rate/interest rate that makes the present value of all the cash flows of a bond equal to its market price.

For a one-year discount bond A, the yield-to-maturity r_A is the solution to

$$(1 + r_A) \times P_0 = \$1000 \tag{5.1}$$

or

$$P_0 = \frac{\$1000}{(1 + r_A)}. \tag{5.2}$$

where P_0 is the known current bond price.

For a two-year discount bond B, the yield-to-maturity r_B is the solution to

$$(1 + r_B)^2 \times P_0 = \$1000 \tag{5.3}$$

or

$$P_0 = \frac{\$1000}{(1 + r_B)^2}. \tag{5.4}$$

For a two-year coupon bond C with coupon payment \$50 per year, the yield-to-maturity r_C is the solution to

$$(1 + r_C)[(1 + r_C) \times P_0 - \$50] - \$50 = \$1000 \tag{5.5}$$

$$P_0 = \frac{\$50}{(1 + r_C)} + \frac{\$50 + \$1000}{(1 + r_C)^2}. \tag{5.6}$$

A spot rate is measured at a given time as the yield-to-maturity on a pure-discount security and can be thought of as the interest rate associated with a spot contract. Spot rates can also be determined in another manner if only coupon-bearing bonds are available for longer maturities. Generally the one-year spot rate (s_1) will be known. If no two-year pure-discount bond exists and only a two-year coupon-bearing bond is available (it has a current market price P_2, a maturity value of M, a coupon payment one year from now equal to C_1, a coupon payment two year from now equal to C_2), then the two-year spot rate (s_2) is the solution to

$$P_2 = \frac{C_1}{(1 + s_2)} + \frac{C_2}{(1 + s_2)^2} + \frac{M}{(1 + s_2)^2}. \tag{5.7}$$

In general we can compute the T year coupon bond price (P) using the following equation conditional on C_t ($t = 1, 2, \ldots, T$) and the T-year spot rate (s_T).

$$P = \frac{C_1}{(1 + s_T)} + \frac{C_2}{(1 + s_T)^2} + \cdots + \frac{C_T}{(1 + s_T)^T} + \frac{M}{(1 + s_T)^T} \tag{5.8}$$

or

$$P = \sum_{i=1}^{T} d_t C_t + d_T M \tag{5.9}$$

where $d_t = \frac{1}{(1+s_T)^t}$, is the discount factor.

When $C_t = C$ and $s_T = r$ for all t we can compute the T year coupon bond price P as follows:

$$P = \frac{C}{r}\left[1 - \frac{1}{[1+r]^T}\right] + \frac{M}{(1+r)^T}. \tag{5.10}$$

Why? The term $\frac{M}{(1+r)^T}$ represents the present value of M paid in period T. Now we focus on $\frac{C}{r}\left[1 - \frac{1}{[1+r]^T}\right]$. Let $\lambda = \frac{1}{(1+r)}$. Then we can write

$$\frac{C}{(1+r)} + \frac{C}{(1+r)^2} + \cdots + \frac{C}{(1+r)^T} \tag{5.11}$$

$$= C(\lambda + \lambda^2 + \cdots + \lambda^T) \tag{5.12}$$

The term $C(\lambda + \lambda^2 + \cdots + \lambda^T)$ can be expressed as

$$C(\lambda + \lambda^2 + \cdots + \lambda^T) = C(S - \lambda^T S) = CS(1 - \lambda^T), \tag{5.13}$$

where $S = \lambda + \lambda^2 + \lambda^3 + \cdots = \frac{\lambda}{1-\lambda} = \frac{\frac{1}{1+r}}{1-\frac{1}{1+r}} = \frac{1}{r}$. We can further simplify $CS(1-\lambda^T)$:

$$C(\lambda + \lambda^2 + \cdots + \lambda^T) = C\frac{\lambda}{1-\lambda}(1-\lambda^T) = \frac{C}{r}\left[1 - \frac{1}{[1+r]^T}\right]. \tag{5.14}$$

Therefore we have

$$P = \frac{C}{r}\left[1 - \frac{1}{[1+r]^T}\right] + \frac{M}{(1+r)^T}. \tag{5.15}$$

Now consider a 30-year bond with 8 % coupon rate and issued at par with the ongoing interest rate of 8%. The price of the bond is

$$P = \frac{\$80}{0.08}\left[1 - \frac{1}{[1+0.08]^{30}}\right] + \frac{\$1000}{(1+0.08)^{30}} = \$1000.00. \tag{5.16}$$

But the interest rate rises to 9% as soon as the bond is issued, the bond price will change to

$$P = \frac{\$80}{0.09}\left[1 - \frac{1}{[1+0.09]^{30}}\right] + \frac{\$1000}{(1+0.09)^{30}} = \$897.26. \tag{5.17}$$

The bond losses 10.27 % of its value. On the other hand, if the interest rate falls to 7% as soon as the bond is issued, the bond price will change to

$$P = \frac{\$80}{0.07}\left[1 - \frac{1}{[1+0.07]^{30}}\right] + \frac{\$1000}{(1+0.07)^{30}} = \$1124.09. \tag{5.18}$$

The bond gains 12.41 % of its value.

One of examples for forward rates is the one-year interest rate from year one to year two ($f_{1,2}$) which the solution to the following problem

$$P = \frac{\$1}{(1+s_1)(1+f_{1,2})} \tag{5.19}$$

with P and s_1 being known. The forward rate could be the rate for a forward contract. That is, the forward rate applies to the contacts made now but for a period "forward" in time.

In practice, the current yield of a bond is often reported. It is calculated by

$$current \;\; yield = \frac{annual \;\; coupon}{price}. \tag{5.20}$$

Be aware that the current yield has its own drawbacks. On one hand, it does not consider the reinvestment return of coupon. On the other hand, it does not consider the difference between the purchasing price and redemption price, which could involve capital gains and losses.

Be aware that the yield-to-maturity is not necessarily the yield the bond investor gets even if he or she plans to hold the bond till it matures. Why? When the bond is issued at par (at the full face value of \$100), the coupon rate is often set at the yield-to-maturity. If the interest rate falls before the bond matures, it is not possible for the bond investor to reinvest the coupon payments at the yield-to-maturity. The actual return from this bond investment may be lower than the yield-to-maturity.

5.7 The Risk Premium and Term Premium

- The risk premium is the additional return on a risky asset over the return on a risk-free asset.
- An Example

  A corporate bond has a higher yield than a government bond.

  The risk premium = the yield on a corporate bond - the yield on a government bond.
- The term structure of risk premiums refers to the relationship between risk premiums and terms-to-maturity.
- The term premium is the additional return on a longer maturity asset over the return on a shorter maturity asset of the identical quality.
- An Example

  The term premium of the US Treasury bill yields.
- The term structure of term premiums refers the relationship between term premiums and terms-to-maturity.

5.8 The Term Structure of Interest Rates

If one plots the interest rates against their terms-to-maturity, you will have a yield curve at one point in time. A yield curve represents the current term structure of interest rates. The term structure of interest rates refers to the relationship between interest rates and terms-to-maturity. The term structure can be (1) upward sloping, (2) flat, and (3) downward sloping.

The term structure contains useful economic information but our understanding about its behaviors is still not perfect. There are three major theories:

- The Liquidity Preference Theory

  This theory suggests that the market for long-term bonds is less liquid than that for short-term bonds. The investors are primarily interested in purchasing more liquid securities. Thus a liquidity premium is required by, and should be awarded to, the investors for accepting less liquid long-term bonds.

  Given the fact is that the term structure may be upward sloping, flat, or downward sloping, this theory is very limited because it can predict only one of possible cases.

- The Market Segmentation Theory

  This theory suggests that the long- and short-term bond markets are separate markets. The demand for and supply of loanable funds interact within each market. Hence it is possible to observe upward sloping, flat, or downward sloping term structures.

- The Unbiased Expectations Theory

  This theory suggests that the forward interest rate (or forward rate) for a one-year pure-discount bond prevailing in year t represents the average opinion of the expected future spot interest rate (or spot rate) for this one-year pure-discount bond prevailing in year t with the expectations formed in year one. More specifically, let the forward rate for a one-year pure-discount bond prevailing in year two with the expectations formed in year one be $f_{1,2}$. This implies that the spot rate for a two year pure-discount bond (s_2) can be viewed to satisfy the following relationship with the spot rate for the one-year pure-discount bond (s_1) in year one and the forward rate for the one-year pure-discount bond prevailing in year two with the expectations formed in year one ($f_{1,2}$):

  $$(1+s_2)(1+s_2) = (1+s_1)(1+f_{1,2}). \tag{5.21}$$

  This implies that s_2 contains the information of s_1 and $f_{1,2}$.

  Example: Let $s_1 = 4\%$ and $s_2 = 5\%$. Please find $f_{1,2}$.

  Approach 1:

  $$f_{1,2} = \frac{(1+s_2)^2}{1+s_1} - 1 = \frac{(1+0.05)^2}{1+0.04} = 0.06009615 \approx 6\%. \tag{5.22}$$

  Approach 2: Simplify

  $$(1+s_2)(1+s_2) = (1+s_1)(1+f_{1,2}) \tag{5.23}$$

  into

  $$1 + 2s_2 + s_2^2 = 1 + s_1 + f_{1,2} + s_1 f_{1,2}, \tag{5.24}$$

  and then into

  $$2s_2 + s_2^2 = s_1 + f_{1,2} + s_1 f_{1,2}. \tag{5.25}$$

  Because s_2^2 and $s_1 f_{1,2}$ are very small, we can remove them simultaneously to get

  $$2s_2 \approx s_1 + f_{1,2} \tag{5.26}$$

  or

  $$s_2 \approx \frac{s_1 + f_{1,2}}{2}. \tag{5.27}$$

Using the above approximation, we have

$$f_{1,2} \approx 2s_2 - s_1 = 2(0.05) - 0.04 = 0.10 - 0.04 = 0.06 = 6\%. \tag{5.28}$$

Example: Let $s_1 = 4\%$ and $s_2 = 3\%$. Please find $f_{1,2}$.

As Approach 2 is simplier, we use it to get

$$f_{1,2} \approx 2s_2 - s_1 = 2(0.03) - 0.04 = 0.02 = 2\%. \tag{5.29}$$

Therefore, we conclude the spot rates of long pure-discount bonds contain the information of forward interest rates of one-year pure-discount bonds in future years. If the spot rates of these long bonds are higher (lower), the forward rates of these one-year pure-discount bonds in future years are higher (lower). The unbiased expectations theory emphasizes the link.

5.9 Bond Attributes and Determinations of Yield Spreads

There are six attributes of a bond are important in bond evaluation:

1. length of time until maturity;
2. coupon rate;
3. call provisions;
4. taxation;
5. marketability;
6. likelihood of default.

Length of time until maturity and coupon rate determine the size and timing of the cash flow and therefore the price of a bond.

Call provisions enables the issuer to redeem the bonds prior to maturity at a call price, which is generally above par. The difference between the call price and the par value is called call premium.

Taxation affects the bond pricing in two ways: (1) bonds differing in taxation status will be priced differently with everything else hold equal (taxable and tax-exempt bonds); and (2) when the tax rate on capital gains is lower than that on coupon payments, deep discount bonds have a tax advantage over small- or non-discount bonds and low-coupon bonds have a tax advantage over high-coupon bonds.

Marketability or liquidity refers to the ease with which an investor can sell an asset without having to make a substantial price concession. In general, the bonds that are actively traded have a smaller bid-ask spread and are more liquid that those that are less or not actively traded.

The likelihood of default will also affect the bond pricing. Generally, a bond with no likelihood of default is called a riskless asset. When a bond with a higher likelihood of default will be priced lower (therefore it has a higher yield) than that with a lower likelihood of default. The bond rating agencies will provide ratings for various bonds.

The difference between the yield on a risky (default free or not default free) bond and the yield on a riskless bond is often called yield spread. This spread can be decomposed further into two parts: (1) default premium (the difference between the yield on a risky and not default free bond and the yield on a risky and default free bond) and (2) risk premium (the difference between the yield on a risky and default free bond and the yield on a riskless bond).

One of the US study on corporate bond pricing finds that the yield spreads are determined by the four factors:

1. The extent to which the firm's net income had varied over the preceding nine years.

2. The length of time that the firm had operated without forcing any of its creditors to take a loss.

3. The ratio of the market value of the firm's equity to the par value of its debt.

4. The market value of the firm's outstanding debt.

5.10 Bond Market Efficiency

There are two general methods of managing bond portfolios: (1) passive methods and (2) active methods.

Passive methods are used by managers who believe that the market is efficient and that security selection and market timing will not be useful in getting above-average returns.

Active methods are used by managers who believe that the market is not efficient and that security selection and market timing will be useful in getting above-average returns.

According to the US and Canadian studies, the US bond markets appear to be highly, but not perfectly, efficient in the semi-strong form (that is, one cannot predict the future price by using all public information). The Canadian bond markets are less active and have some, possibly insignificant, inefficiencies.

More specifically, the past study shows that the knowledge on how Treasury bill prices had changed in the past was of little use in predicting how they would change in the future. Thus, Treasury bill markets are efficient in the weak form (that is, one cannot predict the future price by using past prices).

The bond market efficiency is also studied by examining the accuracy of the interest rate predictions made by experts. Since the experts use various data sources to make a prediction, this type of studies can be viewed as ones for semi-strong form efficiency. The general conclusion from these studies is that it is difficult to consistently forecast interest rates with greater accuracy than a no-change prediction/model.

Semi-strong form of efficiency can also be studied by looking at the bond price reaction to bond rating changes. If the market is efficient in semi-strong form, a rating change announcement should not cause the bond price to change. The US empirical evidence shows that rating increases (decreases) are preceded by bond price increases (decreases). That is, investors are able to anticipate bond rating changes and bond prices are able to reflect the facts causing rating changes.

Semi-strong form of efficiency can be evaluated by checking whether bond price adjustments occur after the announcements of money supply. The evidence is positive.

5.11 Bond-Pricing Theorems

Coupon payments refer to the periodic payments from the issuer to the bond bearer. A bond's coupon rate is the ratio of the dollar amount of the annual coupon payments to the principal of the bond. The term-to-maturity is the amount of time left until the last promised payment is made. The yield-to-maturity is the discount rate/interest rate that makes the present value of all the cash flows of a bond equal to its market price.

When the price of a bond is equal to its par value, the yield-to-maturity will be equal to its coupon rate. When the bond price is less than its par value (the bond is traded at a discount), the yield-to-maturity will be greater than the coupon rate. When the bond price is greater than its par value (the bond is traded at a premium), the yield-to-maturity will be less than the coupon rate.

Let the bond price be

$$P = \sum_{t=1}^{T} C_t(1+r)^{-t} + V(1+r)^{-T}, \tag{5.30}$$

where V is the face value of the bond. The bond-pricing theorems give the quantitative relationships among P, r, V, C, and T. These theorems are as follows:

1. If a bond's market price increases, then its yield must decrease; conversely, if a bond's market price decreases, then its yield must increase. [P vs. r, the first order relationship]

   For example, consider a two-year bond,

   $$P = C(1+r)^{-1} + (C+V)(1+r)^{-2}, \tag{5.31}$$

   we have

   $$\frac{\partial P}{\partial r} = -C(1+r)^{-2} - 2(C+V)(1+r)^{-3} < 0. \tag{5.32}$$

2. A decrease in a bond's yield will raise the bond's price by an amount that is greater in size than the corresponding fall in the bond's price that would occur if there were an equal-sized increase in the bond's yield. [P vs. r, the second order relationship]

   Using the above example,

   $$\frac{\partial^2 P}{\partial r^2} = 2C(1+r)^{-3} + 6(C+V)(1+r)^{-4} > 0. \tag{5.33}$$

3. The percentage change in a bond's price due to a change in its yield will be smaller if its coupon rate is higher. [P vs. r less convex when $C\uparrow$]

   Using the above example,

   $$\begin{aligned} \frac{\partial\left(\frac{\partial P}{P\partial r}\right)}{\partial C} &= \frac{\partial\left(-CP^{-1}(1+r)^{-2} - 2(C+V)P^{-1}(1+r)^{-3}\right)}{\partial C} \\ &= -P^{-1}(1+r)^{-2} - 2P^{-1}(1+r)^{-3} < 0. \end{aligned} \tag{5.34}$$

   This theorem does not apply to bonds with a life of one year or to bonds that have no maturity date, known as consols or perpetuities.

4. If a bond's yield does not change over its life, then the size of its discount or premium will decrease as its life gets shorter.

   This theorem has an alternative but equivalent interpretation: If two bonds have the same coupon rate, par value, and yield, then the one with short life will be traded at a smaller discount or premium than its longer-life counterpart. [$\Delta(P-V)$ vs. Δt when r is fixed]

   There are at least two ways to show the theorem.

   The first way is as follows.

   Recall $P = \sum_{t=1}^{T} C_t(1+r)^{-t} + V(1+r)^{-T}$. Let $C_t = C$ and V and T be fixed. The coupon rate is set at $i = C/V = r$ when the bond is issued. Let r_1 and r_2 as the new interest rates after the issuing, where $r_1 < r < r_2$. Let t and P be the only variables but t is discrete while P is continuous. Let $1 < t_1 < t_2 < T$. Corresponding to t_1 and t_2, there are P_1 and P_2, respectively. If $r_1 < r = i = C/V$, then $P_i - V > 0$ for $i = 1, 2$. That means that the bond is sold at premium. Corresponding to $t_1 < t_2$, we have $P_1 - V > P_2 - V$ because P_1 has more price premiums to justify more coupon payments from t_1 to T at the high coupon rate $(r_1 < r = i = C/V)$ than P_2 has from t_2 to T. Similarly, if $r_2 > r = i = C/V$, then $P_i - V < 0$ for $i = 1, 2$. That means that the bond is sold at discount. Corresponding to $t_1 < t_2$, we have $P_1 - V < P_2 - V$ because P_1 has more price discounts to justify more coupon payments from t_1 to T at the low coupon rate $(r_2 > r = i = C/V)$ than P_2 has from t_2 to T.

   The second way is as follows.

   Using the formula for the annuity, we can write $\sum_{t=1}^{T} C(1+r)^{-t}$ as $\frac{C}{r}\left(1 - \frac{1}{(1+r)^T}\right)$. Then

   $$P = \frac{C}{r}\left(1 - \frac{1}{(1+r)^T}\right) + \frac{V}{(1+r)^T}.$$

   Re call that given $y = a^x$, $\frac{dy}{dx} = a^x \ln(a)$. Therefore, we have

   $$\frac{\partial P}{\partial T} = \left(\frac{C}{r} - V\right)(1+r)^{-T}\ln(1+r).$$

   When $r_1 < r = i = C/V$, where r_1 is the new interest rate, the bond is sold at premium, $[\frac{C}{r_1} - V] > 0$. Hence, $\partial P/\partial T > 0$.

   When $r_2 > r = i = C/V$, where r_1 is the new interest rate, the bond is sold at discount, $[\frac{C}{r_2} - V] < 0$. Hence, $\partial P/\partial T < 0$.

5. If a bond's yield does not change over its life, then the size of its discount or premium will decrease at an increasing rate as it approaches maturity. [$\Delta^2(P-V)$ vs. $\Delta^2 t$ when r is fixed]

   There are at least two ways to show the theorem.

   The first way is as follows.

   Let $C_t = C$, $1 < t_1 < t_2 < t_3 < t_4 < T$ and $t_2 - t_1 = t_4 - t_3 = 1$. It is clear that

   $$\sum_{t=t_1}^{t_2} C(1+r)^{-t} > \sum_{t=t_3}^{t_4} C(1+r)^{-t}. \tag{5.35}$$

Hence, from t_1 to t_2, we have $\Delta P_1 = P_2 - P_1$ and from t_3 to t_4, we have $\Delta P_3 = P_4 - P_3$. The above equation implies $\Delta P_1 > \Delta P_3$. Because V is fixed and $\Delta V = 0$, this inequality can be written as $\Delta(P_1 - V) > \Delta(P_3 - V)$ for $\Delta^2 t$.

The second way is as follows.

Given

$$\frac{\partial P}{\partial T} = \left(\frac{C}{r} - V\right)(1+r)^{-T}\ln(1+r),$$

we can find

$$\frac{\partial^2 P}{\partial T^2} = -\left(\frac{C}{r} - V\right)(1+r)^{-T}[\ln(1+r)]^2.$$

Recall, for $r_1 < r < r_2$, we have $-(C/r_1 - V) < 0$ and $-(C/r_2 - V) > 0$.

Hence, when $r_1 < r$, $\partial^2 P/\partial T^2 < 0$. When $r_2 > r$, $\partial^2 P/\partial T^2 > 0$.

5.12 Convexity

The bond convexity refers the relationship between the bond price and the yield-to-maturity. If we want to plot the relationship in the price (P)-yield (r) space. The relationship will be a downward sloping, convex curve. This relationship is also stated in the first and second bond pricing theorems. The first theorem says that the relationship is an inverse one while the second theorem adds that the relationship is a convex one.

The degree of convexity or curvature is not identical for all bonds. It depends on several factors: (1) the size of the coupon payments [$C \uparrow \Rightarrow$ less convex], (2) the life of the bond [$T \uparrow \Rightarrow$ more convex], and (3) the current market price and prevailing yield, among other things [P and $r \Rightarrow$ location of the convex function].

The convexity is an important concept for bond portfolio management as shown later.

5.13 Duration

The term-to-maturity or the life of a bond is not the same as the duration of a bond. Duration is a measure of the "net-present-value-weighted average of maturities" of the stream of payments associated with a bond. It is defined as

$$D = \frac{\sum_{t=1}^{T} PV(C_t) \times t}{P_0} \tag{5.36}$$

where P_0 is the current bond price and $PV(C_t) = d_t C_t$ is the present value of the cash received at time t (Recall $d_t = \frac{1}{(1+r)^t}$). If the bond is correctly priced, $P_0 = \sum_{t=1}^{T} PV(C_t)$. The alternative expression of duration

$$D = \sum_{t=1}^{T}\left[\frac{PV(C_t)}{P_0} \times t\right] \tag{5.37}$$

can help us understand why the duration is a weighted average of maturities.

Example: Now consider a three-year bond with coupon payment \$80 per year is traded at \$950.25. The interest rate is 10%. What is the duration of the bond?

$$\begin{array}{rrrrr} Tmie & Cash & PV & PV(Cash) & PV(Cash) \times Time \\ 1 & \$80 & .9091 & \$72.73 & \$72.73 \\ 2 & 80 & .8264 & 66.12 & 132.23 \\ 3 & 1080 & .7513 & 811.40 & 2434.21 \\ Total & & & \$950.25 & \$2639.17 \end{array} \tag{5.38}$$

Thus the bond is correctly priced and its duration is 2.78 = \$2639.17/\$950.25 = 2.78.

A zero-coupon bond has a duration equal to its remaining life because

$$D = \frac{PV(C_T)}{P_0} \times T = 1 \times T = T. \tag{5.39}$$

Now examine the role that duration plays to the bond price change when the bond yield changes. Let the bond price change be ΔP and the yield and yield change be r and Δr, respectively. The percentage change in price is $\Delta P/P$. The following relationship holds

$$\frac{\Delta P}{P} \simeq -D\left(\frac{\Delta r}{1+r}\right). \tag{5.40}$$

This relationship indicates that the bond price change will differ as the duration differs given the equal amount of change in the yield. This approximation is very useful for bond portfolio management. Because changes in interest rates will lead to changes of bond prices and hence values of bond portfolios. This relationship is a linear approximation of the convexity. Note that the bond price can be computed as

$$P = \sum_{t=1}^{T} C_t(1+r)^{-t} \tag{5.41}$$

Differentiating P with respect to r gives

$$\frac{dP}{dr} = -\sum_{t=1}^{T} t \cdot C_t(1+r)^{-t-1}. \tag{5.42}$$

Multiplying both sides of the above equation by $\frac{dr}{P_0}$ and rearranging yield

$$\frac{dP}{P_0} = -\left(\frac{\sum_{t=1}^{T} t \cdot C_t(1+r)^{-t}}{P_0}\right)\left(\frac{dr}{1+r}\right). \tag{5.43}$$

Note that the term $\frac{\sum_{t=1}^{T} t \cdot C_t(1+r)^{-t}}{P_0}$ is duration. Thus

$$\frac{dP}{P_0} = -D\left(\frac{dr}{1+r}\right). \tag{5.44}$$

From the above derivation we can see that $\frac{\Delta P}{P} \simeq -D\left(\frac{\Delta r}{1+r}\right)$ is indeed an approximation of $\frac{dP}{P_0} = -D\left(\frac{dr}{1+r}\right)$.

Sometimes the concept of modified duration is used. It is defined as

$$D_M = \frac{D}{1+r}. \tag{5.45}$$

With this concept, the percentage change in bond price is given by

$$\frac{\Delta P}{P} \simeq -D_M \Delta r. \tag{5.46}$$

Because some changes in interest rates will have adverse effects on bond portfolios, it is necessary to get bond portfolio immunized from such changes.

5.14 Immunization

Immunization is a technique used in bond portfolio management. This technique is said to allows a bond portfolio manager to be relatively certain that he or she could meet a given promised stream of cash outflows.

How can immunization be accomplished? Two step must be taken: (1) calculate the duration of the promised outflows and (2) then invest in a portfolio of bonds that have an identical duration.

Note that the duration of a portfolio of bonds is equal to the value-weighted average of the durations of the individual bonds in the portfolio. For example, one-third of a portfolio's funds are invested in bonds having a duration of six years and two-thirds of its funds are in bonds having a duration of three years. Then, the duration of the portfolio is $(1/3)6+(2/3)3 = 4$ years.

Consider a portfolio manager who has a cash outflow of \$ 1 million in two years. The duration of the cash outflow is two years. Ideally, the manager can invest in a bond portfolio with a duration of two years. The zero-coupon bonds maturing in two years will be a best choice. But these bonds are not always available. Let us say that the manager can only invest in three-year bonds and one-year bonds. If he or she decides to invest only in one-year bonds, he or she will face reinvestment-rate risk. That is, if the yield falls, he or she may not be able to get the same reinvestment interest rate such that the interest payments and principal will add up to \$ 1 million. If the manager decides to invest only in three-year bonds, he or she must sell the bonds at the end of the second year. At that time, if the yield rises, he or she must sell the bonds at a lower price so that the obligation of \$ 1 million cash outflow will not be met.

One proposed solution is to invest part of the portfolio's fund in the one-year bonds and the rest in the tree-year bonds. Let the weights for one-year and three-year bonds be w_1 and w_2, respectively. We know that one-year bonds have a duration of 1 year and the three-year bonds have a duration of 2.78 years (as shown above). The manager can form a portfolio such that

$$w_1 + w_2 = 1 \tag{5.47}$$

and

$$(w_1 \times 1) + (w_2 \times 2.78) = 2. \tag{5.48}$$

By solving this equation system, $w_1 = .4382$ and $w_2 = .5618$. Given the current yield is 10%, how much the investor has to invest? $\$1,000,000/(1+.10)^2 = \$826,446$. So $w_1 \times \$826,446 = \$326,149$ can be invested in the one-year bonds and $w_2 \times \$826,446 = \$464,297$ can be invested in the three-year bonds.

What happens if the yield changes?

If the yield falls, the loss from reinvesting in one-year bonds at a lower yield in the second year will be exactly offset by the gain from selling of three-year bonds at a premium after two years.

If the yield rises, the loss from selling of three-year bonds at a discount after two years will be exactly offset by the gain from reinvesting in the one-year bonds at a higher yield in the second year.

Immunization also has its limitations:

1. Default and call risk;
2. Multiple nonparallel shifts in a nonhorizontal yield curve;
3. Rebalancing;

4. Choosing among many candidates.

5.15 Active Management

Active management of bond portfolios is based on the belief that the bond market is not perfectly efficient. There are two general approaches to active management: (1) security selection which involves selecting mispriced bonds and (2) market timing which is implemented by predicting the trend of future yield movements.

Horizon Analysis

The return on a bond over any given holding period is determined by the difference between the price at the beginning of the period and the price at the end of the period and its coupon payments.

The bond price is affected by the yield at both the beginning and end of the period. If one can identify the future changes of yields, he or she can utilize this for increasing returns.

Horizon analysis involves selecting a single-holding period, forecasting the yield at the end of the period, identifying and purchasing the underpriced bond, and selling the bond at the end of the period for a capital gain.

We can consider this example. A bond manager wants to invest in bond for five years and he notes a 4% coupon bond with the term-to-maturity of ten years. If the yield at the beginning of the five-year period is 9% ($r_0 = 9\%$) and the bond price at the beginning of the period is \$67.91 ($P_{0,9\%} = \$67.91$). The bond manager forecasts that the yield-to-maturity at the end of the period will fall to 8% ($r_1 = 8\%$) so that the bond price corresponding to the 8% yield at the end of the period will be \$84.03 ($P_{1,8\%} = \$84.03$) . If the yield-to-maturity does not change ($r_1 = 9\%$), then the bond price corresponding to the 9% yield at the end of the period will be \$80.55 ($P_{1,9\%} = \$80.55$).

In the event that the bond manager forecasts the future yield successfully, the change in bond price will be $P_{1,8\%} - P_{0,9\%} = \$84.03 - \$67.91 = \$16.12$, which can be decomposed into the time effect ($P_{1,9\%} - P_{0,9\%} = \$12.64$) and yield change effect ($P_{1,8\%} - P_{1,9\%} = \$84.03 - \$80.55 = \$3.48$).

In addition, if the coupon payments is \$4 in each year the bond holder can invest \$4 immediately at 8% per year, the total value at the end of the period is

$$\$4(1.08^4 + 1.08^3 + 1.08^2 + 1.08^1 + 1) \approx \$23.47 \tag{5.49}$$

where \$20 is from coupon payments and \$3.47 is the interest on the coupon payments.

Thus if the bond manager can forecast the future yield ($r_1 = 8\%$), the total return on this investment will be

$$\frac{\$16.12 + \$23.47}{\$67.91} \approx 58.30\%. \tag{5.50}$$

Bond Swaps

The purpose of bond swaps is to use underpriced bonds to replace overpriced bonds to achieve higher returns.

In substitution swap, the bond manager exchanges a bond for a perfect substitute to take the price advantage in an imbalanced market.

In inter-market swap, the bond manager moves out of one market and move into another market in order to exploit the advantages from the projected changes in two markets.

In rate anticipation swap, the bond manager changes bond holdings to profit from the general movement in the overall market rates.

In pure yield pickup swap, the bond manager changes bond holdings to get a better yield.

Contingent Immunization

The basic idea of contingent immunization is to pursue active management as long as favorable results can be obtained. If unfavorable results are expected, the portfolio will be immediately immunized.

Riding the Yield Curve

Riding the yield curve is a method of bond portfolio management for maintaining liquidity. Normally the bond manager uses only short-term fixed-income securities for this purpose; that is, the manager purchases these securities, hold them to maturity and reinvest in these type securities again.

When certain conditions hold, it is possible to use the method of riding the yield curve. The conditions are: (1) the yield curve is upward sloping and (2) the upward sloping yield curve will remain its shape when the first batch of short-term fixed-income securities mature. The bond manager can purchase securities with somewhat longer term-to-maturity and sell them before they mature so that they can capture capital gains.

Be aware that when the yield curve changes adversely, this method may reduce returns. Riding the yield curve method also have higher transaction costs.

5.16 Review Questions

1. What is the money market? What instruments are considered money market instruments?

2. Please explain the major bond issuers in the bond market.

3. Please explain how to calculate the present value of a cash flow from time 1 to time T (e.g., 10 years).

4. What is risk premium? What is term premium? How do they differ?

5. What are the three major theories about the term structure of interest rates?

6. Please explain the liquidity preference theory.

7. Please explain the market segmentation theory.

8. Please explain the unbiased expectations theory.

9. What are the six attributes of a bond? Please explain each briefly.

10. Could you summarize the four factors that influence the yield spread of the U.S. corporate bonds?

11. What do we mean when we say the market is efficient? Please explain the weak form, semi-strong form, and strong form of market efficiency.

12. Please explain the five bond pricing theorems in words and graphs.

13. What is convexity?

14. What is duration?

15. How to immunize interest rate risk using the concept of duration?

16. Show some examples for active bond management.

Chapter 6

Stock Analysis

6.1 Learning Objectives

- The corporate form and common shares
- Cash dividends
- Preemptive rights
- Common stock betas
- Growth versus value
- Capitalization-of-income method of valuation
- The zero-growth model
- The constant-growth model
- The multiple-growth model
- Valuation based on a finite holding period
- Model based on price-earning ratios
- Sources of earning growth
- A three-stage dividend discount method
- Stock valuation based on earnings
- Determinants of dividends
- The information content of dividends
- Accounting earnings versus economic earnings
- Price-earning ratios

6.2 The Corporate Form

A corporate is a business entity with a charter, or certificate of incorporation, by the governments. The charter specifies the rights and obligations of shareholders.

The share certificates are issued to the holders as the proof of the ownership. The shares are issued through primary distributions and secondary distributions. The share holders have the right to receive dividend payments, voting materials, financial reports and so on.

Shareholders' equity include common shares (par value), capital contributed in excess of par value (additional capital obtained by selling at the price above par value), and cumulative retained earnings (residual earnings). This is also the book value of the equity. The book value per share is computed by dividing the book value of the equity by the number of shares outstanding.

Treasury shares refer those common shares that have been issued and then bought back by a corporation in the open market or through a tender offer. These shares differ from the shares outstanding because there are no voting right or rights to receive dividends offered to the outstanding shares.

Restricted shares are those equity shares whose voting rights are in some way diminished or curtailed. The example of this type of shares are subordinate voting shares and non-voting shares.

Some large corporation also issue preferred shares. The preferred share is priced higher because the owner is entitled higher and more regular cash dividends.

6.3 Cash Dividends

Cash dividends are payments made in cash to shareholders. There are four important dates about dividends: declaration date, ex-dividend date (on this date onwards, shareholders will not receive dividends), date of record, and payment date.

Stock dividends are dividends in stocks to shareholders. The dividends-paying corporation need to record this increase of common stocks and reduce retained earnings.

Stock split is an accounting transaction that increases the amount of shares and reduce the par value of a share proportionally. Reverse stock split is a reversed transaction of stock split.

Stock dividends and split (1) shows the evidence of corporation growth and (2) makes shares more accessible. Empirical studies show that investments made before stock split announcements generate abnormal returns.

6.4 Preemptive Rights

Shareholders have inherent rights to maintain their proportionate ownership of the corporation. These rights are called preemptive rights that entitle shareholders to purchase new common shares in proportion to the number of shares that they currently own when the corporation issues new shares.

Sometimes shareholders may have oversubscription privilege, that is the opportunity given to shareholders who have excised their rights in a rights offering to buy shares that were not purchased in the offering.

In a rights offering, the price at which holders of rights can purchase new shares is called the subscription price.

Rights are distributed in a manner similar to cash dividends. There is an ex-right date. Before the ex-right date, the value of a right, R, is computed as follows:

$$C_0 - (RN + S) = R \tag{6.1}$$

where C_0 is the "right-on" market price of the stock, R is the value of a right, N is the number of rights needed to buy one share, and S is the subscription price. The above equation can be simplified into

$$R = \frac{C_0 - S}{N + 1}. \tag{6.2}$$

On or after the ex-rights date, the value of a right can be computed as follows:

$$C_e - (RN + S) = 0. \tag{6.3}$$

This can be simplified into

$$R = \frac{C_e - S}{N}. \tag{6.4}$$

6.5 Common Stock Betas

The common stock betas are computed by many brokerage firms as indicators of market risk. The market model discussed before can be use to estimate betas using historical data. Recall the market model is

$$r_{it} = \alpha_i + \beta_i r_{It} + \epsilon_{it} \tag{6.5}$$

Given the historical data on r_{it} and r_{It}, the ordinary least squares method can be used to estimate α_i and β_i. We can use "$\widehat{\ }$" to denote the estimates. They are computed by the ordinary least squares method as

$$\widehat{\beta}_i = \frac{\sum_{t=1}^{T}(r_{it} - \overline{r}_i)(r_{It} - \overline{r}_I)}{\sum_{t=1}^{T}(r_{It} - \overline{r}_I)^2} \tag{6.6}$$

and

$$\widehat{\alpha}_i = \overline{r}_i - \widehat{\beta}_i \overline{r}_I, \tag{6.7}$$

where $\overline{r}_i = \frac{\sum_{t=1}^{T} r_{it}}{T}$ and $\overline{r}_I = \frac{\sum_{t=1}^{T} r_{It}}{T}$.

6.6 Growth versus Value

A growth stock is a stock that has experienced or is expected to experience rapidly increasing earnings per share. It is often characterized as having low earning-to-price (E/P) and low book-value-to-market-value (BV/MV) ratios.

The price here refers to the market price of the stock. The book value refers to the book value of the equity based on the most recent balance sheet information. The earning information is taken from the most recent income statement. The earnings per share is then obtained by dividing the total earnings by the number of shares outstanding.

A value stock is a stock that has a relative low price with reference to its underlying value. This can be caused by a number of factors. Generally, it is often characterized as having higher E/P and higher BV/MV ratios.

In addition to the classification of stocks into growth and value stocks, it is possible and useful to consider the size effects. The firms can be classified into small, medium, and large capitalizations.

Empirical findings suggest that portfolios with higher BV/MV ratios tend to have higher monthly returns, that portfolios with higher E/P ratios tend to have higher monthly returns, and that smaller portfolios tend to have higher monthly returns than large portfolios.

6.7 Capitalization-of-Income Method of Valuation

The capitalization-of-income method of valuation is an approach to valuing financial asset. This method values assets based on the cash flows that the assets can generate in the future. More specifically, if we know the discount rate k that is the interest rate used to compute the present value of future cash flows (C_t, $t = 1, 2, \ldots$), then the value of an asset (V)is computed by

$$V = \sum_{t=1}^{\infty} \frac{C_t}{(1+k)^t}. \tag{6.8}$$

Whether or not a financial asset is viewed underpriced, fairly-priced, or overpriced? We can compare its value V with its price P. The difference is called the net present value (NPV) is computed as

$$NPV = V - P = \sum_{t=1}^{\infty} \frac{C_t}{(1+k)^t} - P. \tag{6.9}$$

A financial asset is underpriced if

$$\sum_{t=1}^{\infty} \frac{C_t}{(1+k)^t} > P; \tag{6.10}$$

it is overpriced if

$$\sum_{t=1}^{\infty} \frac{C_t}{(1+k)^t} < P. \tag{6.11}$$

The internal rate of return is a discount rate k^* that can make the value of an asset equal to its price, that is

$$0 = \sum_{t=1}^{\infty} \frac{C_t}{(1+k^*)^t} - P. \tag{6.12}$$

Common shares that pay dividends can be evaluated by the dividend discount method (DDM). That is, we use D_t for the dividend payment in period t instead of C_t. Then the price of a common share can be computed by

$$V = \sum_{t=1}^{\infty} \frac{D_t}{(1+k)^t}. \tag{6.13}$$

The growth of dividends can be computed by

$$g = \frac{D_t - D_{t-1}}{D_{t-1}}. \tag{6.14}$$

6.8 The Zero-Growth Model

When there is a zero growth of dividends $(D_0 = D_1 = D_2 = D_3 = \ldots)$, we can replace all D_t's with D_0 in

$$V = \sum_{t=1}^{\infty} \frac{D_t}{(1+k)^t} \tag{6.15}$$

so that

$$V = \sum_{t=1}^{\infty} \frac{D_0}{(1+k)^t}. \tag{6.16}$$

This can be simplified into

$$V = \frac{D_0}{k}. \tag{6.17}$$

Thus the internal rate of return should be

$$k^* = \frac{D_0}{P}. \tag{6.18}$$

6.9 The Constant-Growth Model

It is possible that the dividend grows at a constant rate g. In that case, the dividend payment in period t is a function of the dividend payment in period $t-1$ and the growth rate g:

$$D_t = D_{t-1}(1+g) \tag{6.19}$$

where $g > 0$. In this case the valuation of the common share is given by

$$V = \sum_{t=1}^{\infty} \frac{D_0(1+g)^t}{(1+k)^t}. \tag{6.20}$$

By using $S = \sum_{t=1}^{\infty} \lambda^t = \frac{\lambda}{1-\lambda}$ where $|\lambda| < 1$, with $\lambda = \frac{1+g}{1+k}$ the equation for V can be simplified into

$$V = D_0 \left(\sum_{t=1}^{\infty} \frac{(1+g)^t}{(1+k)^t} \right). \tag{6.21}$$

Since the second term on the right hand side of the above equation can be simplified further into

$$\sum_{t=1}^{\infty} \frac{(1+g)^t}{(1+k)^t} = \frac{1+g}{k-g}. \tag{6.22}$$

Thus,

$$V = D_0 \left(\frac{1+g}{k-g} \right) = \frac{D_1}{k-g}. \tag{6.23}$$

The internal rate of return is

$$k^* = \frac{D_0(1+g)}{P} + g \tag{6.24}$$

or

$$k^* = \frac{D_1}{P} + g. \tag{6.25}$$

6.10 The Multiple-Growth Model

The multiple-growth model is the model in which dividends are assumed to grow at different rates over specifically defined time periods. For example, dividends from an asset from period 1 to T, $D_1, D_2, \ldots, D_T$, grow at irregular rates while those from period T onwards, $D_{T+1}, D_{T+2}, \ldots$ grow at a constant rate g. The total value of the asset V can be computed as the sum of the present value (V_{T-}) of $D_1, D_2, \ldots, D_T$ and the present value (V_{T+}) of $D_{T+1}, D_{T+2}, \ldots$:

$$V_{T-} = \sum_{t=1}^{T} \frac{D_t}{(1+k)^t} \tag{6.26}$$

and

$$V_{T+} = \frac{D_{T+1}}{(k-g)(1+k)^T}. \tag{6.27}$$

That is,

$$V = V_{T-} + V_{T+} = \sum_{t=1}^{T} \frac{D_t}{(1+k)^t} + \frac{D_{T+1}}{(k-g)(1+k)^T}. \tag{6.28}$$

The internal rate of return in this case has no explicit solution so that we generally write the condition as

$$P = \sum_{t=1}^{T} \frac{D_t}{(1+k^*)^t} + \frac{D_{T+1}}{(k^* - g)(1+k^*)^T}. \tag{6.29}$$

6.11 Valuation Based on a Finite Holding Period

Sometimes investors want to hold an asset for a period time. The evaluation of asset value requires the information of the dividend payments in the period and the market price at the end of the period. Let D_1 be the expected dividend and P_1 the expected selling price at $t = 1$. The value of the asset will be

$$V = \frac{D_1}{1+k} + \frac{P_1}{1+k}. \tag{6.30}$$

This is consistent with our previous model. Because P_1 is equal to

$$P_1 = \sum_{t=2}^{\infty} \frac{D_t}{(1+k)^{t-1}}. \tag{6.31}$$

Substituting the above into the equation for V yields

$$V = \frac{D_1}{1+k} + \frac{1}{1+k} \sum_{t=2}^{\infty} \frac{D_t}{(1+k)^{t-1}} \tag{6.32}$$

which in turn is equal to

$$V = \sum_{t=1}^{\infty} \frac{D_t}{(1+k)^t}. \tag{6.33}$$

6.12 Models Based on Price-Earnings Ratios

In reality, analysts often forecast the future earnings and then translate this information into the forecasted price. The way to do this is (1) to forecast the future earnings (E_1) and (2) to multiply the "normal" industry price-earnings ratio (P_1/E_1) to the earnings (E_1); that is, $P_1 = P_1/E_1 \times\ E_1$.

The expected return (r) can be computed then as a function of the capital gain or loss (price change), expected dividends (D_1) , and purchasing price (P_0):

$$r = \frac{(P_1 - P_0) + D_1}{P_0}. \tag{6.34}$$

Further some analysts want to compare the projected price-earnings ratio with the industry norm. More specifically, let the future earnings be $E_1, E_2, \ldots$, the future payout ratios be p_1,$p_2, \ldots$, the growth rate of earnings be $g_{e1}, g_{e2}, \ldots$. We have

$$\begin{aligned} E_1 &= E_0(1+g_{e1}) \\ E_2 &= E_1(1+g_{e2}) = E_0(1+g_{e1})(1+g_{e2}) \\ E_3 &= E_2(1+g_{e3}) = E_0(1+g_{e1})(1+g_{e2})(1+g_{e3}) \end{aligned} \tag{6.35}$$

Note that if the payout ratio is fixed ($p_1 = p_2 = \cdots = \overline{p}$), then the growth rate of earnings is identical to the growth rate of dividends. But here we allow the payout rate to vary over time. Hence, we need to focus on the growth rate of earnings rather than the growth rate of dividends as we have done previously. The value of the asset is

$$V = \frac{p_1[E_0(1+g_{e1})]}{(1+k)^1} + \frac{p_2[E_0(1+g_{e1})(1+g_{e2})]}{(1+k)^2} \tag{6.36}$$

$$+\frac{p_3[E_0(1+g_{e1})(1+g_{e2})(1+g_{e3})]}{(1+k)^3} + \cdots. \tag{6.37}$$

Dividing both sides of the above equation by E_0 gives

$$\frac{V}{E_0} = \frac{p_1[(1+g_{e1})]}{(1+k)^1} + \frac{p_2[(1+g_{e1})(1+g_{e2})]}{(1+k)^2} \tag{6.38}$$

$$+\frac{p_3[(1+g_{e1})(1+g_{e2})(1+g_{e3})]}{(1+k)^3} + \cdots. \tag{6.39}$$

Thus, other things being equal, the normal price-earnings ratio will be higher if

1. The greater the expected payout ratios (p_1,$p_2, \ldots$,);
2. The greater the expected growth rates in earnings per share ($g_{e1}, g_{e2}, \ldots$,);
3. The smaller the required rate of return (k).

If $V > P$, then a share is underpriced. If $V < P$, the share is overpriced. If we divide both sides of these two inequalities by E_0, we have $V/E_0 > P/E_0$ indicating the share is underpriced and $V/E_0 < P/E_0$ indicating the share is overpriced.

We can discuss several cases where P/E can be used.

The zero-growth model

Assume that the payout ratio is one ($p = 1$) and earnings are constant ($E_0 = E_1 = E_2 = \ldots.$). Then the value of the share is

$$V = \frac{E_0}{k}. \tag{6.40}$$

Thus the price-earnings ratio should be

$$\frac{V}{E_0} = \frac{1}{k}. \tag{6.41}$$

The constant-growth model

Assume that the constant growth rate of earnings (g_e) and the payout ratio (p) are both constant. Then the value of the share is

$$V = pE_0\left(\frac{1+g_e}{k-g_e}\right). \tag{6.42}$$

Thus the price-earnings ration should be

$$\frac{V}{E_0} = p\left(\frac{1+g_e}{k-g_e}\right). \tag{6.43}$$

6.13 Sources of Earnings Growth

Recall the payout ratio in period t is p_t so that $p_t E_t$ is the dividend payout in the same period. $(1-p_t)$ is called the retention ratio. $(1-p_t)E_t$ represents the earnings that are not paid out to shareholders but are invested by the firm as new investment I_t; that is

$$I_t = (1-p_t)E_t. \tag{6.44}$$

The earnings grow because new investment I_t gives a return r_t. Thus the earnings in period $t+1$ can be viewed as the sum of the earnings in period t and the additional earnings generated by the new investment made in period t:

$$E_{t+1} = E_t + rI_t. \tag{6.45}$$

By substituting the equation for I_t into the above equation, we have

$$E_{t+1} = E_t[1 + r_t(1-p_t)]. \tag{6.46}$$

As have shown before, the earnings grow at the rate of g_{et} in period t such that

$$E_t = E_{t-1}(1+g_{et}), \tag{6.47}$$

$$E_{t+1} = E_t(1+g_{et+1}), \tag{6.48}$$

and so on. By comparing the earnings in period $t+1$, one can find that the earnings growth in period $t+1$ is identical to the product of the returns on the new investment and the retention ratio in period t; that is

$$g_{et+1} = r_t(1-p_t). \tag{6.49}$$

The above analysis can be readily applied to the valuation model. For example, if we would like to use the constant-growth valuation model

$$V = D_0\left(\frac{1+g_e}{k-g_e}\right), \tag{6.50}$$

then the value can be given by

$$V = D_0\left(\frac{1+r(1-p)}{k-r(1-p)}\right). \tag{6.51}$$

The above expression shows that the value of the share is positively related to the return r. Recall

$$\frac{V}{E_0} = p\left(\frac{1+g_e}{k-g_e}\right). \tag{6.52}$$

This now can be expressed as

$$\frac{V}{E_0} = p\left(\frac{1+r(1-p)}{k-r(1-p}\right). \tag{6.53}$$

Thus the price-earnings ratio is very much dependent upon r and p.

6.14 Stock Valuation Based on Earnings

A firm will consider the investment and dividends decisions and evaluate earnings over time:

$$I_0, I_1, I_2, \ldots \tag{6.54}$$

$$E_1, E_2, \ldots \tag{6.55}$$

$$D_1, D_2, \ldots \tag{6.56}$$

There are three possibilities in period t after the initial investment:

$$(a) E_t = D_t + I_t \qquad (b) E_t < D_t + I_t \qquad (c) E_t > D_t + I_t \tag{6.57}$$

In case (a), the earnings can support both dividends and new investment. In case (b), the earnings will not be sufficient for dividends and new investment. Hence new capital $(D_t + I_t - E_t)$ is required for new business opportunities. The new capital can be obtained in the form of equity and/or bond financing (depending on the debt-equity ratio). In case (c), the earnings are sufficient for both dividends and new investment and part of the earnings can be used for reduction of existing capital.

Investment decisions are very much dependent on the growth of the economy, on the sector within which the firm operates, and on the strategic position that the firm has.

Dividends and investment decisions must be balanced. Higher dividends may constrain business growth. Low dividends may attract less capital.

The ways in which additional capital is acquired (debt and/or equity financing) will be affected by the existing debt-equity ratio. When this ratio is very low, the stock is attractive and the debt financing can be accommodated. When this ratio is very high, the stock is less attractive and the debt financing will be more difficult.

What will be the dividends level so that the shareholders will be better off? To answer this question, assume that a shareholder currently holds 1% of the common shares of the firm and wishes to maintain this percentage ownership in the future.

Consider time 0 and case (a). This shareholder will receive $.01D_0 = .01(E_0 - I_0)$ and can expect to receive $.01D_0 = .01(E_0 - I_0) = .01E_0 - .01I_0$.

Consider case (b). In this case the shareholder should invest in additional funds in common shares (if the equity financing is used). Let the additional equity be F_t such that

$$E_t + F_t = D_t + I_t. \tag{6.58}$$

The shareholder should invest $.01F_0$ at time 0. Thus, the net amount that the shareholder receives at time 0 is $.01D_0 - .01F_0$, that is

$$.01D_0 - .01(D_0 + I_0 - E_0) = .01E_0 - .01I_0. \tag{6.59}$$

Thus, the amount received in case (b) is the exactly same as in case (a). This is because the additional dividends received is offset by the additional amount invested to keep the shareholder's share in the capital.

Consider case (c). At time 0, the firm will buy back shares because there are less profitable business opportunities or because of some other reasons. Accordingly, the shareholder must sell some shares to the firm in order to avoid having an increased ownership position in the firm. As $E_0 > D_0 + I_0$, the total amount for buying back shares is R_0 such that

$$E_0 = D_0 + I_0 + R_0 \tag{6.60}$$

or

$$R_0 = E_0 - D_0 - I_0. \tag{6.61}$$

The number of the shares that the shareholder needs to sell back to the firm to maintain 1% position in the firm is $.01R_0 = .01(E_0 - D_0 - I_0)$. At time 0, the shareholder receives $.01D_0$ as the dividends and $.01R_0$ as the cash from the sales of some shares. The total amount is

$$.01D_0 + .01(E_0 - D_0 - I_0) = .01E_0 - .01I_0. \tag{6.62}$$

Thus, the amount that the shareholder receives is the same.

So if the shareholder wishes to keep the same position, the results will be the same at time 0 or in the future. After all, the earnings will determine the value of the shareholders $.01V$ for a given discount factor k if the shareholder keeps the same position. More specifically,

$$.01V = \frac{.01(E_0 - I_0)}{(1+k)^0} + \frac{.01(E_1 - I_1)}{(1+k)^1} + \frac{.01(E_2 - I_2)}{(1+k)^2} + \cdots \tag{6.63}$$

By multiplying both sides of the above equation by $\frac{1}{.01}$ we have

$$V = \frac{(E_0 - I_0)}{(1+k)^0} + \frac{(E_1 - I_1)}{(1+k)^1} + \frac{(E_2 - I_2)}{(1+k)^2} + \cdots \tag{6.64}$$

Clearly, the earnings will determine the value of the firm.

If the shareholder does not wish to keep the same position, things will be somewhat complex. The price at which shares are newly issued or bought back by the firm relative to the price at which original shares were issued become important factors. For example if the price of newly issued shares is much lower than the price of original shares and the shareholder does not buy additional shares to keep the ownership position, the ownership position and dividends of this shareholder will be diluted. If the price of bought back shares is much lower than the price of original shares and the shareholder does not reduce the holding to keep the same ownership position, the ownership position and dividends of this shareholder will be increased.

6.15 Determinants of Dividends

Since the earnings are variable over time, the payout ratio is determined over the longer run. Generally, dividends are fixed in a dollar amount. The adjustments upward and downward are made as the management of the firm projects an increasing or a decreasing earnings in the future.

Empirically, the existing research shows that firms are more likely to increase their dividends even if they had lower earnings in the previous year. When firms experienced two consecutive poor earnings, they tend to lower their dividends.

6.16 The Information Content of Dividends

It is reasonable to believe that the management of a firm has more information about the future earnings than the public do. When there is an incentive the management would like to seek to convey the information to the public.

The dividend announcement is a useful signaling device.

The dividend initiations and omissions also contain substantial information about the earnings.

6.17 Accounting Earnings versus Economic Earnings

Accounting earnings can be computed by the difference between the firm's revenues and its expenses. They can also be viewed as the change in the firm's book value of the equity plus dividends paid to shareholders.

The firm's accounting earnings divided by the number of its common shares outstanding is called earnings per share (EPS).

EPS divided by the book value per share is called the return on equity (ROE).

Economic earnings are somewhat different. They may be defined as the change in the economic value of the firm plus dividends paid to shareholders. The change in the economic value can be measured by the change in the market value of the firm's common shares.

The book value and market value of a firm are often different.

6.18 Price-Earnings Ratios

The price-earnings (PE) ratio can be used to determine if a stock is underpriced or overpriced.

The PE ratio is one of the useful indicators for selecting value and growth stocks.

Historically, the PE ratio of the market is not stable. The changes in PE ratios can be viewed as changes in permanent and transitory components of earnings. But generally both price and earnings of the market tend to go up over time. The price tends to move in the direction in which the market anticipates the future earnings will move. But the price will not react to the earning announcement fully and it will drift for a few months after the announcement.

Analysts appear to forecast earnings better than sophisticated mechanical models. Analysts tend to overestimate when forecasting earnings per share. The management can forecast earnings more accurately than analysts.

Alternatively, the EV/EBITDA ratio can be used. EV refers to enterprise value and indicates the value of a company's business rather than the company. EV is calculated as market cap plus debt, minority interest and preferred shares, minus total cash and cash equivalents. EBITDA refers to earnings before interest, tax, depreciation, and amortization.

The main advantage of the EV/EBITDA ratio over the PE ratio ratio is that it is unaffected by a company's capital structure. The stock of a company with a high (low) EV/EBITDA ratio can be interpreted as overvalued (undervalued). Therefore, this ratio can be used for direct comparison across the companies in the same sector.

6.19 Review Questions

1. What is the basic idea in the capitalization-of-income method of valuation?

2. Using the capitalization-of-income method of valuation, please define the situation where a financial asset is under-valued or over-valued.

3. What is the zero-growth model for stock prices?

4. What is the constant-growth model for stock prices? What is the limitation of this model?

5. What is the multiple-growth model for stock prices?

6. If the stock price is determined by the capitalization-of-income method of valuation, does it matter if the stock is sold in the second year? Why?

7. What is the payout ratio? What is the retention ratio?

8. How to define the growth rate of earnings? What factors affect the growth rate of earnings?

9. If we use the dividend discount model to evaluate the stock price, how would the following factors affect the price-earning ratio of a stock? (1) the payout ratio, (2) the growth rate of earnings, and (3) the required rate of return on capital.

Chapter 7

Financial Statements and Economic Conditions

7.1 Learning Objectives

- Analysis of Financial Statements
- Analysis of Economic Conditions

7.2 Analysis of Financial Statements

7.2.1 What Are the Financial Statements?

The major financial statements are the balance sheet, income statement, and statement of cash flows. Each firm should have them. The publicly-listed companies should have these statements made available to the public including shareholders and potential shares holders.

The balance sheet of a firm gives a snap shot of the assets that the firm controls and shows how these assets are financed via borrowing (loan and bonds) and owner's equity (capital, shares, and retained earnings).

The income statement contains information on sales, expenses, and earnings over a period of time.

The statement of cash flows combines the information from both the balance sheet and income statement to show how cash moves from operating activities (shown in the income statement) and from changes (in assets, liabilities, and owner's equity) in the balance sheet (shown from the changes of the balance sheets in two points in time).

There are three major entries in the statement of cash flows:

(1) cash flows from operating activities, which are defined as

$$Cash\ Flows\ from\ Operating\ Activities =$$

$$= Net\ Income + Noncash\ Revenue\ and\ Expenses$$

$$+\ Changes\ in\ Net\ Working\ Capital\ Items$$

(2) cash flows from investing activities, which are defined as increases (decreases) of own non-current fixed assets and equity investment in other firms. If these cash flows are negative, these represent the use of funds. Otherwise, these represent the source of funds.

(3) cash flows from financing activities, which are defined as increases (decreases) in liabilities and owner's equity.

As there are a number of ways that cash flows can be measured, we try to differentiate them here. *Traditional cash flow* equals net income plus depreciation expenses and deferred taxes. *Cash flow from operations* is the traditional cash flow adjusted for changes in operating (current) assets and liabilities that use or provide cash. *Free cash flow* is a modified concept of cash flow which recognizes some investing and financing activities. The *EBITDA* (earnings before interest, taxes, depreciation, and amortization) is another measure given in the income statement.

Typically, it is more useful and informational to show the *relative* financial statements where each entry is presented as certain percentage of the total quantity (total assets, total sales, and total cash flows). These can then be compared with the whole economy, a relevant industry, major competitors in the same industry, and past historical records.

7.2.2 Ratio Analysis of Financial Statements

In general, analysts focus on four dimensions in their analysis which are given below:

1. Internal Liquidity (Solvency)
2. Operating Performance
   (a) Operating Efficiency
   (b) Operating Profitability
3. Risk
   (a) Business Risk
   (b) Financial Risk
   (c) External Liquidity Risk
4. Growth

7.2.3 Internal Liquidity(Solvency)

$$Current\ Ratio = \frac{Current\ Assets}{Current\ Liabilities}.$$

$$Quick\ Ratio = \frac{Cash + Marketable\ Securities + Receivables}{Current\ Liabilities}.$$

$$Cash\ Ratio = \frac{Cash\ +\ Marketable\ Securities}{Current\ Liabilities}.$$

$$Receivable\ Turnover = \frac{Net\ Annual\ Sales}{Average\ Receivables},$$

where *Average Receivables* is defined as the average of the beginning and ending receivables values over the period (e.g. year).

$$Average\ Receivable\ Collection\ Period = \frac{365}{Annual\ Receivable\ Turnover}.$$

$$Inventory\ Turnover = \frac{Costs\ of\ Goods\ Sold}{Average\ Inventory},$$

where *Average Inventory* is defined as the average of the beginning and ending inventories over the period (e.g. year).

$$Average\ Inventory\ Processing\ Period = \frac{365}{Annual\ Inventory\ Turnover}.$$

$$Payable\ Turnover\ Ratio = \frac{Cost\ of\ Goods\ Sold}{Average\ Trade\ Payables}.$$

$$Payable\ Payment\ Period = \frac{365}{Payable\ Turnover}.$$

7.2.4 Operating Performance

Operating Efficiency Ratios

$$Total\ Asset\ Turnover = \frac{Net\ Sales}{Average\ Total\ Net\ Assets},$$

where *Average Total Net Assets* is defined the average of the beginning and ending net assets, which represent gross assets minus depreciation on fixed assets.

$$Fixed\ Asset\ Turnover = \frac{Net\ Sales}{Average\ Net\ Fixed\ Assets},$$

where *Average Net Fixed Asssets* is defined as the average of the beginning and ending net fixed assets.

$$Equity\ Turnover = \frac{Net\ Sales}{Average\ Equity},$$

where *Average Equity* is defined as the average of the beginning and ending equities.

Operating Profitability Ratios

$$Gross\ Profit\ Margin = \frac{Gross\ Profit}{Net\ Sales},$$

where *Gross Profit* equals net sales minus the cost of goods sold.

$$Operating\ Profit\ Margin = \frac{Operating\ profit}{Net\ Sales},$$

where *Operating profit* is defined as gross profit minus sales, general, and administrative (SG&A) expenses and is often referred to as earnings before interest and taxes (EBIT). A measure that is alternative to EBIT is to add back depreciation (D) and amortization (A) to EBIT. This is referred to as earnings before interest, taxes, depreciation, and amortization (EBITDA). This is a biased cash flow estimate.

$$Net\ Profit\ Margin = \frac{Net\ Income}{Net\ Sales},$$

where *Net Income* refers after-tax net income.

$$Return\ on\ Total\ Invested\ Capital = \frac{Net\ Income + Interest\ Expense}{Average\ Total\ Invested\ Capital},$$

where *Average Total Invested Capital* refers to the average of the beginning and ending interest-bearing debt and owner's equity.

$$Return\ on\ Total\ Equity = \frac{Net\ Income}{Average\ Total\ Equity},$$

where *Average Total Equity* refers to the average of the beginning and ending equity (including preferred stocks).

$$Return\ on\ Owner's\ Equity = \frac{Net\ Income}{Average\ Common\ Equity},$$

where *Average Common Equity* refers to the average of the beginning and ending owner's equity (represented by common stocks).

$$Return\ on\ Equity\ (ROE) = \frac{Net\ Income}{Common\ Equity}.$$

ROE is often used in the context of the DuPont System where ROE can be decomposed into three components: *Profit Margin*, *Total Asset Turnover*, and *Financial Leverage*. These are defined below:

$$Profit\ Margin = \frac{Net\ Income}{Net\ Sales},$$

$$Total\ Asset\ Turnover = \frac{Net\ sales}{Total\ Assets},$$

and

$$Financial\ Leverage = \frac{Total\ Assets}{Common\ Equity}.$$

It is apparent that

$$ROE = Profit\ Margin \times Total\ Asset\ Turnover \times Financial\ Leverage$$

or

$$ROE = \frac{Net\ Income}{Net\ Sales} \times \frac{Net\ Sales}{Total\ Assets} \times \frac{Total\ Assets}{Common\ Equity}.$$

This is the famous DuPont equation, which says that ROE is a function of *profit margin*, *total asset turnover*, and *financial leverage*. Because the net income over the net sales ratio is a function of *tax burden*, *interest burden*, and *EBIT%* as shown below

$$\frac{Net\ Income}{Net\ Sales} = \frac{Net\ Income}{Income\ before\ Taxes} \times \frac{Income\ before\ Taxes}{Income\ before\ Interest\ Payments\ \&\ Taxes}$$
$$\times \frac{Income\ before\ Interest\ Payments\ \&\ Taxes}{Net\ Sales},$$

we can have the five component decomposition of the famous DuPont equation:

$$ROE = \frac{Net\ Income}{Income\ before\ Taxes} \times \frac{Income\ before\ Taxes}{Income\ before\ Interest\ Payments\ \&\ Taxes}$$
$$\times \frac{Income\ before\ Interest\ Payments\ \&\ Taxes}{Net\ Sales}$$
$$\times \frac{Net\ Sales}{Total\ Assets} \times \frac{Total\ Assets}{Common\ Equity}.$$

The above extension further suggests that ROE is a function of *tax burden*, *interest burten*, *EBIT %*, *profit margin*, *total asset turnover*, and *financial leverage*.

7.2.5 Risk

Business Risk

$$Business\ Risk$$
$$= \textit{Coefficient of Variation of Operating Earnings (OE)}$$

$$= \frac{Standard\ Deviation\ of\ OE}{Mean\ of\ OE},$$

where the standard deviation and mean are computed using monthly/quarterly data over five or ten years.

$$Sales\ Volatility$$

$$= \textit{Coefficient of Variation of Sales (S)}$$

$$= \frac{Standard\ Deviation\ of\ S}{Mean\ of\ S},$$

where the standard deviation and mean are computed using monthly/quarterly data over five or ten years.

The above measure should be adjusted for growth companies by considering replace OE/S with the percentage change of OE/S (%ΔOE/%ΔS).

$$Operating\ Leverage = \frac{\sum \left| \frac{\%\Delta OE}{\%\Delta S} \right|}{No.\ of\ Periods}.$$

Financial Risk

Financial risk is intimately related to business risk. Some important ratios are given as follows.

$$Debt\ Equity\ Ratio = \frac{Total\ Long\ Term\ Debt}{Total\ Equity},$$

where *Total Long Term Debt* includes *Nocurrent Liabilities + Deferred Taxes + Present Value of Lease Obligations.* Please note that financial risk analysis should give a full consideration of leased assets and their depreciation.

$$Long\ Term\ Debt\ Total\ Capital\ Ratio = \frac{Total\ Long\ Term\ Debt}{Total\ Long\ Term\ Capital},$$

where *Total Long Term Capital* includes all long term debt, any preferred stock, and total equity.

$$Long\ Term\ Debt\ Total\ Capital\ Ratio = \frac{Total\ Long\ Term\ Debt}{Total\ Long\ Term\ Capital},$$

where *Total Long Term Capital* includes all long term debt, any preferred stock, and total equity.

$$Total\ Debt\ Total\ Capital\ Ratio = \frac{Total\ Interest\ Bearing\ Debt}{Total\ Invested\ Capital},$$

or

$$Total\ Debt\ Total\ Capital\ Ratio = \frac{Capitalized\ Leases\ +\ Noncurrent\ Liabilities}{Total\ Interest\ Bearing\ Debt\ +\ Shareholders'\ Equity}.$$

$$Interest\ Coverage\ Ratio = \frac{Income\ Before\ Interest\ and\ Taxes(EBIT)}{Debt\ Interest\ Charges},$$

or

$$Interest\ Coverage\ Ratio = \frac{Net\ Income\ +\ Income\ Taxes\ +\ Interest\ Expenses}{Interest\ Expense}.$$

$$Fixed\ Financial\ Cost\ Coverage\ Ratio = \frac{EBIT\ +\ Implied\ Lease\ Interest}{Gross\ Interest\ Expense\ +\ Implied\ Lease\ Interest}.$$

$$Cash\ Flow\ Coverage\ Ratio$$
$$= \frac{\begin{array}{c}Net\ Cash\ Flow\ from\\ Operating\ Activities\end{array}\ +\ Interest\ Expense\ +\ Implied\ Lease\ Interest}{Interest\ Expense\ +\ Implied\ Lease\ Interest}.$$

$$Cash\ Flow\ Long\ Term\ Debt\ Ratio$$
$$= \frac{Cash\ Flow\ from\ Operating\ Acvitities}{Book\ Value\ of\ Long\ Term\ Debt\ +\ Present\ Value\ of\ Lease\ Obligations}.$$

$$Cash\ Flow\ Total\ Debt\ Ratio$$
$$= \frac{Cash\ Flow\ from\ Operating\ Acvitities}{Total\ \ Long\ Term\ Debt\ +\ Interest\ Bearing\ Current\ Liabilities}.$$

External Liquidity Risk

Liquidity refers to the ability to buy or sell an asset quickly. When external liquidity risk presents, it will be difficult to buy or sell the asset quickly. Liquidity is determined by the number of shares and the market value of these shares. The trading turnover is an good indicator of liquidity and it is defined as

$$Trading\ Turnover = \frac{Number\ of\ Shares\ Traded}{Number\ of\ Shares\ Outstanding}.$$

Another measure of liquidity is the bid–ask spread which is defined as the difference between the bid and ask prices. A small spread indicates liquidity while a larger one reflects the lack of liquidity.

7.2.6 Growth

The growth of the underlying business of a firm affects the future growth of the firm value. But the growth must be financed by new investments which will increase the business and its efficiency.

The payout ratio (PR) captures the percentage of after-tax earnings paid out as dividends and that is defined by

$$PR = \frac{Dividends\ Paid\ Out}{Operating\ Income\ after\ Taxes}$$

while the retention ratio (RR) shows the percentage of after-tax earnings retained for investment purposes and that is given by

$$RR = 1 - PR.$$

In the firm, RR determines how much new investment will be financed by after-tax earnings and ROE measures the effectiveness of the new investment is measured by ROE. We can measure the growth of the firm g by

$$g = (RR)(ROE).$$

and ROE can be decomposed in the DuPont System as explained previously.

7.2.7 Specific Uses of Financial Ratios

Stock Valuation Models

For stock valuation models, consider the following indicators over the 5-10 year period.

1. Financial Ratios
   (a) Average debt equity
   (b) Average interest coverage
   (c) Average dividend payout
   (d) Average return on equity
   (e) Average retention rate
   (f) Average market price to book value
   (g) Average market price to cash flow
   (h) Average market to sales
2. Variability Measures
   (a) Coefficient of variation of operating earnings
   (b) Coefficient of variation of sales
   (c) Coefficient of variation of net income
   (d) Systematic risk (beta)
3. Non-ratio Variables
   (a) Average growth rate of earnings

Systematic Risk Models

For systematic risk models, consider the following indicators over the 5 year period.

1. Financial Ratios
   (a) Dividend payout
   (b) Total debt total assets
   (c) Cash flow total debt
   (d) Interest coverage
   (e) Working capital total assets
   (f) Current ratio
2. Variability Measures
   (a) Coefficient of variation of net earnings
   (b) Coefficient of variation of operating earnings
   (c) Coefficient of variation of operating profit margins
   (d) Operating earnings beta (company earnings relative to aggregate earnings of the capital market portfolio)
3. Non-ratio Variables
   (a) Asset size
   (b) Market value of stock outstanding

Bond Rating Models

For bond rating models, consider the following indicators.

1. Financial Ratios
   (a) Long term debt total assets
   (b) Total debt total capital
   (c) Net income plus depreciation (cash flow) long term senior debt
   (d) Cash flow total debt
   (e) Earnings before interest and taxes (EBIT) interest expense (fixed charge coverage)
   (f) Cash flow from operations plus interest interest expense
   (g) Market value of stock par value of bonds
   (h) Net operating profit sales
   (i) Net income owners' equity (ROE)
   (j) Net income total assets (ROA)
   (k) Working capital sales
   (l) sales net worth (equity turnover)
2. Variability Measures
   (a) Coefficient of variation of sales
   (b) Coefficient of variation of net earnings
   (c) Coefficient of variation of return on assets
3. Non-ratio Variables
   (a) subordination of the issue
   (b) Size of the firm (total assets)
   (c) Issue size
   (d) Par value of all publicly traded bonds of the firm

Default/Bankruptcy Models

For default/bankruptcy models, consider the following indicators.
Financial ratios

1. Cash flow total debt
2. Cash flow long term debt
3. Sales total assets (*)
4. Net income total assets
5. EBIT total assets (*)
6. Total debt total assets

7. Market value of stock book value of debt (*)
8. Working capital total assets (*)
9. Retained earnings total assets (*)
10. Current ratio
11. Working capital sales

Please note that (*) indicates the ratios used in the well-known Altman Z-score model (Altman, 1968).

7.3 Analysis of Economic Conditions

7.3.1 Macroeconomic Conditions

Macroeconomic conditions refers to the general economic environment in which a firm operates. Macroeconomic conditions include both domestic and international macroeconomic conditions. Macroeconomic conditions reflect whether or not the economy is growing and at what rate it is growing. In order to appreciate the where the economy is heading for, we need to understand what happen to prices, wages, interest rates, asset prices, and exchange rates in the labor, goods and services, money and capital, and foreign exchange markets, respectively. We also need to understand the monetary and fiscal policies which will ultimately influence the economy.

In the goods and services market, the aggregate demand for and supply of goods and services interact so that the equilibrium price and output level are determined. If the aggregate quantity demanded is too high (low) relative to the aggregate quantity supplied, we will expect a rise (fall) in price and therefore inflation (deflation).

The dynamics in the goods and service market will also find its impact on the labor market. The hot (cold) goods and services market will lead the active (depressed) labor market and higher (lower) wages and salaries.

The dynamics in the goods and service market will also find its impact on the capital market. The capital market plays an important role of financial intermediation. The demand for and supply of loanable funds will determine interest rates. When loanable funds are too little (much) relative to the demand, interest rates rise (fall).

In the capital market, stocks and bonds are traded at prices that approximate their present values of future cash flows. For a given cash flow, high (lower) current/future interest rates will depress (buoy) asset prices.

Both monetary policy and fiscal policy can also influence the economy.

Monetary policy can influence the economy by changing the very short term interest rate so that the term structure of interest rates and the exchange rates will be affected. Lower (higher) interest rates will boost (suppress) the aggregate demand.

Lower (higher) interest rates will depreciate (appreciate) the local currency with reference to the foreign currency and further increase (decrease) exports and decrease (increase) imports. Ultimately these changes will regulate the aggregate demand and international trade and capital flows.

Fiscal policy can also influence the economy by boosting (suppressing) the aggregate demand with higher (lower) public transfer and expenditure and lower (higher) taxation.

Within a reasonable range, asset values should reflect the present values of cash flows generated by the underlying assets and discounted by current and future interest rates.

Generally, increasing (decreasing) cash flows and/or falling (rising) interest rates will cause asset values to rise (fall).

Stocks are financial assets which are claims to underlying real assets. Therefore stock prices reflect values of the underlying assets. Stock prices are closely related to, and influenced by, domestic and international macroeconomic conditions.

7.3.2 Economic Series

Macroeconomic conditions can be measured by cyclical indicators, which can be further classified into leading, coincident and lagging indicators (Conference Board Indicators):

1. Leading indicators
   (a) Average weekly hours of manufacturing workers
   (b) Average weekly initial claims for unemployment insurance
   (c) Real value of manufacturers' new orders for consumer goods and materials
   (d) Index of consumer expectations
   (e) Index of 500 common stock prices
   (f) Manufacturers' new orders, non-defense capital goods
   (g) Index of new private housing starts authorized by local building permits
   (h) Vendor performance (the percentage of companies receiving delivery later than industry average)
   (i) Real money supply, M2
   (j) Interest rate spread, ten-year Treasury bond interest rate less the federal fund rate
2. Coincident indicators
   (a) Number of employees on non-agricultural payrolls
   (b) Personal income less transfer payments
   (c) Index of industrial production
   (d) Manufacturing and trades sales
3. Lagging indicators
   (a) Average duration of unemployment
   (b) Ratio of manufacturing and trade inventories to sales
   (c) Percentage change in the labor cost per unit of output in manufacturing
   (d) Average prime rate charged by commercial banks
   (e) Commercial and industrial loans outstanding
   (f) Ratio of consumer installment credit outstanding to personal income
   (g) Change in the consumer price index (inflation) for services

Chapter 8

Options

8.1 Learning Objectives

- Types of option contracts
- Option trading
- Valuation of options
- Market completeness and options
- The binomial option pricing model
- The Black-Scholes model for call options
- The valuation of put options
- Strategies of option market makers

8.2 Types of Option Contracts

Two most basic types of contracts are known as calls and puts.

Call Options

A call option gives the buyer the right to buy ("call away") a specific number of shares of a specific company from the option writer at a specific purchase price at any time up to and including a specific date.

The call buyer has to pay for the call writer for having the right. The amount paid is called the premium.

The existing call contracts can be traded. The premium changes as the market conditions changes.

Options clearings.

Put Options

A put option gives the buyer the right to sell ("put away") a specific number of shares of a specific company to the option writer at a specific selling price at any time up to and including a specific date.

Similarly, the put buyer has to pay for the put writer for having the right. The amount paid is also called premium.

When the option buyer and writers are trading, they in fact establish the contract relationship with the option cleaning authorities and hence they have to honor their contractual obligations.

In the event that the stock-split occurs, the option contract will be modified automatically to reflect the stock-split. The cash dividend will have no impact on the option contract although it will affect the premium.

An at-the-money call (put) option is an option whose exercise price is roughly equal to the market price of its underlying asset. An out-of-the-money call (put) option is a call (put) option whose exercise price is greater than (less than) the market price of its underlying asset. An in-the-money option is a call (put) option whose exercise price is less than (greater than) the market price of its underlying asset.

8.3 Option Trading

The regular options and long-term equity participation securities are traded according to different cycles.

The buyers and writers can place their orders as market, limit, and stop orders with their brokers. In general, the option buyers will not be able to use margins while the option writers should have the backing assets to meet the financial obligation specified in option contracts.

The option trading commission is somewhat lower. The option settlement takes a shorter time (one day generally).

In the exchange, the market-makers for options specialize in certain options. They play two roles as: dealers and brokers. As dealers, they set the bid and ask prices and keep an inventory of the options and the underlying stocks that are assigned to them. They buy into and sell from the inventory as required to maintain a market. As brokers, they keep the limit order book and execute the orders in it as market prices move up and down.

8.4 Valuation of Options

Valuation at Expiration

At expiration, options have intrinsic value IV, the value if the option is exercised immediately. Let the market price of the underlying stock be P_s and the exercise price of the option E. Then the intrinsic value of a call option IV_c can be expressed as

$$IV_c = \max\{0, P_s - E\}. \tag{8.1}$$

The intrinsic value of a put option IV_p can be expressed as

$$IV_p = \max\{0, E - P_s\}. \tag{8.2}$$

Profit-and-Loss Profile of Call and Puts

We will discuss the following cases

(a) Buy a call

(b) Write a call

(c) Buy a put

(d) Write a put

(e) Buy a put and a call (i.e., straddle)

(f) Write a put and a call (i.e., straddle)

(g) Buy a stock

(h) Short sell a stock

(i) Buy a stock and write a call

(j) Short sell a stock and buy a call

Discussion:

Cases (a) and (b) are mirror images to each other. The same can be said for cases (c) and (d). For case (a) the profit of a call option, π_c, can be expressed as a function of the call premium P_c, the market price of the underlying stock P_s and the exercise price of the option E:

$$\begin{aligned} \pi_c &= IV_c - P_c \\ &= \max\{0, P_s - E\} - P_c \\ &= \max\{-P_c, P_s - E - P_c\}. \end{aligned} \tag{8.3}$$

For case (c) the profit of a put option, π_p, can be expressed as a function of the put premium P_p, the market price of the underlying stock P_s and the exercise price of the option E:

$$\begin{aligned}\pi_p &= IV_p - P_p \\ &= \max\{0, E - P_s\} - P_p \\ &= \max\{-P_p, E - P_s - P_p\}.\end{aligned} \tag{8.4}$$

The option trading is a zero-sum game. Hence. It is not surprising that cases (b) and (d) are mirror images of cases (a) and (c)

Cases (e) and (f) are more complex option strategies. Case (e) has a maximum loss of $P_c + P_p$ (plus commissions). The straddle buyer expects the large price movements for the underlying stock before the options are expired. Case (f) is an mirror image of case (e). Theoretically, the straddle writer has limited gain, that is $P_c + P_p$ (minus commissions), when the price of the underlying stock does not change a lot.

Cases (g) and (h), buy a stock or short sell a stock, are mirror images to each other. The loss may occur when price falls below (rise above) the purchasing price when the investor has a long (short) position. This indicates the potential risk.

Cases (i) and (j) represent the cases where the long and short positions are combined with some option strategies. Case (i) represents a cover call writing strategy. A cover call writing strategy works as follows. If the price of the underlying stock rises to or above the exercise price of the option, the call buyer will exercise the call option and the call writer must sell the underlying stock. If the price falls below the exercise price, the call option expires worthlessly so that the call buyer loses and the call writer keeps the premium. The cover call writing is often used when the investor is moderately bullish about the underlying stock. Case (j) represents a combination of short selling and a call buying strategies. When the investor expects the fall of the price of the underlying stock but feels that there is some chance that the price may rise, he or she may short the stock and buy a call at the same time. Using this combination, the investor can make a profit when the price falls and a limited loss (P_c plus commission) when price rises.

8.5 Market Completeness and Options

There is a theoretical literature on market incompleteness. If we have n future states but we do not have n state-dependent securities and therefore n state-dependent payoffs, then it is possible that when one of the future states realizes, there will be no corresponding payoff. The real life example is that farmers may have average or better crops when there is an average or sufficient rainfall but they will get virtually nothing when there is a drought.

In a complete market, we have n state-dependent securities in the n future state world. Now assume we only have three future states s_1, s_2, and s_3 and three state-dependent securities θ_1, θ_2, and θ_3. The payoffs can be illustrated by a matrix

$$\begin{array}{cccc} & \theta_1 & \theta_2 & \theta_3 \\ s_1 & a_{11} & a_{12} & a_{13} \\ s_2 & a_{21} & a_{22} & a_{23} \\ s_3 & a_{31} & a_{32} & a_{33} \end{array} \tag{8.5}$$

A specific example of the above is

$$\begin{array}{cccc} & \theta_1 & \theta_2 & \theta_3 \\ s_1 & 1 & 0 & 0 \\ s_2 & 0 & 1 & 0 \\ s_3 & 0 & 0 & 1 \end{array} \tag{8.6}$$

In this complete market, if state s_i occurs, security θ_i will pay 1 but other securities will pay 0. More generally, we may have various types of payoffs such as

$$\begin{array}{cccc} & \theta_1 & \theta_2 & \theta_3 \\ s_1 & 1 & 0 & 9 \\ s_2 & 3 & 5 & 4 \\ s_3 & 0 & 0 & 1 \end{array} \tag{8.7}$$

or

$$\begin{array}{cccc} & \theta_1 & \theta_2 & \theta_3 \\ s_1 & 1 & 0 & 0 \\ s_2 & 2 & 1 & 0 \\ s_3 & 3 & 0 & 1 \end{array} . \tag{8.8}$$

From the above two payoff matrices, we cannot tell immediately whether or not a market is complete. In these cases, how can we tell if a market is complete? The general strategy is to name the payoff as matrix $\mathbf{A} = \{a_{ij}\}$. If the determinant of $\mathbf{A}$, $|A|$, does not equal to 0, then the inverse of $\mathbf{A}$, $\mathbf{A}^{-1}$, exists. If a payoff matrix has an inverse, we can always convert it into the identity matrix by premultiplying $\mathbf{A}$ by $\mathbf{A}^{-1}$. That is, $\mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$.

Further, if we know that the market is incomplete, would it be possible to make it complete? The answer is positive. Let us consider an example where there is one security θ_1 but there are three future states. This market is obviously incomplete.

$$\begin{array}{cccc} & \theta_1 & ? & ? \\ s_1 & 1 & 0 & 0 \\ s_2 & 2 & 0 & 0 \\ s_3 & 3 & 0 & 0 \end{array} . \tag{8.9}$$

Can we make this market complete by using options? We can use two call options θ_2 and θ_3 at the strike prices 1 and 2, respectively with θ_1 as the underlying security. Then we have the new payoff matrix with addition of θ_2 and θ_3:

$$\begin{array}{cccc} & \theta_1 & \theta_2 & \theta_3 \\ s_1 & 1 & 0 & 0 \\ s_2 & 2 & 1 & 0 \\ s_3 & 3 & 2 & 1 \end{array} . \tag{8.10}$$

As the determinant of $\mathbf{A}$ is not equal to 0, we can say that the matrix is invertible and that the market is now complete. Now we turn to the story of farmers. When drought leads to no yield to farmers, they need some kind of drought insurance policies. If they have such insurance policies, they would receive some compensation when drought occurs. These insurance policies make farmers future prospects somewhat certain when the most undesirable state occurs—in theoretical sense, these insurance policies make the market complete!

8.6 The Binomial Option Pricing Model

There are two types of options. A European option is an option that can only be exercised on its expiration date while an American option is an option that can be exercised any time through its expiration date. Normally, the options on individual stocks are American options while the some options on stock exchange indices are European options.

The pricing model of a European option is easier than the pricing model of an American option. The binomial option pricing model (BOPM) is for an European option. In the following we can use an specific example to explain the idea using the call option.

Let the price of ABC at time $t = 0$ be \$100. Let the expiration date be one year, that is $t = T$. One can image that at $t = T = 1$ there are two possible states: in one state the price of ABC is \$125 and in another the price is \$80. If one is to buy a call option at $t = 0$ with an exercise price \$100, what will be the fair price?

To solve the option pricing problem, we need to consider other possible investment opportunities. In addition to the stock investment into ABC, it is also possible to invest in a risk-free bond with the interest rate 8%. If one is to invest \$100 into the bond, the risk-free bond will grow to approximately \$108.33=\$100$e^{.08T}$. The payoff profile can be given below

Security	Payoff(up)	Payoff(down)	Current Price
Share ABC	\$125	\$80	\$100
Bond	\$108.33	\$108.33	\$100
Call Option ABC \$100	\$25	\$0	???

How can we price the call option? If we can replicate a portfolio which gives the payoffs of the call option for ABC \$100, then the price of the portfolio must be the current price of the option. Otherwise, it will generate an arbitrage opportunity. Under the condition of no arbitrage opportunity, we can price the option directly.

We claim that we can form this portfolio using a proportion of stocks, N_s , and a proportion of bonds, N_b so that the payoffs of the option equals $N_sP_s + N_bP_b$ in both up state and down state. The question now is to how to determine N_s and N_b. Based on the information given in the above table, we have

$$\$125N_s + \$108.33N_b = \$25 \tag{8.11}$$

for the up state and

$$\$80N_s + \$108.33N_b = \$0 \tag{8.12}$$

for the down state. This is a system of two linear equations with two unknowns. We can get

$$N_b = -\frac{\$80N_s}{\$108.33} \tag{8.13}$$

from the second linear equation in the system and substitute it to the first equation. This gives us

$$(\$125 - \$80)N_s = \$25 \tag{8.14}$$

or

$$N_s = .5556. \tag{8.15}$$

By substituting the above to any of the two equations in the system we get

$$N_b = -.4103. \tag{8.16}$$

Now we have both N_s and N_b so that we can verify if the payoffs of the portfolio are the same of the option. In the up state,

$$\$125 \times .5556 + \$108.33 \times (-.4103) = \$69.45 - \$44.45 = \$25. \tag{8.17}$$

In the down state,

$$\$80 \times .5556 + \$108.33 \times (-.4103) = \$44.45 - \$44.45 = \$0. \tag{8.18}$$

Indeed, the portfolio can replicate the payoffs of the option exactly. The value of the option, therefore must be

$$V_c = \$100 \times .5556 - \$100 \times .4103 = \$100 \times .1453 = \$14.53. \tag{8.19}$$

The Hedge Ratio of the Call Option

Recall that the portfolio, which mimics the payoffs of the call option, is composed of .5556 shares of ABC and \$41.04 borrowed money (in form of short selling of bonds). For every \$1 change in the price of ABC, there will be a \$.5556 change in the value of the portfolio or the call option value. This relationship is called the option's hedge ratio. The hedge ratio of the call option can be computed by

$$h_c = \frac{P_{cu} - P_{cd}}{P_{su} - P_{sd}} = \frac{\$25 - \$0}{\$125 - \$80} = .5556. \tag{8.20}$$

Here the subscript u refers to "up state" while the subscript d refers to "down state." This h_c also represents the amount of shares of ABC must be purchased. Simultaneously, an amount must be risklessly borrowed by short selling bonds; this amount is

$$B = PV(h_c P_{sd} - P_{cd}) = e^{-0.08}(.5556 \times \$80 - \$0) = \$41.03 \tag{8.21}$$

where $PV(x)$ represents the present value of x. The value of the call option is given by

$$V_c = h_c P_s - B = h_c P_s - PV(h_c P_{sd} - P_{cd}) = \$55.56 - \$41.03 = \$14.53. \tag{8.22}$$

This is identical to the result computed previously.

We can also use the binomial option pricing model to price the put option (let the put option of ABC expires $t = T = 1$ with an exercise price \$100). At the end of the period, there are two possible ending states

(a) the price goes to \$125 and the put option price will be \$0;

(b) the price goes to \$80 and the put option price will be \$20.

Compute the hedge ratio of the put option

$$h_p = \frac{P_{pu} - P_{pd}}{P_{su} - P_{sd}} = \frac{\$0 - \$20}{\$125 - \$80} = -.4444. \tag{8.23}$$

This negative number indicates that a rise in the price of the stock will lower the price of the put option.

Compute the "borrowing" amount B

$$B = PV(h_p P_{sd} - P_{pd}) = e^{-0.08}\,(-.4444 \times \$80 - \$20) = -\$51.28. \tag{8.24}$$

This is a negative number indicating "lending."

To replicate the put option, one sells short .4444 shares of ABC and lends \$51.28 at the rate of 8%. The net cost of this portfolio is \$51.28 – \$44.44 = \$6.84. This is the fair value of the put option.

Table 8.1: Arbitrage Table Showing the Call-Put Parity Condition

Transaction	Current Date Cash Flow	Future $P_s^* \leq E$	Date $P_s^* > E$
Write call	P_c	—	$E - P_s^*$
Buy put	$-P_p$	$E - P_s^*$	—
Buy stock	$-P_s$	P_s^*	P_s^*
Borrow	Ee^{-rT}	$-E$	$-E$
Total	0	0	0

Given the call (put) option hedge ratio h_c (h_p), we can find the put-call parity condition:

$$h_c - 1 = h_p. \tag{8.25}$$

In our example,

$$.5556 - 1 = -.4444. \tag{8.26}$$

Now let us discuss the call-put parity condition which gives the insight to how the pricing of a call option is connected to the pricing of a put option when both share the same exercise price and expiration date. Consider two strategies: (a) buy a share of ABC and a put option for ABC \$100—a protective put strategy and (b) buy a call option \$100 and invest an amount of money in the risk-free asset that is equal to the present value of the exercise price. At the expiration date, two strategies have the identical payoffs: When the price is below \$100, both strategies give a payoff of \$100. When the price is equal to \$100, both have the same payoff of \$100. When the price is above \$100, both strategies end up with one share of ABC with the above \$100 market price. Since these two strategies have the identical payoffs, both portfolios must cost the same amount in equilibrium. Let P_p and P_c be the prices of the put and call options, respectively. The relationship between the market prices of a call option and a put option with the same exercise price and expiration date is:

$$P_p + P_s = P_c + E/e^{rT}. \tag{8.27}$$

The above parity can be further explained by the arbitrage argument via Table 8.1. Let P_s^* to denote the future price of a stock. As we can see from the table that the future payoffs of the set of transactions are zero. Hence the total price of these transactions are zero; that is

$$P_c - P_p - P_s + E/e^{rT} = 0. \tag{8.28}$$

This is an equivalent way to write the parity condition.

8.7 The Black-Scholes Model for Call Options

If one allows the number of periods before the expiration to increase, the binomial option pricing model will converges to the Black-Scholes option pricing model (BSOPM).

Let

V_c = price of the call option

P_s = current market price of the underlying stock

E = exercise price of the option

r = continuously compounded risk-free rate of return expressed on an annual basis

T = time remaining before expiration, expressed as a fraction of a year

σ = risk of the underlying common stock, measured by the standard deviation of the continuously compounded annual rate of return on the stock

E/e^{rT} represents the present value of the exercise price where a continuous discount rate is used. $\ln(P_s/E)$ is the natural logarithm of P_s/E. $N(d_1)$ and $N(d_2)$ are the probabilities of the standard normally distributed outcomes (with mean 0 and variance 1) that are less than d_1 and d_2 respectively.

Given the above information, the Black-Scholes option pricing formula is

$$V_c = N(d_1)P_s - \frac{E}{e^{rT}}N(d_2) \tag{8.29}$$

where

$$d_1 = \frac{\ln(P_s/E) + (r + .5\sigma^2)T}{\sigma\sqrt{T}} \tag{8.30}$$

$$\begin{aligned} d_2 &= \tfrac{\ln(P_s/E)+(r-.5\sigma^2)T}{\sigma\sqrt{T}} \\ &= d_1 - \sigma\sqrt{T} \end{aligned} \tag{8.31}$$

One difficulty for computing V_c is the function $N(\cdot)$ that is called the cumulative standard normal distribution. It can be approximated in the following formula:

$$x = 1 - z(1.330274y^5 - 1.821256y^4 + 1.781478y^3 - .356538y^2 + .3193815y) \tag{8.32}$$

where

$$y = \frac{1}{1 + .2316419|d|} \tag{8.33}$$

and

$$z = .3989423e^{-d^2/2}. \tag{8.34}$$

If $d > 0$, then $N(d) = x$; if $d < 0$, then $N(d) = 1 - x$.

Example: $T = 60/365 = .16438$. $E = \$50$. $P_s = 45$. $r = 10\%$. $\sigma = 30\%$.

$$d_1 = \frac{\ln(45/50) + (.1 + .5 \times (.3)^2) \times .16438}{.3 \times \sqrt{.16438}} = -.67025 \tag{8.35}$$

$$d_2 = -.67025 - .3\sqrt{.16438} = -.79189 \tag{8.36}$$

$$N(d_1) = N(-.67025) = 1 - .74865 = .25134 \tag{8.37}$$

$$N(d_2) = N(-.79189) = 1 - .78579 = .21421 \tag{8.38}$$

$$V_c = .25134 \times 45 - \frac{50}{e^{-.1\times.16438}} \times .21421 = .7746. \tag{8.39}$$

Comparison between the BSOPM and the BOPM

BOPM

$$V_c = h_cP_s - B \tag{8.40}$$

where $B = PV(h_cP_{sd} - P_{od})$.

BSOPM

$$V_c = N(d_1)P_s - \frac{E}{e^{rT}}N(d_2). \tag{8.41}$$

We can find the following corresponding relationship:

The hedge ratio of the call option:

$$h_c \Leftrightarrow N(d_1) \tag{8.42}$$

is the number of shares that an investor would need to purchase in executing an investment strategy that was designed to have the same payoffs as the call option.

$$B \Leftrightarrow \frac{E}{e^{rT}} N(d_2) \tag{8.43}$$

is the amount of money that the investor borrows as the other part of the strategy. That means that $EN(d_2)$ is the face amount of the loan and $\frac{E}{e^{rT}}N(d_2)$ is the present value of the loan.

Now discuss static analysis:

1. $P_s \uparrow \Rightarrow V_c \uparrow$
2. $E \uparrow \Rightarrow V_c \downarrow$
3. $T \uparrow \Rightarrow V_c \uparrow$
4. $r \uparrow \Rightarrow V_c \uparrow$
5. $\sigma \uparrow \Rightarrow V_c \uparrow$

Of these five factors, P_s, E, and T are known. r can be estimated by using the yield-to-maturity on a Treasury bill having a maturity date close to the expiration date of the option. σ is not readily available. It must be estimated.

Estimating σ from historical prices: (1) collect the stock prices for n period (e.g. 52 weeks); (2) compute the weekly return using $r_t = \ln(P_{st}/P_{st-1})$ for $t = 1, 2, \ldots, n$); (3) compute the per-period average return using

$$r_{av} = \frac{1}{n}\sum_{t=1}^{n} r_t \tag{8.44}$$

the pre-period standard variance using

$$s^2 = \frac{1}{1-n}\sum_{t=1}^{n}(r_t - r_{av})^2; \tag{8.45}$$

(4) estimate the annual (52 weeks) variance σ^2 using $52s^2$; and (5) find the estimated annual volatility $\sqrt{52s^2}$.

Another way to estimate the risk of a stock is to assume the option pricing model is correct and the option is priced to reflect the market consensus about the stock's risk. When P_s, E, T, r, and V_c are given, σ can be solved as an unknown from the BSOPM. The volatility computed as such is called implicit or implied volatility.

Please note that the BSOPM has a few limitations: (1) It only applies to European options not American options and (2) It does not consider the dividend payment.

For a non-dividend-paying stock, it can be shown that it is unwise for an investor holding an American call to exercise it prior to maturity. Thus, one can use the BSOPM for American options.

Generally, other things being equal, the greater the amount of the dividends to be paid during the life of a call option, the lower the value of the call option will be. This is because the greater the dividend that a firm declares, the lower its stock price will be. Since options are not "dividend protected," this lower stock price will lead to the lower value of the call option (and a higher value for put options).

8.8 The Black-Scholes Model for Put Options

Recall

$$P_p + P_s = P_c + E/e^{rT}. \tag{8.46}$$

Rearrange the equation for P_p:

$$P_p = P_c + E/e^{rT} - P_s. \tag{8.47}$$

Recall

$$V_c = N(d_1)P_s - \frac{E}{e^{rT}}N(d_2). \tag{8.48}$$

Replace V_c with P_c:

$$P_c = N(d_1)P_s - \frac{E}{e^{rT}}N(d_2). \tag{8.49}$$

Substitute the above into the equation for P_p:

$$P_p = N(d_1)P_s - \frac{E}{e^{rT}}N(d_2) + E/e^{rT} - P_s. \tag{8.50}$$

This can be simplified into

$$P_p = \frac{E}{e^{rT}}\left(1 - N(d_2)\right) - P_s\left(1 - N(d_1)\right). \tag{8.51}$$

Because $1 - N(d_2) = N(-d_2)$ and $1 - N(d_1) = N(-d_1)$. Replace P_p with V_c and the BSOPM for a put option can be written as

$$V_p = \frac{E}{e^{rT}}N(-d_2) - P_sN(-d_1). \tag{8.52}$$

Now discuss static analysis:

1. $P_s \uparrow \Rightarrow V_p \downarrow$
2. $E \uparrow \Rightarrow V_p \uparrow$
3. $T \uparrow \Rightarrow V_p \uparrow$
4. $r \uparrow \Rightarrow V_p \downarrow$
5. $\sigma \uparrow \Rightarrow V_p \uparrow$

8.9 Strategies of Option Market Makers

Few, market makers if any, simply buy calls or sell puts when they are bullish and buy puts or sell calls when they are bearish. While most market makers will “scalp” or “leg” into spreads on a short-term basis, trying to take advantage of moves in the underlying prices is not generally their long-term strategy. The risk of simply taking directional bets, or taking on any one kind of exposure for that matter, is just too great; those who do don’t survive over the long run.

All market makers attempt to control the risks of their positions, most of them by spreading options against other options or the underlying stock or index futures. Nearly every market maker is looking for a synthetic arbitrage trade—a trade that can be combined with other trades to produce a profit with very low risk.

To do this the market maker must know what is mispriced. In addition he needs to know how to hedge away the unwanted risks. If the market maker can enter two or more offsetting trades that cancel out the risk and do this for a net profit then he solved both problems.

Relative Pricing and Arbitrage Spreads

Market makers often don't need to worry about whether an option is actually over priced or under priced in some absolute sense. What matters is whether an option is mispriced relative to the underlying stock or to other options at any given point in time so they can create a spread and reduce the risk of buying or selling the option.

There are few basic arbitrage spreads that determine the price relationships that the underlying stocks and their various options should have to each other. When these price relationships don't hold, there is an opportunity for profit. Market makers quickly learn to think in terms of synthetic equivalents—alternative ways of constructing a position. By buying relatively under priced and selling relatively overpriced combinations of puts, calls and stock that have the same risk exposure at the same time, the market maker can take advantage of any mispricing and cancel out his risk. In fact, this basic pricing technique is fundamental part of the way that market makers operate. Looking at option positions in terms of synthetic equivalents tells the market maker his alternatives. It is also key to understanding market makers ability to buy and sell options when the market appears to be heading in one direction and presumably no one else would take the other side.

An example:

Suppose ABC shares are trading at $10.

1. If you own the stock, you gain and lose a dollar for every dollar the stock rises/falls above/below $10

2. If you own a $10 call, at expiry the call is worth a dollar for every dollar above $10

3. If you're short a $10 put, you position is has lost a dollar for every dollar below $10

So the combination of 2 and 3 a long call and a short put, is synthetically equivalent to 1, owning stock. If the stock price rises a dollar the call is worth a dollar for every dollar above $10 and the short put is worthless. If the stock falls, the long call is worthless and the short put loses a dollar for every dollar below $10.

At expiry the combination of 2 & 3 (long call, short put) will show the same net gain or loss with any change in the stock price. Thus by buying one and selling the other you can eliminate the most significant form of position risk, exposure to the direction of price movement. Buying stock and selling synthetic stock, or the reverse, results in no net direction exposure. The positions cancel because what you make on one you lose on the other.

Not only is there a synthetic equivalent for owning stock, there is a synthetic equivalent for any option or stock position (see Table 8.2).

Conversions and Reversals

The two most basic forms of option arbitrage are the "conversion" and reverse conversion or "reversal". If a market maker can buy stock and sell synthetic stock (or the reverse) for a net price difference that will more than cover his costs, then the combination of trades ought to make a profit with no directional risk. What matters is not the price of the call or the put or the stock itself in isolation, but the relative prices of the offsetting pieces.

An example:

Table 8.2: Synthetic Equivalent Securities

Position	Synthetic Equivalent
Long Stock	Long Call + Short Put
Short Stock	Short Call + Long Put
Long Call	Long Stock + Long Put
Long Put	Short Stock + Long Call
Short Call	Short Stock + Short Put
Short Put	Long Stock + Short Call

Suppose a market maker finds the 10 calls, expiring in one month, trading at 45c and the puts at 35c with the underlying trading at \$10. The market maker simply puts the three pieces together- selling the call, buying the put and buying the stock. He takes in 10c and at the same time hedges a way his exposure to any changes in the price of the stock. Lets assume that the carrying the stock until expiry costs \$10 x 5% x (30/365) = 4c. his net profit on the position, assuming no other costs and no other risks is about 6 cents, which can earned with no price exposure. All calculations should be multiplied by 1000 because options cover 1000 shares.

There is no reason to think of a conversion exclusively in terms of buying share s and selling them synthetically. From Table 8.2, it is clear that a conversion can be viewed in terms of other pieces. A conversion can be either a long call and a short synthetic call, or a short put and a long synthetic put, as well as long stock and synthetic short stock.

The opposite strategy, a reverse conversion or reversal, can be established if the call and put prices were out of line in the opposite direction. If, for example, the 10 call were offered at 40c and the put bid at 38c, market maker could buy the under priced call, sell the expensive put and sell the stock short for a net debit of 2c. He could then earn interest on the \$10 he received from the sale of the stock to generate a net positive return with no directional exposure.

As the current level of interest rates determines whither a conversion or reversal will be profitable these spreads are known as interest rate plays. Using his own appropriate current interest rate, a market maker calculates his cost of carry for the position on including the receipt of the dividend (long stock) or the payment of one(short stock). He then knows the size of the credit or debit that would make a conversion or reversal profitable, and can examine current option prices with those values in mind.

Market makers are subject to interest rate risk prior to expiry. An increase in rates, increases the cost of carry, a reduction in rates, reduce the size of the interest earned. For this reason market makers generally try to balance the number of conversions and reversals.

Other Risks

Changes in implied volatility levels and to dividends are other risks that market makers deal with on an ongoing basis. Whilst neither of these should be underestimated none is as great as directional price risk. Competition among market makers often forces them to accept risks just to be included in trades, however most will not accept directional risk for more than a very short time.

Source: http://www.asx.com.au (retrieved on November 22, 2007)